5-1-2007

Distributed Cooperative Framework and Algorithms for wireless Network Performance Optimization

Emily Hong Qi
Portland State University

Recommended Citation

Qi, Emily Hong, "Distributed Cooperative Framework and Algorithms for wireless Network Performance Optimization" (2007). *Dissertations and Theses.* Paper 6142.

DISTRIBUTED COOPERATIVE FRAMEWORK AND ALGORITHMS FOR WIRELESS NETWORK PERFORMANCE OPTIMIZATION

by

EMILY HONG QI

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
ELECTRICAL AND COMPUTER
ENGINEERING

Portland State University
©2007

DISSERTATION APPROVAL

The abstract and dissertation of Emily Hong Qi for the Doctor of Philosophy in Electrical and Computer Engineering were presented May 1, 2007, and accepted by the dissertation committee and the doctoral program.

COMMITTEE APPROVALS: [redacted]

Fu Li, Chair

[redacted]

James Morris

[redacted]

Douglas Hall

[redacted]

Xiaoyu Song

[redacted]

Kwok-Wai Tam
Representative of the Office of Graduate Studies

DOCTORAL PROGRAM APPROVAL: [redacted]

Malgorzata Chrzanowska-Jeske, Director
Electrical and Computer Engineering
Ph.D. Program

ABSTRACT

An abstract of the dissertation of Emily Hong Qi for the Doctor of Philosophy in Electrical and Computer Engineering presented May 1, 2007.

Title: Distributed Cooperative Framework and Algorithms for Wireless Network Performance Optimization

In Wireless Local Access Networks (WLANs), the Medium Access Control (MAC) protocol is the primary element that determines the efficiency of sharing the limited communication bandwidth of the wireless channel. IEEE 802.11 MAC uses the contention-based Distributed Coordination Function (DCF) as a fundamental medium access mechanism. However, the dynamic nature of the wireless environment creates mobility challenges of maintaining maximum channel capacity, of obtaining optimal throughput and latency, and of retaining good security in a distributed wireless network.

This dissertation first introduces a set of parameters to characterize the medium status and radio environment, and a mechanism for mobile devices to exchange measurements in order to obtain broad and comprehensive knowledge of the wireless environment. Then the dissertation proposes a distributed cooperative wireless architecture and framework, and three cooperative algorithms to optimize wireless network performance. The cooperative algorithms allow wireless devices to cooperatively adjust configurations and optimize operations based on the characteristics of the environment.

The first algorithm adaptively adjusts the contention window size to reduce the number of collisions as the number of mobile devices increases, in order to reach maximum channel utilization. However, if a channel reaches the saturated state, the throughput per user decreases significantly. Therefore, the second algorithm discussed in this dissertation is to select the best Access Point (AP) in overlapped AP coverage areas to balance network loads and maximally utilize the network capacity. When the mobile device transitions from one AP to another AP, it may take milliseconds to seconds due to required re-association and re-authentication with the new AP. Thus, the third cooperative algorithm optimizes the device transition to provide an acceptable balance of latency and security. The corresponding simulation or experiment results that demonstrate a significant improvement of wireless network performance are explained for each algorithm.

Forgery and confidentiality are major concerns for distributed radio resource measurement and cooperation. Thus, this dissertation concludes with an analysis of security threats to radio resource measurement and cooperation, and proposes an action frame protection scheme to ensure secure distributed cooperative wireless networks.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Professor Fu Li for his supervision and guidance during the research and thesis preparation. In addition, special thanks to Prof. James Morris, Prof. Douglas Hall, Prof. Xiaoyu Song and Prof. Kwok-Wai Tam for their review of the thesis and guidance. Thanks also to Dr. Jesse Walker, Ms. Jolinda Osborne, Dr. Charlie Tai and Dr. George Wang for their generous support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACK	acknowledgment
AP	access point
BSS	basic service set
CCA	clear channel assessment
CBC	cipher-block chaining
CCM	counter mode with CBC-MAC
CCMP	counter mode with CBC-MAC Protocol
CF	contention free
CFP	contention-free period
CP	contention period
CS	carrier sense
CSMA/CA	carrier sense multiple access with collision avoidance
CTS	clear to send
CW	contention window
DCF	distributed coordination function
DIFS	distributed (coordination function) inter-frame space
DSSS	direct sequence spread spectrum
DSSS-OFDM	PHYs using DSSS-OFDM modulation under 19.7 rules
EAP	Extensible Authentication Protocol (IETF RFC 3748 [B23])
EAPOL	Extensible Authentication Protocol over LANs (IEEE 802.1X™-2004)

EDCA	enhanced distributed channel access
ESS	extended service set
FH	frequency hopping
FHSS	frequency-hopping spread spectrum
GTK	group temporal key
HC	hybrid coordinator
HCC	hyperbolic congruence code
HCCA	HCF control led channel access
HCF	hybrid coordination function
IBSS	independent basic service set
MAC	medium access control
MIB	management information base
MPDU	MAC protocol data unit
MSDU	MAC service data unit
NAV	network allocation vector
NIC	network interface card
PC	point coordinator
PCF	point coordination function
PHY	physical layer
PMK	pairwise master key
PTK	pairwise transient key
QoS	quality of service

RCPI	received channel power indicator
RF	radio frequency
RIC	resource information container
RPI	receive power indication
RSNI	received signal to noise indicator
RSSI	receive signal strength
RTS	request to send
SNMP	simple network management protocol
STA	station
SSID	service set identifier
TC	traffic category
TS	traffic stream
TU	time unit
TX	transmit or transmitter
TXOP	transmission opportunity
WLAN	wireless LAN
WEP	wired equivalent privacy
Wi-Fi	wireless fidelity (a promulgated term by the Wi-Fi Alliance)
WM	wireless medium
WPA	Wi-Fi Protected Access

1. INTRODUCTION

In recent years, Ethernet has been the predominant local area network technology for supporting distributed computing. The proliferation of portable and laptop computers has spurred the need for Local Area Network (LAN) technology to support wireless connectivity. IEEE 802.11 based wireless devices and networks are becoming increasingly popular and are beginning to be deployed nearly everywhere—from the office to the home to public hotspots in cafes and restaurants. IEEE 802.11 wireless local area networks, or WLANs, operate in the unlicensed radio-frequency (RF) spectrum, making them quick and inexpensive to deploy. Businesses and consumers enjoy both freedom and economy as a diverse set of devices become Wi-Fi ready.

1.1 Introduction to IEEE 802.11

In 1997 the IEEE adopted IEEE Std. 802.11-1997, the first wireless LAN (WLAN) standard [1]. This standard defines the media access control (MAC) and physical (PHY) layers for a LAN with wireless connectivity. It addresses local area networking where the connected devices communicate via radio with other devices that are within close proximity. The standard is similar in most respects to the IEEE 802.3 Ethernet standard, as described in Figure 1. Specifically, the 802.11 standard addresses:

- Functions required for an 802.11 compliant device to operate either in a peer-to-peer fashion or be integrated with an existing wired LAN

- Operation of the 802.11 device within possibly overlapping 802.11 wireless LANs and the mobility of this device between multiple wireless LANs
- MAC level access control and data delivery services to allow upper layers of the 802.11 network
- Several physical layers signaling techniques and interfaces
- Confidentiality of user data being transferred over the wireless media

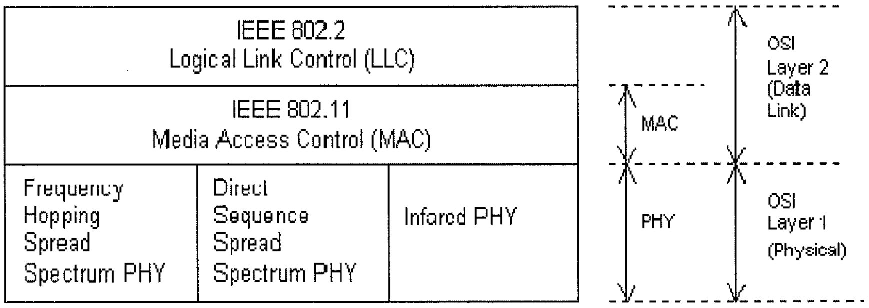

IEEE 802.2
Logical Link Control (LLC)
IEEE 802.11
Media Access Control (MAC)
Frequency Hopping Spread Spectrum PHY
Direct Sequence Spread Spectrum PHY
Infared PHY
MAC
PHY
OSI Layer 2 (Data Link)
OSI Layer 1 (Physical)

Figure 1: IEEE 802.11 standards mapped to the OSI reference model

1.1.1 Physical (PHY) Layer

The 802.11 PHY is the interface between the MAC and the wireless media where frames are transmitted and received. The PHY provides three functions. First, the PHY provides an interface to exchange frames with the upper MAC layer for transmission and reception of data. Secondly, the PHY uses signal carrier and spread spectrum modulation to transmit data frames over the media. Thirdly, the PHY provides a carrier-sense indication back to the MAC to verify activity on the media.

The 802.11 standard [1] provides three different PHY definitions. Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) support 1 and 2 Mbps data rates. An extension to the 802.11 architecture, the IEEE

802.11a [2], defines different multiplexing techniques that can achieve data rates up to 54 Mbps. Another extension to the standard, IEEE 802.11b [2], defines 11 Mbps and 5.5 Mbps data rates (in addition to the 1 and 2 Mbps rates) utilizing an extension to DSSS called High Rate DSSS (HR/DSSS). The IEEE 802.11b also defines a rate shifting technique where 11 Mbps networks may fall back to 5.5 Mbps, 2 Mbps, or 1 Mps under noisy conditions or to inter-operate with legacy 802.11 PHY layers. In June 2003, a third modulation standard 802.11g [3] was ratified. This standard works in the 2.4 GHz band (like IEEE 802.11b) but operates at a maximum raw data rate of 54 Mb/s, or about 24.7 Mb/s net throughputs (like IEEE 802.11a).

In January 2004, IEEE announced that it had formed a new 802.11 Task Group n (TGn) to develop a new amendment to the IEEE 802.11 standard for wireless local area networks. The real data throughput is estimated to reach a theoretical 540 Mbit/s, and should be up to 50 times faster than IEEE 802.11b, and up to 10 times faster than IEEE 802.11a or IEEE 802.11g. IEEE 802.11n [4] builds upon previous IEEE 802.11 standards by adding MIMO (multiple-input multiple-output). MIMO uses multiple transmitter and receiver antennas to allow for increased data throughput via spatial multiplexing and increased range by exploiting the spatial diversity. In March 2007, the IEEE 802.11n draft passed the working group letter ballot, and a new study group in IEEE 802.11 was formed to study a very high throughput technology that can deliver (at a minimum) an actual data throughput of at least 1 Gbps.

1.1.2 Media Access Control (MAC) Layer

The 802.11 MAC layer provides functionality to allow reliable data delivery for the upper layers over the wireless PHY media. The data delivery itself is based on an asynchronous, best-effort, connectionless delivery of MAC layer data. There is no guarantee that the frames will be delivered successfully. The 802.11 MAC provides a control led access method to the shared wireless media called Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA). CSMA/CA is similar to the collision detection access method deployed by 802.3 Ethernet LANs. The third function of the 802.11 MAC is to protect the data being delivered by providing security services including authentication, authorization, integrity, and confidentiality.

1.2 Distributed Wireless Network

The dynamic nature of the wireless environment creates challenges for mobile devices to maintain maximum channel capacity and obtain great throughput and low latency in a distributed wireless network. Research has begun to address this challenge by defining cooperative algorithms to improve mobile station performance. The throughput and delay limits were discussed in [5][6] . It motivated research to seek link adaptation, to optimize the wireless link. In [7] [8], data rate adaptation for IEEE 802.11a and IEEE 802.11b is discussed. In [9][10] , an adaptive back-off algorithm is also obtained from theoretical analysis. Most research is based on a perfect channel and the assumption that every mobile user always has traffic ready to send. However, in a real wireless environment channel load condition and signal quality issues including interference and noise are dynamically changed and are difficult to predict.

To improve the WLAN capacity and scalability, enterprises deploy a large number of Access Points (AP), resulting in High-Density 802.11 networks. High density networks cause devices to transition among APs during on-going communication even without client movement. Improved WLAN security and QoS as introduced in IEEE 802.11i [11] and IEEE 802.11e [12], however, have introduced additional delay to device transition that increases the possible transition time from a few hundred milliseconds to several seconds. Indeed, IEEE 802.11i adds a minimum of two round trips and a series of cryptographic operations to each transition, and IEEE 802.11e adds at least one round trip and a complex bandwidth allocation algorithm. The latency introduced by these new services can greatly exceed the real-time constraints imposed by VoIP.

This research focuses on distributed wireless network optimizations based on the real time radio environment data and the cooperation between AP and mobile devices. This dissertation proposes a set of parameters to characterize the radio resource environment. The dissertation analyses the medium sensing and radio environment parameters, and introduces a distributed cooperative wireless framework to enable wireless devices to adjust configurations adaptively. The dissertation also proposes three cooperative algorithms to improve throughput and reduce the latency while retaining good security. By coordinating the radio resource measurement protocol defined in the IEEE 802.11 Task Group k [13], the cooperative algorithms enable Access Point and mobile device collaboration and improvement of the overall wireless network performance.

The first algorithm presented in the dissertation adaptively configures the contention window size to reduce the number of collisions as the number of mobile devices increases; and thus allow devices to maintain the maximum channel utilization. The fraction of channel bandwidth used by successfully transmitted messages gives a good indication of the overheads required by the MAC protocol to perform its coordination task among stations. This fraction is known as the utilization of the channel, and the maximum value it can attain is known as the capacity of the MAC protocol. In this dissertation a performance model and the theoretical limit of the IEEE 802.11 MAC protocol capacity are derived to show that if a station has an exact knowledge of the wireless network status and radio environment, it is possible to adjust its configuration parameters to achieve a protocol capacity very close to its theoretical limit, and furthermore, to improve the wireless network performance.

However, if the channel reaches its saturated state, the throughput per user decreases significantly regardless of adaptation of contention windows size. The second algorithm proposed in this dissertation is a cooperative scheme for load balancing to achieve higher throughput. 802.11-based networks can be set up in two modes – Infrastructure and ad-hoc. In the Infrastructure mode, mobile devices associate with APs, and traffic between mobile devices is routed via these APs. In the ad-hoc mode, clients establish an ad-hoc network by communicating directly with each other. The infrastructure mode of operation presents an interesting challenge. Due to the probabilistic/problematic nature of DCF, the performance of a client depends heavily on network capacity and the number of mobile devices competing for access.

Cooperative load balancing and AP selection algorithm balances network load and maximally utilizes the network capacity.

After selecting a better AP, the mobile client needs to transition from the old AP to the new AP. To accelerate the mobile device transition without compromising security, the third algorithm explores an AP and mobile device cooperative transition scheme that preserves both security and latency without disrupting existing transfers as we proposed to IEEE 802.11r draft [14]. The dissertation also analyzes the efficacy of the transition optimization, and then identifies security flaws in the current design and proposes simple corrections.

Authentication, Authorization, Confidentiality and Integrity are major concerns for radio source measurement and cooperation. This dissertation concludes with an analysis of the security threats to distributed radio resource measurement and cooperation, and then proposes a management frame protection scheme by extending the existing IEEE 802.11i security scheme. The goal of the proposed security techniques is to protect IEEE 802.11 Action Frames without introducing a new encryption and key management scheme. The protection scheme proposed is backward compatible with the existing Action Frames used widely in other IEEE 802.11 amendments, so it extends automatically to protect many other management messages besides those of radio resource measurement and cooperation messages.

1.3 Outline of This Dissertation

This chapter serves as an introduction to the entire dissertation. A brief synopsis of the remaining chapters follows.

Chapter 2 contains the overview of IEEE 802.11 MAC protocols and the different topologies incorporated to accommodate the unique characteristics of the IEEE 802.11 wireless LAN standard and amendments. A performance analysis of 802.11 MAC protocol based on the OPNET simulation is also given at the end of the chapter.

Chapter 3 introduces local medium sensing and radio measurement, and discusses the radio resource measurement exchange between the mobile stations or mobile station and AP by introducing IEEE 802.11k. Based on the radio sensing and resource monitoring, a distributed cooperative wireless architecture is presented.

Chapters 4, Chapter 5 and Chapter 6 present three distributed cooperative algorithms to optimize the wireless network performance.

Chapter 4 includes an analytical performance model derived for channel utilization of Wireless LANs for an ideal channel (no interference, no hidden nodes). The analytical model shows that the maximum channel utilization can be achieved when an optimum contention window is selected. With the adaptive contention window Size algorithm, the number of collisions is reduced as the number of users increases to maintain the maximum channel utilization. The adaptive contention window size algorithm is also validated against simulated performance at the end of the chapter.

If the channel reaches its saturated state, the throughput per user decreases significantly. Chapter 5 presents an algorithm that selects the best AP in overlapped infrastructure BSS to balance network load and utilize the network capacity.

Chapter 6 describes cooperative transition optimizations, along with a security and latency evaluation of the existing design. Furthermore, a security analysis of the new transition process and a test bed to evaluate the efficacy of the design are presented at the end of the chapter.

Chapter 7 contains a threat analysis, and suggests an extension of IEEE 802.11i to protect Action Frames. Since Action Frames are widely used in IEEE 802.11 amendments, the Action Frame protection mechanism extends automatically to these messages, as well.

Chapter 8 concludes of the dissertation and outlines some future research that the author is interested in pursuing.

2 THE 802.11 MAC PERFORMANCE ANALYSIS

A number of characteristics are unique to the wireless environment (as compared to a wired LAN) and must be taken into consideration by wireless LAN design. The physical characteristics of a wireless LAN introduce range limitations and unreliable media, dynamic topologies where stations move about, interference from outside sources, and the inability of every device to 'hear' every other device within the WLAN.

The 802.11 standard specifies a common Medium Access Control (MAC) Layer to address these characteristics. In general, the MAC Layer manages and maintains communications among 802.11 stations by coordinating access to a shared radio channel and by utilizing protocols that enhance communications over the radio medium. Medium Access, Scanning, Authentication, Association, and Fragmentation are the major functions of the 802.11 MAC Layer.

This chapter begins with an overview of the IEEE 802.11 MAC architecture and the different topologies incorporated to accommodate the unique characteristics of the IEEE 802.11 wireless LAN standard, and then proceeds to focus on the Basic Service Set (BSS) transition procedure and the Distributed Contention Based Access. A performance analysis based on OPNET simulation results is also presented. The performance analysis paves the way for the introduction of the distributed cooperative architecture in chapter 3.

2.1 IEEE 802.11 MAC Architecture

2.1.1 Components and Services

The 802.11 architecture is comprised of several components and services that interact to provide station mobility transparent to the higher layers of the network stack.

Wireless LAN Station

The station (STA) is the most basic component of the wireless network. A station is any device that contains the functionality of the 802.11 protocol--MAC, PHY, and a connection to the wireless media. Typically the 802.11 functions are implemented in the hardware and software of a network interface card (NIC). A station could be a laptop PC, handheld device, or an Access Point. Stations may be mobile, portable, or stationary, and all stations support the 802.11 station services of authentication, de-authentication, privacy, and data delivery.

Basic Service Set (BSS)

IEEE 802.11 defines the Basic Service Set (BSS) as the basic building block of an 802.11 wireless LAN. The BSS consists of a group of any number of stations. The 802.11 topology can be an independent Basic Service Set (IBSS) as well an Infrastructure Basic Service Set (BSS).

Independent Basic Service Set (IBSS)

The most basic wireless LAN topology is a set of stations which have recognized each other and are communicating via the wireless media in a peer-to-peer fashion. This form of network topology is referred to as an Independent Basic Service

Set (IBSS) or an Ad-hoc network, as depicted in Figure 2. In an IBSS, the mobile stations communicate directly with each other, though every mobile station may not be able to communicate with every other station due to the range limitations. There are no relay functions in an IBSS; therefore, all stations need to be within range of each other and communicate directly.

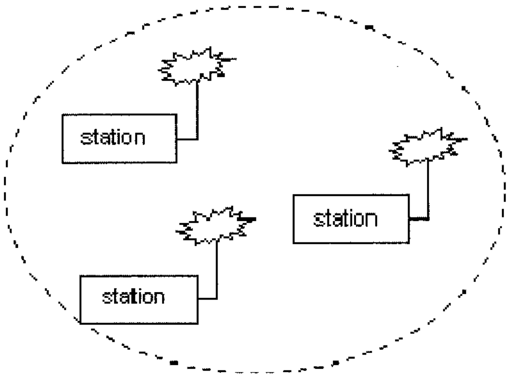


Figure 2: Independent Basic Service Set (IBSS)

Infrastructure Basic Service Set

An Infrastructure Basic Service Set is a BSS with a component called an Access Point (AP), see Figure 3. The access point provides a local relay function for the BSS. All stations in the BSS communicate with the access point and no longer communicate directly. All frames are relayed between stations by the access point. This local relay function effectively doubles the range of the IBSS. The access point may also provide connection to a distribution system.

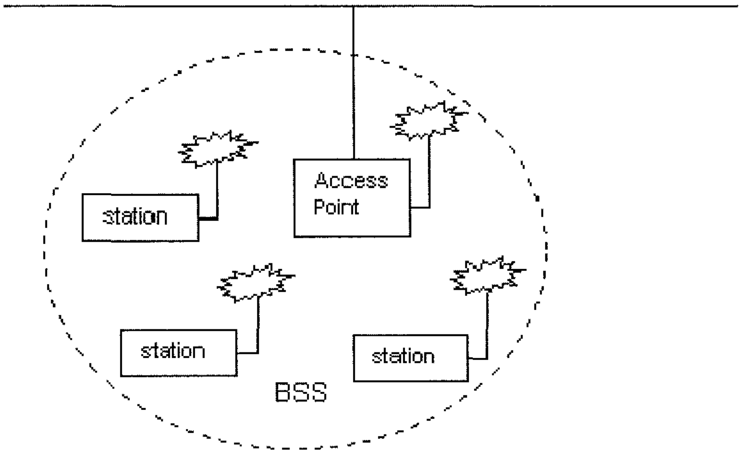


Figure 3: Infrastructure Basic Service Set

Distribution System (DS)

The distribution system (DS) is the means by which an access point communicates with another access point to exchange frames for stations in their respective BSSs, and exchanges frames with a wired network. As IEEE 802.11 describes it, the distribution system is not necessarily a network nor does the standard place any restrictions on how the distribution system is implemented. Thus the distribution system may be a wired network like 803.2 or a special purpose box that interconnects the access points and provides the required distribution services.

Extending coverage via an Extended Service Set (ESS)

IEEE 802.11 extends the range of mobility to an arbitrary range through the Extended Service Set (ESS) as shown in Figure 4. An extended service set is a set of infrastructure BSSs, where the access points communicate among themselves via a DS. The DS is the backbone of the LAN and may be constructed of either a wired LAN or wireless network.

Typically the distribution system is a thin layer in each access point that determines the destination for traffic received from a BSS. The distribution system determines if traffic should be relayed back to a destination in the same BSS, forwarded on the distribution system to another access point, or sent into the wired network to a destination not in the extended service set. Communications received by an access point from the distribution system are transmitted to the BSS to be received by the destination mobile station.

Network equipment outside of the extended service set views the ESS and all of its mobile stations as a single MAC-layer network, with all stations viewed as physically stationary. Thus, the ESS hides the mobility of the mobile stations from everything outside the ESS. This level of indirection provided by the 802.11 architecture allows existing network protocols that have no concept of mobility to operate correctly with a wireless LAN where there is mobility.

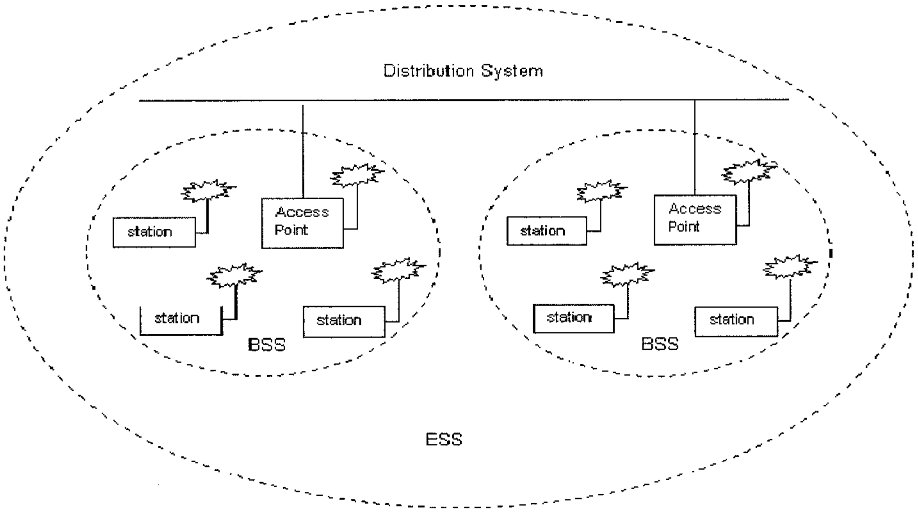


Figure 4: Extended Service Set (ESS)

2.1.2 MAC Layer Functions

The 802.11 standard specifies a common medium access control (MAC) Layer, which provides a variety of functions that support the operation of 802.11-based wireless LANs. In general, the MAC Layer manages and maintains communications between 802.11 stations (radio network cards and access points) by coordinating access to a shared radio channel and utilizing protocols that enhance communications over a wireless medium. Often viewed as the "brains" of the network, the 802.11 MAC Layer uses an 802.11 Physical (PHY) Layer, such as 802.11b or 802.11a, to perform the tasks of carrier sensing, transmission, and receiving of 802.11 frames.

The following summarizes the primary 802.11 MAC functions, especially as they relate to infrastructure wireless LANs.

Medium Access Functions

Before transmitting frames, a station must first gain access to the medium, which is a radio channel that stations share. The 802.11 standard defines two forms of medium access, distributed coordination function (DCF) and point coordination function (PCF). DCF is mandatory and based on the carrier-sense multiple accesses with collision avoidance (CSMA/CA) protocol. With DCF, 802.11 stations contend for access and attempt to send frames when there is no other station transmitting. If another station is sending a frame, stations are polite and wait until the channel is free.

An important aspect of the DCF is a random back-off timer that a station uses if it detects a busy medium. If the channel is in use, the station must wait a random period of time before attempting to access the medium again. This back-off ensures that multiple stations wanting to send data don't transmit at the same time. The random delay causes stations to wait different periods of time and minimizes the chance of all of them sensing the medium at exactly the same time, finding the channel idle, transmitting, and colliding with each other. The back-off timer significantly reduces the number of collisions and corresponding retransmissions, especially when the number of active users increases. With radio-based LANs, a transmitting station cannot listen for collisions while sending data, mainly because the station cannot have its receiver on while transmitting the frame. As a result, the receiving station needs to send an acknowledgement (ACK) if it detects no errors in the received frame. If the sending station does not receive an ACK after a specified period of time, the sending station will assume that there was a collision (or RF interference) and retransmit the frame. The next section provides a detailed study of DCF.

For supporting time-bounded delivery of data frames, the 802.11 standard defines the optional point coordination function (PCF) where the access point grants access to an individual station to the medium by polling the station during the contention free period. Stations cannot transmit frames unless the access point polls them first. The period of time for PCF-based data traffic (if enabled) occurs alternately between contention (DCF) periods. The access point polls stations according to a polling list, then switches to a contention period when stations use DCF. This process enables support for both synchronous (i.e., video applications) and asynchronous (i.e., e-mail and Web browsing applications) modes of operation.

Scanning

The 802.11 standard defines both passive and active scanning, whereby a radio Network Interface Card (NIC) searches for access points. Passive scanning is mandatory where each NIC scans individual channels to find the best access point signal. Periodically, access points broadcast a Beacon, and the radio NIC receives these Beacons while scanning, and takes note of the corresponding signal strengths. The Beacons contain information about the access point, including service set identifier (SSID), supported data rates, etc. The radio NIC can use this information along with the signal strength to compare access points and decide which to use.

Optional active scanning is similar, except that the radio NIC initiates the process by broadcasting a Probe Request frame. Access points within range may reply with a Probe Response. Active scanning enables a radio NIC to receive immediate response from access points, without waiting for a beacon transmission. The issue, however, is that active scanning imposes additional overhead on the network because of the transmission of Probe Requests and the corresponding Response frames.

Authentication

Authentication is the process of proving identity, and the 802.11 standard specifies two forms: Open System Authentication and Shared Key Authentication.

Open system authentication is a mandatory two step process. A radio NIC first initiates the process by sending an authentication request frame to the access point. The access point replies with an authentication response frame containing approval or disapproval of authentication indicated in the Status Code field in the frame body.

Shared Key Authentication is an optional four step process that was intended to demonstrate that an authenticating device possesses the correct Wired Equivalent Privacy (WEP) key. The radio NIC starts by sending an authentication request frame to the access point. The access point responds with an authentication response message conveying a challenge text in its frame body. The radio NIC uses its WEP key to encrypt the challenge text and then sends it back to the access point in another authentication frame. The access point decrypts the challenge text and compares it to the initial text. If the text is equivalent, then the access point assumes that the radio NIC has the correct key. The access point finishes the sequence by sending an authentication frame to the radio NIC with the approval or disapproval. Shared Key Authentication is flawed, however, and has been deprecated. In particular, WEP encryption is based on XORing a key stream with plaintext, so an attacker can extract the WEP key stream by XORing the public challenge with the public response, and use the result to authenticate without ever possessing the WEP key used to generate the key stream.

(Re)Association

Once authenticated, the radio NIC must associate with the access point before sending data frames. Association is necessary to synchronize the radio NIC and access point with important information, such as supported data rates. It also causes the access point to reconfigure the DS to forward the STA's data traffic.

The radio NIC initiates an association by sending an Association Request frame containing elements such as SSID and supported data rates. The access point responds by sending an Association Response frame containing an association Identifier, along with other information regarding the access point. Once the radio NIC and access point complete the association process, they can send data frames to each other.

WEP

With the optional WEP enabled, the wireless NIC will encrypt the body (not header) of each frame before transmission using a common key, and the receiving station will decrypt the frame upon receipt using the common key. The 802.11 standard specifies a 40-bit key and misuses encryption, which makes 802.11 wireless LANs vulnerable to attack; see, e.g., [15] [16][17]. IEEE 802.11i standard [11] improves IEEE 802.11 security by incorporating 802.1X authentication and stronger encryption into the standard as described in the next section.

RTS/CTS

The optional request-to-send and clear-to-send (RTS/CTS) function allows the access point to control use of the medium for stations activating RTS/CTS. With most radio NICs, users can set a maximum frame length threshold whereby the radio NIC will activate RTS/CTS. For example, a frame length of 1,000 bytes will trigger RTS/CTS for all frames larger than 1,000 bytes. The use of RTS/CTS alleviates hidden node problems, where two or more radio NICs cannot hear each other and they are associated with the same access point.

If the radio NIC activates RTS/CTS, it will first send a RTS frame to the access point before sending a data frame. The access point will then respond with a CTS frame, indicating that the radio NIC can send the data frame. With the CTS frame, the access point will provide a value in the duration field of the frame header that holds off other stations from transmitting until after the radio NIC initiating the RTS can send its data frame. This feature avoids collisions between hidden nodes. The RTS/CTS handshake continues for each frame, as long as the frame size exceeds the threshold set in the corresponding radio NIC.

Fragmentation

The optional fragmentation function enables an 802.11 station to divide data packets into smaller frames. This is done to avoid needing to retransmit large frames in the presence of RF interference. It was also introduced to make it easier for data and voice to coexist at low speeds by chopping large data packets into fragments; for instance, at 1 Mbps, the lowest 802.11b speed, a single full sized 802.11 frame (about 2600 octets) consumes the entire channel for 21 milliseconds, whereas each voice call must access the channel every 20 milliseconds. The bits errors resulting from RF interference are likely to affect a single frame, and it requires less overhead to retransmit a smaller frame than a larger one. As with RTS/CTS, users can generally set a maximum frame length threshold whereby the radio NIC will activate fragmentation. If the frame size is larger than the threshold, the radio NIC will break the packet into multiple frames, with each frame no larger than the threshold value.

2.2 QoS and Security Enhancements

One of the amendments to the IEEE 802.11 Media Access Control (MAC) layer is IEEE 802.11e [12]. IEEE 802.11e as of late 2005 was approved as a standard that defines a set of Quality of Service (QoS) enhancements for LAN applications. The standard is considered of critical importance for delay-sensitive applications, such as Voice over Wireless IP and Streaming Multimedia. The 802.11e enhances the DCF and the PCF, through a new coordination function: the Hybrid Coordination Function (HCF). Within the HCF, there are two methods of channel access, similar to those defined in the legacy 802.11 MAC: HCF Control led Channel Access (HCCA) and Enhanced Distributed Channel Access (EDCA). Both EDCA and HCCA define Traffic Classes (TC). For example, emails could be assigned to a low priority class, and Voice over Wireless WLAN (VoWLAN) could be assigned to a high priority class. In addition to HCCA, EDCA and TXOP, 802.11e specifies additional optional protocols for enhanced 802.11 MAC layer QoS.

IEEE 802.11i [11], is an amendment to the 802.11 standard specifying security mechanisms for wireless networks. The draft standard was ratified on 24 June 2004, and supersedes the previous security specification, Wired Equivalent Privacy (WEP), which was shown to have severe security weaknesses. Wi-Fi Protected Access (WPA) had previously been introduced by the Wi-Fi Alliance as an intermediate solution to WEP insecurities. WPA implemented a subset of 802.11i. The Wi-Fi Alliance refers to their approved, interoperable implementation of the full 802.11i as WPA2. IEEE 802.11i makes use of the Advanced Encryption Standard (AES) block cipher; WEP and WPA use the RC4 stream cipher.

The IEEE 802.11i architecture contains the following components: 802.1X for authentication (entailing the use of EAP and an authentication server), RSN for keeping track of associations, and AES-based CCMP to provide confidentiality, integrity origin authentication, and the four-way handshake.

The authentication process leaves two considerations: the access point (AP) still needs to authenticate itself to the client station (STA), and keys to encrypt the traffic need to be derived. The earlier EAP exchange has provided the shared secret key Pairwise Master Key (PMK). This key, however, is designed to last the entire session and should be exposed as little as possible. Therefore the four-way handshake is used to establish another key called the Pairwise Transient Key (PTK). The PTK is generated by concatenating the following attributes: PMK, AP nonce (ANonce), STA nonce (SNonce), AP MAC address and STA MAC address. The product is then put through a cryptographic hash function.

The handshake also yields the Group Temporal Key (GTK), used to decrypt multicast and broadcast traffic. The GTK used in the network may need to be updated due to the expiry of a preset timer. When a device leaves the network, the GTK also needs to be updated. This is to prevent the device from receiving any more multicast or broadcast messages from the AP.

2.3 BSS Transition Procedures

A BSS network is also called an infrastructure network as described in the previous section. A Basic Service Set (BSS) is a "network segment." A BSS consists of a single AP, together with the group of associated 802.11 STAs. During transition, a mobile station leaves one BSS and joins another. A transition requires two execution phases: a Discovery phase and a Transition phase as depicted in Figure 5.

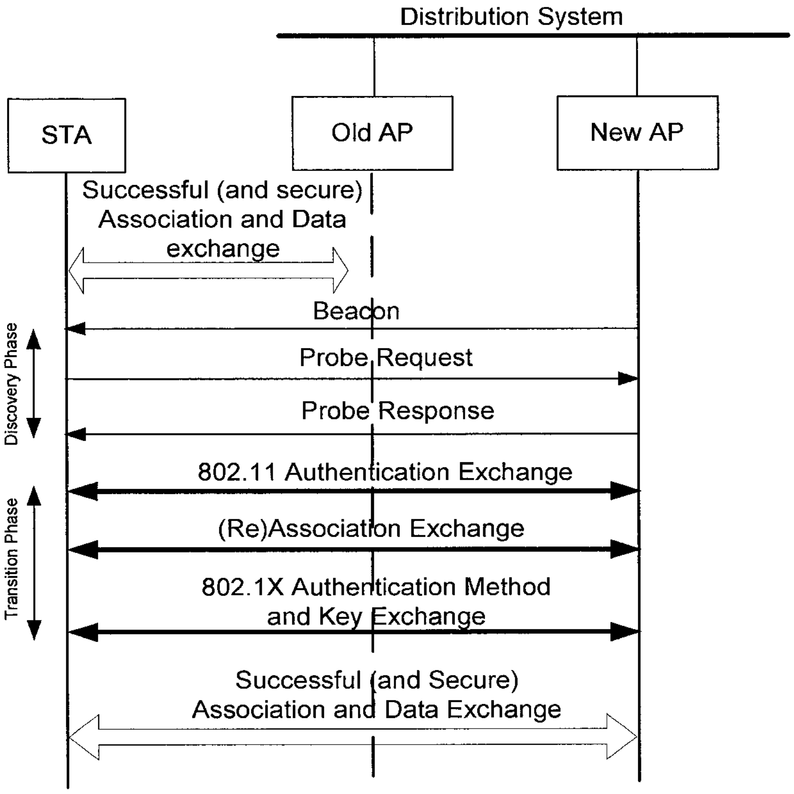

Distribution System
STA
Old AP
New AP
Successful (and secure) Association and Data exchange
Beacon
Probe Request
Probe Response
802.11 Authentication Exchange
(Re)Association Exchange
802.1X Authentication Method and Key Exchange
Successful (and Secure) Association and Data Exchange
Discovery Phase
Transition Phase

Figure 5: BSS transition procedure

2.3.1 Discovery Phase

The discovery consists of scanning to locate candidate APs, and deciding on the target AP with which to associate. A mobile station continuously scans for all of the APs within its range. Scanning is either active or passive. Passive scanning is accomplished by listening for Beacon frames, which an AP typically sends at 100 milliseconds intervals to announce its presence. Each Beacon provides timing for its BSS, and also advertises the BSS configuration and policy. A mobile station listens for Beacons on each radio channel in succession, thereby identifying APs that are using a particular channel. A mobile station actively scans by broadcasting a Probe Request message on each radio channel. An AP that receives a Probe Request replies with a Probe Response frame, which reports the same policy and configuration as does the Beacon.

For each AP discovered, the mobile station accumulates roaming triggers that are used as input to AP selection. Roaming triggers consist of the measured signal strength and signal-to-noise ratio of each AP known to be within range, as well as the frame loss rate of the current association. When the station has identified one or more candidate APs, it selects a new target AP based on the information accumulated during the scanning process.

Scanning dominates the discovery phase latency. This latency is a function of the number of radio channels available and the waiting time for each channel. In a typical enterprise environment, scanning can take up to 1 second, but a scanning latency of 200 to 300 milliseconds is typical [18]. Figure 6 depicts a typical scanning process.

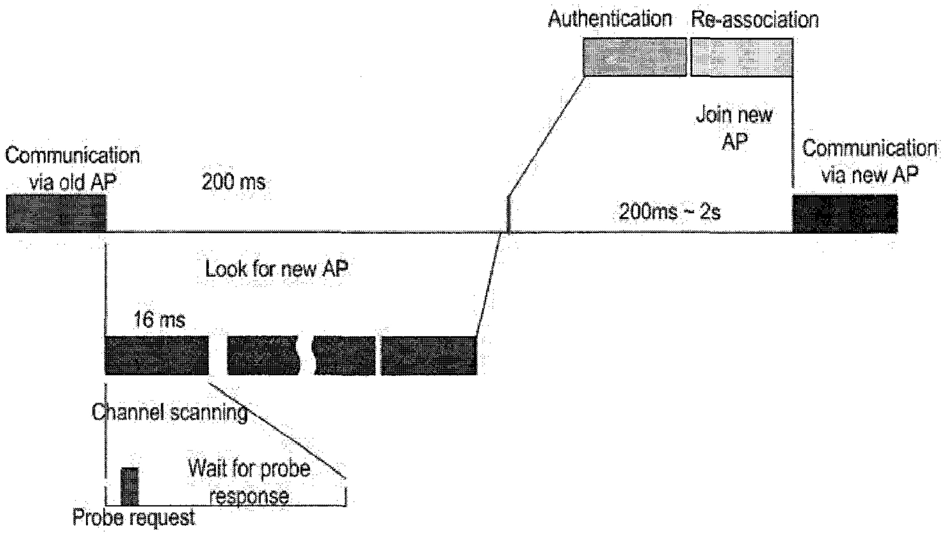


Figure 6: An example of BSS transition time

2.3.2 TRANSITION PHASE

Once a mobile station identifies a suitable candidate AP, the station breaks its association with the current AP and then re-associates with the targeted AP. The mobile station performs the following transition steps:

1. The mobile station stops data transmission to its current AP.
2. The mobile station switches its radio to the channel used by the targeted AP.
3. The mobile station completes a Reassociation Request/Response exchange with the targeted AP.
4. If Reassociation succeeds, the targeted AP is the mobile station's new AP, and the mobile station authenticates and performs 802.11i key management with the new AP, to secure its new link.

5. The mobile station requests that the new AP allocate bandwidth to maintain the quality-of-service required by its applications.
6. When these steps are completed, data flow resumes between mobile station and the infrastructure, now via the new AP.

The introduction of IEEE 802.11i security and the negotiation of QoS using 802.11e have increased this transition time from a few milliseconds to a few seconds.

2.4 Contention Based Medium Access

The fundament of 802.11 MAC is Distributed Coordination Function (DCF), which will be covered in detail in this section as a means of optimizing the IEEE 802.11 performance.

2.4.1 Overview

Networks can be divided into two categories: those using point-to-point connections and those using multi-access channels. In any multi-access channel network, the key issue is how to determine who gets to use the channel when there is competition for it. Consider a conference call where six people, on six different telephones, are all connected so that each one can hear and talk to all the others. It is very likely that when one of them stops speaking, two or more will start talking at once, resulting in chaos. The protocols used to determine who goes next on a multi-access channel belong to the sub-layer of the data link layer called the Medium Access Control (MAC) sub-layer. The MAC sub-layer is especially important in LANs, nearly all of which use a multi-access channel as the basis of their communication.

Many algorithms for allocating a multiple access channel are known. They can be categorized as distributed contention based medium access and centralized contention-free access. This dissertation will concentrate on distributed contention based medium access, because it is in these scenarios that the congestion problems could arise.

One of the most widely used distributed contention based access is Carrier Sense Multiple Access with Collision Detection (CSMA/CD). It is commonly used on LANs in the MAC sub-layer. CSMA/CD, as well as many other LAN protocols, uses the conceptual model of Figure 7. At the point marked t0, a station has finished transmitting its frame. Any other station having a frame to send may now attempt to do so. If two or more stations decide to transmit simultaneously, there will be a collision. Collisions can be detected by looking as the power or pulse width of the received signal and comparing it to the transmitted signal. After a station detects a collision, it aborts its transmission, waits a random period of time, and then tries again, assuming that no other station, in the meantime, has started transmitting.

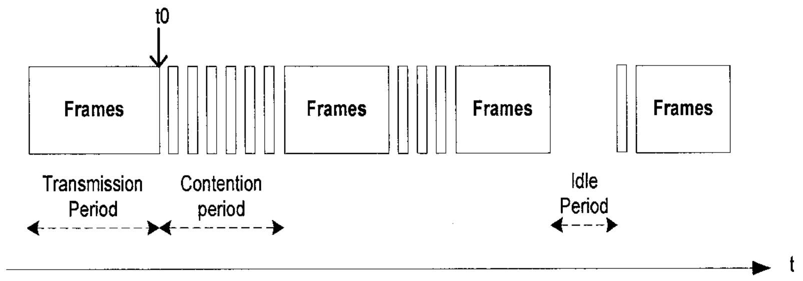
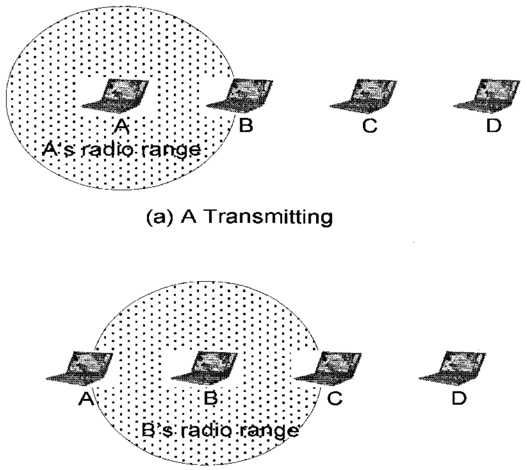


Figure 7: Transmission, contention, and idle states

As the number of portable computing and communication device grows, so does the demand to connect them to the outside world. To achieve true mobility, portable computers use radio (or infrared) signals for communication. A system of portable computers that communicate by radio can be regarded as a wireless LAN. These LANs have somewhat different properties than conventional LANs and require special MAC sub-layer protocols.

Due to the inherent flexibility of the random access system (e.g. random access allows unconstrained movement of mobile hosts), the IEEE 802.11 standard committee decided to adopt a random access CSMA-based scheme for WLANs. In this scheme there is no collision detection capability, due to a WLAN STA's inability to listen while sending. This is because a WLAN STA usually has one antenna for both sending and receiving.

A naïve approach to using a wireless LAN might be to try CSMA: just listen for other transmissions and only transmit if no one else is doing so. The trouble with this protocol is that interference occurs at the receiver, not at the sender. To see the nature of the problem, consider Figure 8 that illustrates four wireless stations. The radio range is such that A and B are within each other's range and can potentially interfere with one another. C can also potentially interfere with both B and D, but not with A.



(a) A Transmitting

(b) B Transmitting

Figure 8: A Wireless LAN

If C senses the medium, it will not hear A because A is out of range, and thus C will falsely conclude that it can transmit. If C does start transmitting, it will interfere with B, wiping out the frame from A. The problem of which a station is not able to detect a potential competitor for the medium because the competitor is too far away is called the ***hidden station problem***. If C senses the medium, it will hear an ongoing transmission and falsely conclude that it may not send to D, when in fact such a transmission would cause bad reception only in the zone between B and C, where neither of the intended receivers is located. This is called the ***exposed station problem***. The root of the problem is that before starting a transmission, a station really wants to know whether there is activity around the receiver. CSMA merely tells the station whether or not there is activity around the station sensing the carrier. With a wire, all signals propagate to all stations, so only one transmission can take place at once anywhere in the system. In a system based on short-range radio waves, multiple transmissions can occur simultaneously if they all have different destinations and these destinations are out of range of one another.

2.4.2 Distributed Coordination Function

The Distributed Coordination Function (DCF), as the basic access mechanism of the IEEE 802.11 MAC, achieves automatic medium sharing between compatible stations through the use of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) [20]. Before a station starts transmission, it senses the wireless medium to determine if it is idle. If the medium appears to be idle, the transmission may proceed, or else the station will wait until the end of the in-progress transmission. The CSMA/CA mechanism requires a minimum specified gap/space between contiguous frame transmissions. A station will ensure that the medium has been idle for the specified inter-frame interval before attempting to transmit.

The Distributed Inter-Frame Space (DIFS) is used by stations operating under the DCF to transmit data frames. A station using the DCF has to follow two medium access rules:

- The station is allowed to transit only if its carrier sense mechanism has determined that the medium has been idle for at least DIFS time, and
- In order to reduce the collision probability among multiple stations accessing the medium, the station will select a random back-off interval after deferral or prior to attempting to transmit another frame after a successful transmission.

One important characteristic of the IEEE 802.11 MAC is that an Acknowledgment (ACK) frame will be sent by the receiver upon successful reception of a data frame. It is only after receiving an ACKframe correctly that the transmitter assumes successful delivery of other corresponding data frame. The short Inter-Frame Space (SIFS), which is smaller than DIFS, is the time interval between reception of a data frame and transmission of its ACKframe. A basic medium access method is illustrated in Figure 9. Use of this small gap between transmissions within the frame exchange sequence prevents other stations – which are required to wait for the medium to be idle for a longer gap (e.g., at least DIFS time) – from attempting to use the medium, thus giving priority to completion of the in-progress frame exchange sequence.

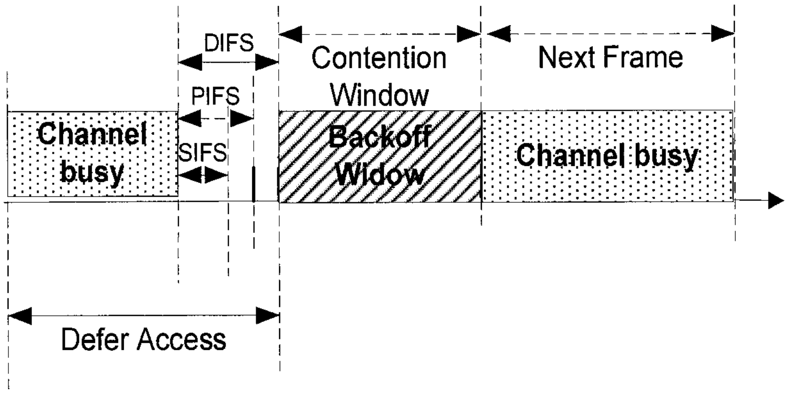


Figure 9: DCF channel access

The timing of successful frame transmissions is illustrated in Figure 10. On the other hand, if an Ack frame is received in error, i.e., received with an incorrect frame check sequence (FCS), the transmitter will re-contend for the medium to retransmit the frame after an EIFS (Extended Inter Frame Space) interval, as shown in Figure 11.

However, if no ACK frame is received within an SIFS interval, due possibly to an erroneous reception of the preceding data frame, as shown in Figure 12, the transmitter will contend again for the medium to retransmit the frame after an ACK timeout.

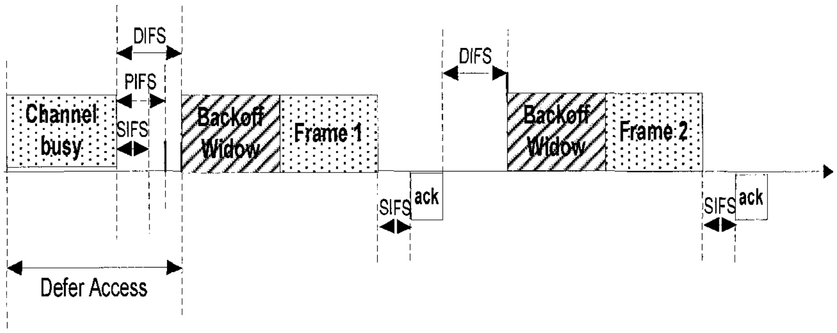


Figure 10: Timing of successful frame transmissions under the DCF

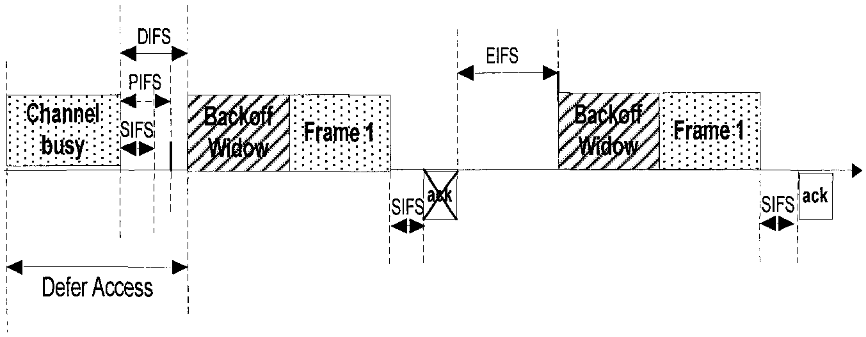


Figure 11: Frame retransmissions due to ACK failure

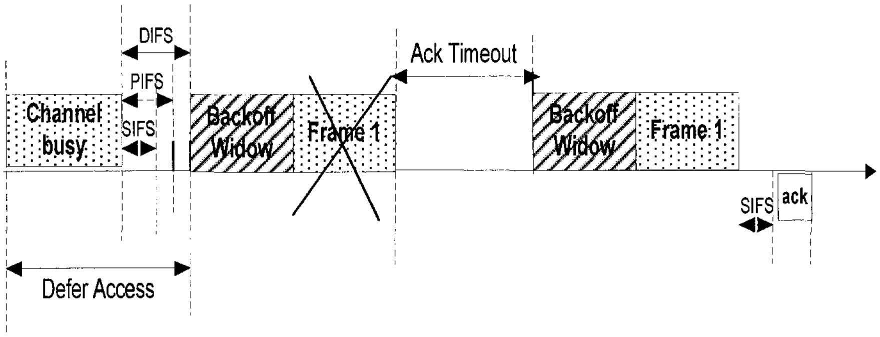


Figure 12: Frame retransmissions due to an erroneous data frame reception

If the medium is busy upon transmitting a data frame or an ACK, the transmission must be deferred until the end of the ongoing transmission. In this case, a random back-off interval is selected. To select the random back-off count υ, each station maintains a contention windows (*CW*) value. The back-off count υ is determined as a random integer drawn from a uniform distribution over the interval [0, *CW*]. The *CW* size is initially assigned *CWmin* (W), and it is increased exponentially when a transmission fails. After any unsuccessful transmission attempt, another back-off is performed with a new *CW* value determined as follows:

$$CW \leftarrow 2 \cdot (CW + 1) - 1.$$

Once *CW* reaches the value of *CWmax,* it remains at the value of *CWmax* until it is reset. The *CW* is reset to *CWmin* after a successful transmission or after reaching the maximum retry limit.

The back-off count τ is the number of idle "slots" which the station must wait for until it is allowed to transmit – there is specified and understood slot duration. The value is decremented by one for each idle slot detected. The back-off timer suspends when the medium becomes busy before υ reaches zero. The timer resumes only after the medium has been idle longer than the designed inter-frame space interval. The station starts transmitting the frame when the back-off timer reaches zero. The back-off method is used to minimize collisions and maximize throughput at both low and high network utilization.

DCF also provides an alternative way of transmitting data frames that involves transmission of RTS (Request-To-Send) and CTS (Clear-To-Send) Control frames prior to actual data transmission. This mechanism, known as RTS/CTS, is shown in Figure 13. RTS and CTS are used to reserve the channel between the transmitter and receiver. An RTS frame is transmitted by a station which needs to transmit a data packet. The receiving station responds with a CTS frame. The rules for transmission of RTS and CTS frames are the same as those of the data frame and the acknowledgement frame. RTS and CTS frames contain a duration field that tells the period of time the channel is to be reserved for transmitting the data frame. This information is picked up by other stations in the area that are sensing the channel. It helps them to construct a Network Allocation Vector (NAV) – the period of time a station is required to be kept silent. The technique is referred as virtual carrier sense mechanism and is used to reduce contentions due to hidden terminals.

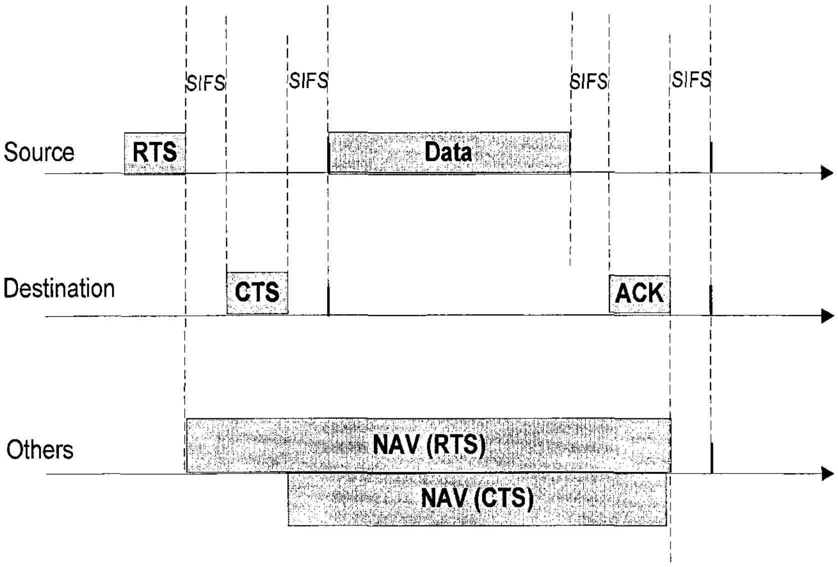


Figure 13: RTS/CTS access mechanism

The RTS/CTS mechanism is very effective in terms of system performance, especially when large packets are considered, as it reduces the length of the frames involved in the contention process. In fact, assuming perfect channel sensing by every station, collision may occur only when two (or more) packets are transmitted within the same slot time. If both transmitting stations employ the RTS/CTS mechanism, collision occurs only on the RTS frames, and it is detected early by the transmitting stations because of the lack of CTS responses.

The IEEE 802.11 standard requires that a data frame be discarded by the transmitter's MAC after a certain number of unsuccessful transmission attempts. If the length of a data frame is less than or equal to *dot11RTSThreshold*, the number of transmission attempts is limited by *dot11shortRetryLimit*, or else the maximum number of transmission attempts is set to *dot11LongRetryLimit*. The default values of *dot11ShortRetryLimit* and *dot11LongRetryLimit* are 7 and 4, respectively.

2.5 Performance Analysis

In order to evaluate the performance of current 802.11 MAC protocol design, a study was conducted of three synthetic networks in which every station could directly communication with every other node (thereby eliminating the possibility of hidden terminals).

Networks with 2, 10, and 20 stations in addition to the test station were considered, and a perfect channel was assumed. Each packet was comprised of 1K bytes plus MAC and PHY layer headers. Each station generated packets for the MAC layer in accordance with a Poisson process. Every station used the same rate in order to Control the offered load through the common packet generation rate. It was also assumed that 802.11b used direct sequence spread spectrum (DSSS) at the physical layer and used parameters according to the IEEE 802.11 specification.

For example, a network layout with 3 stations is depicted in Figure 14. All of the stations sent FTP traffics to "Server" through "AP_level_2". Client 0 was the reference client for the throughput and Medium Access Delay (MAD) analysis.

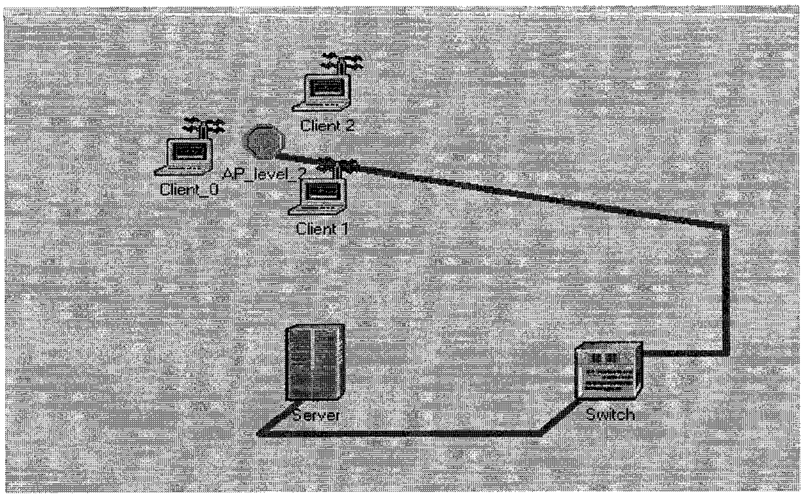


Figure 14: Network layout with 2 stations

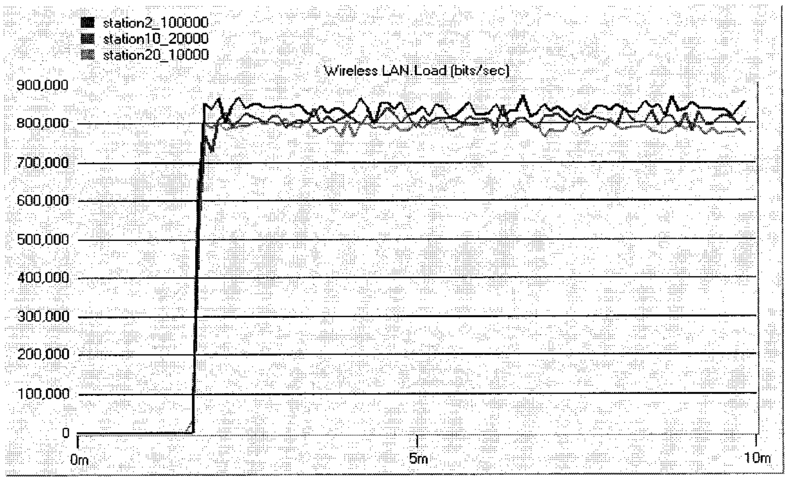


Figure 15: The global traffic loads

<u>Impact of number of users</u>

In the following simulation, global traffic load remained unchanged, and a comparison was made between the throughput and Medium Access Delay of Client 0 with varied active station count in the BSSs (2 stations, 10 stations, and 20 stations). FTP file size and inter request time of each of clients were set as described in Table 1.

Table 1: Traffic load for 2, 10, and 20 station scenarios

Traffic Pattern	**File Size (bytes)**	**Inter Request Time (seconds)**
Client 0 Setting	constant (20000)	exponential (2)
2 stations scenarios: Client 1 to Client 2 settings	constant (100000)	exponential (2)
10 stations scenarios: Client 1 to Client 10 settings	constant (20000)	exponential (2)
20 stations scenarios: Client 1 to Client 20 settings	constant (10000)	exponential (2)

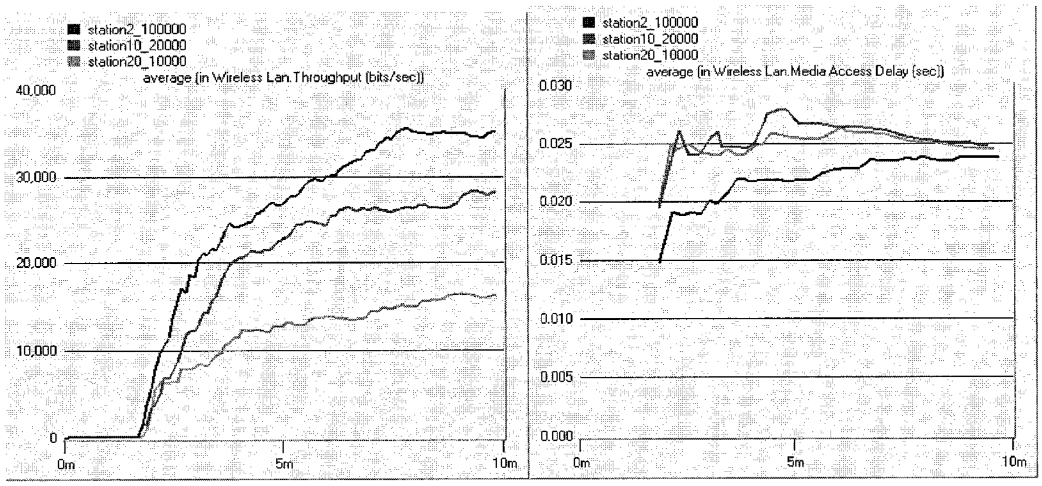


Figure 16: The throughput and the Medium Access Delay of Client 0

Figure 15 shows the global traffic load of three scenarios. 3 scenarios had the constant value that was expected. Figure 16 compares the throughput and the Medium Access Delay of Client 0 at 2 station, 10 station and 20 station scenarios. Note that x-axis is time in the unit of minute in Figure 15 , Figure 16, Figure 17, Figure 18, and Figure 19.

From Figure 16, the following conclusions can be derived:

- As the total station count increases, the throughputs of Client 0 decrease 20% to 60% even though the global traffic load is unchanged.
- The Medium Access Delay of Client 0 is almost same when the total active station count increases. The global traffic load maintains the same as well.

Impact of traffic Load

In the following simulation, 3 active stations (2 stations + Client 0) network topology and 11 active stations (10 stations + Client 0) network topology were examined. For each network topology, Client 0 traffic load was kept unchanged and the traffic load of the remaining clients was increased. The throughput and Medium Access Delay of Client 0 were compared in different traffic load scenarios in each of network topologies.

In order to make them comparable, the study generated 10000 loads, 50000 loads and 100000 loads for 3 active station scenario, and generated 2000 loads, 10000 loads, and 20000 loads for 11 active station scenario.

Figure 17 shows global traffic loads. The global traffic loads for each of the scenarios varied, as was expected.

Figure 18 compares the throughput of Client 0 with different traffic load in 2 station scenario (left graph) and different traffic load in 10 station scenario (right graph).

Figure 19 compares the MAD of Client 0 with different traffic load in 2 station scenario and with different traffic load in 10 active station scenario.

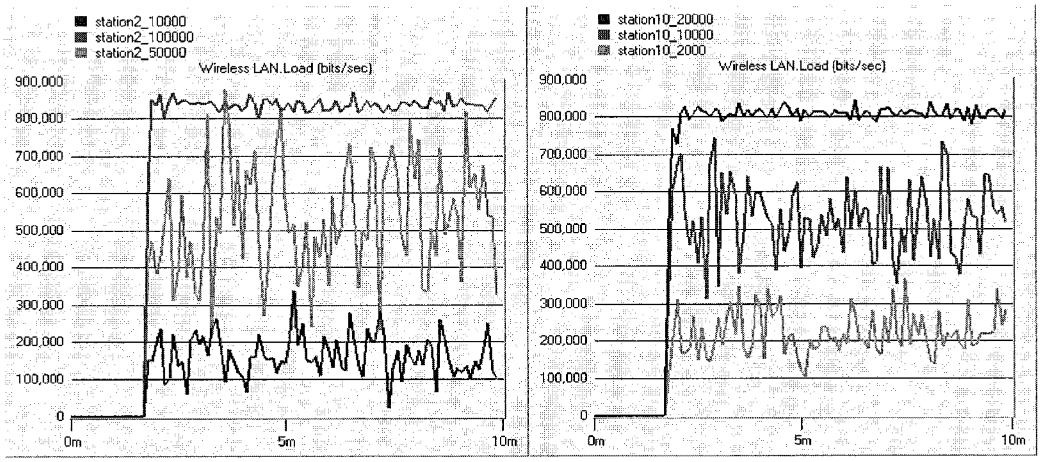


Figure 17: The global traffic load in 2 and 10 station scenarios

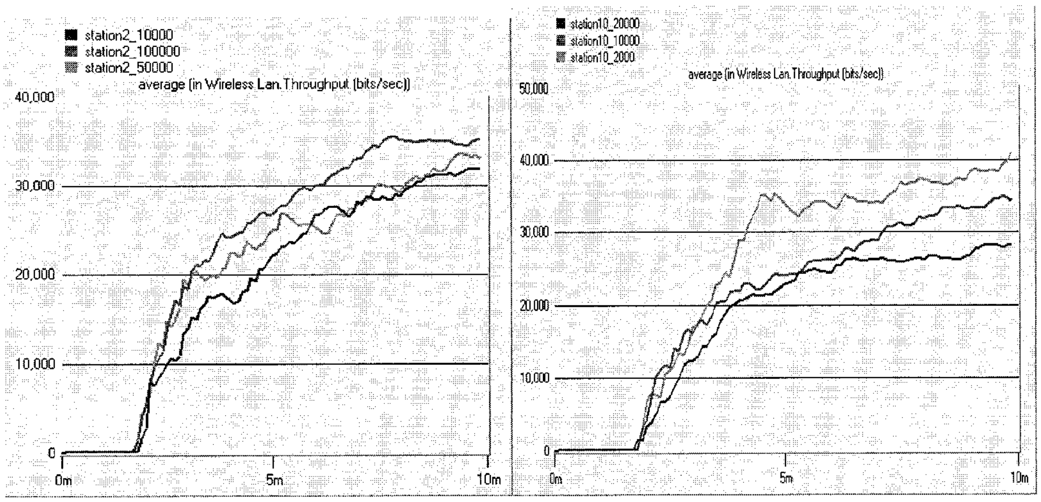


Figure 18: The throughput of Client 0 in 2 and 10 station scenarios

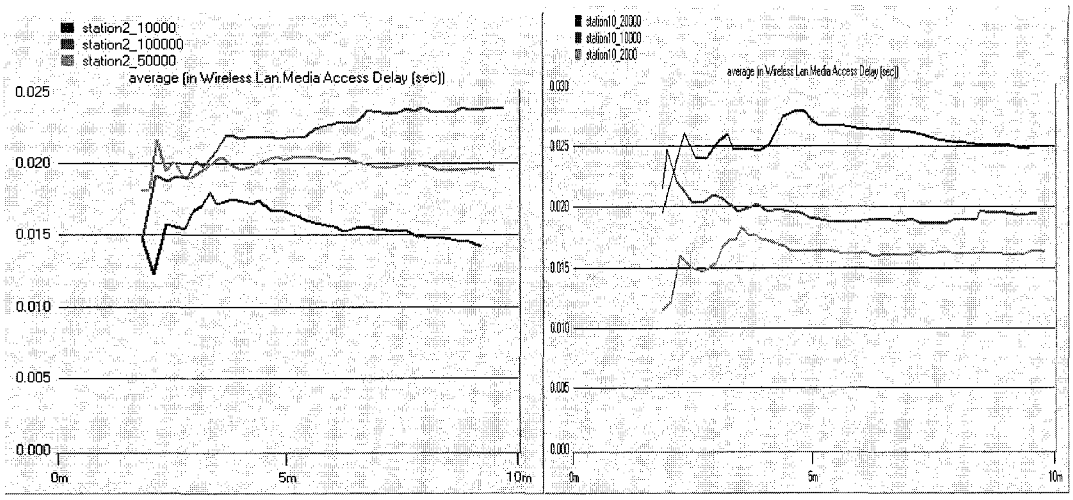


Figure 19: The Medium Access Delay of Client 0 in 2 and 10 station scenarios

Figure 18 illustrates that the throughput of Client 0 is not relevant to the change of traffic load if there are only 2 stations, but the throughput of Client 0 decrease significantly as the traffic load increases when there are 10 stations. As the traffic load increases, the Medium Access Delay of Client 0 increases as well, in both 2 active station and 10 active station scenarios.

Simulation Conclusion

The simulation results lead to the following conclusions:

- The channel load influence on the throughput becomes more significant as the number of active stations count increases. The channel load is a critical factor to the throughput of mobile station when the total number of active stations exceeds certain of a number (10 stations in the study).
- The channel load always has a significant influence on the Medium Access Delay.
- The total active station count has a significant influence on the throughput of the observed mobile station, but not always on Medium Access Delay.

2.6 SUMMARY

Wireless networking has a promising future with IEEE 802.11 leading the way as the standard for adoption in local networking environments. IEEE 802.11 MAC layer needs to address mobility, security, reliability while keeping compatibility with 802-type legacy networks. In today's 802.11 WLANs, devices use Distributed Coordination Function (DCF) to access wireless channel. The performance of this function heavily depends on channel condition, channel load and the number of mobile devices competing for the wireless channel. As the number of mobile devices connected to wireless networks increases, devices need to be aware of the environmental factors that are critical for managing performance. As the mobile device roams across BSSs to maintain the connectivity and high throughput due to the channel capacity and limited AP coverage, the transition latency of the mobile device is restricted by real-time application and good security. The dynamic nature and mobility characteristics of the wireless environment create challenges for mobile users wishing to get and maintain great wireless performance.

3 RADIO SENSING AND DISTRIBUTED COOPERATION

The dynamic nature of the wireless environment creates challenges for a distributed wireless network in maintaining great wireless performance and capacity. Substantial research work has begun to address this challenge by defining medium sensing and resource measurement [20] [21]. The medium sensing and resource measurement characterizes the dynamic wireless environment and improves the environmental adaptation for wireless devices. The environment data can be measured locally, as well as remotely. The industry has taken steps to address this challenge by establishing the IEEE 802.11 Task Group k (TGk). TGk is defining the extensions to existing 802.11 standards for Radio Resource Measurement (RRM) [13].

This chapter first introduces local medium sensing and radio measurement, and then discusses the radio resource measurement exchange between the mobile stations. Finally, based on the radio sensing and resource monitoring, a distributed cooperative wireless framework are proposed.

3.1 Medium Sensing and Radio Measurement

The environmental data can be used to adjust wireless device configuration parameters in order to improve network performance. It can also be analyzed by the wireless device in order to estimate signal quality, channel load and other conditions before a mobile client joins the wireless environment or while roaming from one environment to another. The environment data is measured in PHY and MAC. It can be used in PHY and MAC to adaptively optimize the MAC and PHY setting, as well as be transferred to upper-layer drivers for further analysis.

3.1.1 CHANNEL LOAD

As described in Chapter 2, the Distributed Coordination Function (DCF) is used as the mechanism to access the medium. It has been shown that the performance of this mechanism strongly depends on the channel busy fraction and the number of competing (active) users. The throughput per user decreases, resulting in lower performance, when the channel load increases. By sensing the channel load, STA can select a lightly loaded AP to improve throughput. The Channel load measurement can also help the network balance the traffic load to achieve high overall network throughput.

The Channel load can be expressed by the Channel Busy Fraction. The Clear Channel Assessment CCA Busy Fraction contains the fractional duration over which CCA indicates the channel is busy during the measurement duration. It is defined as Ceiling (255 × [Duration CCA indicated channel was busy]/[measurement duration]).

Figure 20 depicts how Clear Channel Time (*Tcc*) is computed. The STA measures the length of each CCA interval *Tcca*:

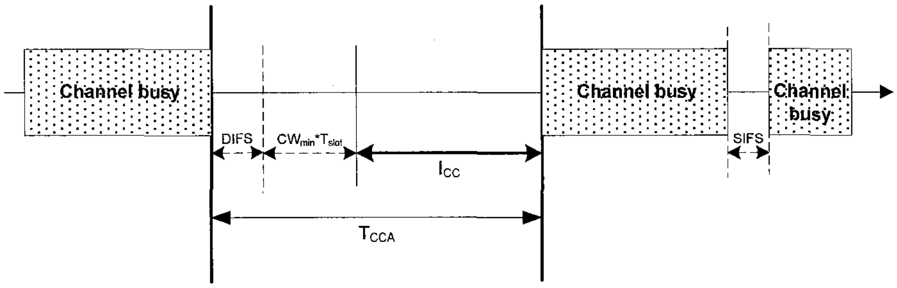


Figure 20: Clear channel time measurement

If (*Tcca* > *DIFS* + *CWmin* × *Tslot*) then there is a clear opportunity to transmit a packet, and we increase Scc accumulator:

$$Scc \leftarrow Scc + (Tcca - (DIFS + CWmin \times Tslot)).$$

Channel Busy fraction = 255 × (*Ttesting* – *Scc*) / *Ttesting*.

where,

$$Ttesting = dot11ChannelUtilizationBeaconIntervals \times dotBeaconPeriod \times 1024$$

To Sync with the Medium Time calculation, it may be necessary to set *Ttesting* = 1 second.

The algorithms showing how to use channel load to improve channel capacity and mobile station throughput will be discussed in Chapter 4 and Chapter 6.

3.1.2 Receiving Channel Power and Noise

.

The Signal to Noise and Interference Ratio (SNIR) is critical to the mobile station's throughput. Since a higher data rate modulation scheme requires a higher SNIR threshold, the different thresholds can form a set of rank for SNIR value. For example, in 802,11b, five ranks are defined:

Rank 0: $SNIR < 0$ dB , not supported data rate

Rank 1: $0dB < SNIR \leq 5dB$, for data rate of 1 Mbps

Rank 2: $5dB < SNIR \leq 7.5dB$, for data rate of 2 Mbps

Rank 3: $7.5dB < SNIR \leq 12.5dB$, for data rate of 5.5 Mbps

Rank 4: $12.5dB < SNIR$, for data rate of 11 Mbps

However, measuring SNIR is more complicated in an unlicensed than in a licensed band, because of the pattern of noise and interference. This section introduces two signal-to-noise measurements. They are the receiving channel power indication and the noise histogram (non-802.11 signal).

Received Channel Power Indicator (RCPI)

The RCPI indicator is a measure of the received RF power in the selected channel, measured at the antenna connector. This parameter is a measure by the PHY sublayer of the received RF power in the channel. The PHY measures the RCPI over the entire received frame and preamble. RCPI is a monotonically increasing, logarithmic function of the received power level, defined in dBm. The Received Channel Power Indicator (RCPI) parameter can be defined as an 8 bit value in the range from 0 through 220, with indicated values rounded to the nearest 0.5 dB. Accuracy for each measurement is +/- 5dB. The measurement assumes a receiver noise equivalent bandwidth of 22 MHz.

Noise Histogram

The Noise Histogram Measurement contains the Received Power Indicator (RPI) densities observed in the channel for the several RPI levels. To compute the RPI densities, the STA measures the RPI in the specified channel as a function of time over the measurement duration when NAV is equal to 0 (when virtual CS mechanism indicates idle channel) except during frame transmission and reception. The time resolution of the RPI measurements is in microseconds. The RPI densities are computed for each of the possible RPI values using ⌈(255 × ([Duration receiving at RPI value] / (1024 × [Measurement Duration] – [NAVBUSY]))⌉. NAVBUSY is the total time in microseconds that NAV is non-zero during the Measurement Duration. An algorithm on how to use receiving channel power and noisy histogram to optimize fragmentation size, transmit rate and RTS/CTS threshold is discussed in Chapter 5.

3.1.3 Transmit QoS metrics

Transmit QoS metrics are measured per destination and per traffic category identifier or per traffic streaming identifier. The Peer QoS STA Address contains the 6 byte MAC address from the Address 1 field of the measured Data frames. The Traffic Identifier indicates the TC or TS of the measured traffic. The Transmit QoS metrics can be characterized by several parameters:

- Transmitted MSDU Count,
- MSDU Failed Count,
- MSDU Discarded Count,
- MSDU Multiple Retry Count,
- QoS CFPolls Lost Count,

- Average Queue Delay,
- Average Transmit Delay, and
- Delay histogram

These parameters relate transmissions to the QoS STA given in the Peer QoS STA Address field.

The Transmitted MSDU Count field contains the number of MSDUs for the TC, or TS given by the Traffic Identifier successfully transmitted in the measurement duration.

The MSDU Discarded Count field contains the number of MSDUs for the TC, or TS given by the Traffic Identifier discarded due either to the number of transmit attempts exceeding *dot11ShortRetryLimit* or the number of MSDUs for the TC, or TS given by the Traffic Identifier that are successfully transmitted after more than one retransmission attempt.

The QoS CFPolls Lost Count field contains the number of QoS (+)CF-Poll frames transmitted where there was no response from the QSTA. QoS CFPolls Lost Count is returned only if the reporting QSTA is a QAP and the Traffic Identifier is for TS. If unused, QoS CFPolls Lost count is set to 0. .

Average Queue Delay is the average queuing delay of the frames (MSDUs) that are passed to the MAC during the measurement duration for the indicated Peer QSTA Address and the indicated Traffic Identifier. The Queue Delay is measured from the time the MSDU is passed to the MAC until the point at which the first, or only, fragment is ready for transmission. The Queue Delay is expressed in TUs.

Average Transmit Delay is the average delay of the frames (MSDUs) that are successfully transmitted during the measurement duration for the indicated Peer QoS STA Address and the indicated Traffic Identifier. Delay is measured from the time the MSDU is passed to the MAC until the point at which the entire MSDU has been successfully transmitted, including receipt of the final ACK from the peer QSTA if the QoSAck service class is being used. Average Transmit delay is expressed in TUs.

Bin 0 Range indicates the delay range of the first bin (Bin 0) of the Transmit Delay Histogram, expressed in TUs. It is also used to calculate the delay ranges of the other 5 bins making up the histogram. The delay range for each bin increases in a binary exponential fashion as follows:

B_0 duration:

$0 \leq \text{Delay} < B_0$, for i =0;

Bi duration:

$2^{i-1} * B_0 \leq \text{Delay} < 2^i * B_0$ for $1 < i < 5$

Note: For instance, if Bin 0 Range is 10ms, the bin durations should be defined in Table 2.

Table 2: Delay Definitions for a Transmit QoS Report

Bin	Measured Delay (TUs)
0	Delay <10
1	10 ≤ Delay < 20
2	20 ≤ Delay <40
3	40 ≤ Delay <80
4	80 ≤ Delay <160
5	Delay ≥ 160

To compute the value reported in Bin *i*, B*i*, $0 \leq i < 5$, of the Transmit Delay Histogram, the STA initializes all bin counts to zero. For each MSDU successfully transmitted, the measured Transmit Delay determines which bin count is incremented. If the measured delay has a duration time *t* within Bin *i*, then the frame count in Bin *i* is increased by one, up to a ceiling value of 2^32 – 1. Transmit Delay is measured from the time the MSDU is passed to the MAC until the point at which the entire MSDU has been successfully transmitted, including receipt of the final ACK from the peer QSTA if the QoSAck service class is being used. During the Transmit QoS Metrics Measurement, a histogram is generated that represents the distribution of Transmit Delay.

3.2 Measurement Coordination

An emerging IEEE standard, 802.11k [13], aims to provide client radio sensing to wireless-LAN Access Points and wireless network infrastructure information to from Access Point to client. The 802.11k draft standard defines a set of measurement requests and reports that detail MAC and PHY 2 client statistics. In most cases, access points or WLAN switches ask clients to report data, but in some cases client might request data from Access Points.

Draft standard IEEE 802.11k defines a number of measurement messages:

- the Radio Measurement Request,
- the Radio Measurement Report,
- the neighbor Report,
- the RCPI Report, and

- the AP Channel Report.

The Radio Measurements include the Beacon measurement, Frame measurement, Channel Load measurement, Noise Histogram measurement, Location Configuration Information, and STA Statistics measurement. 802.11k permits a station to send a Radio Measurement Report frame in response to Radio Measurement Request. The Radio Measurement Request, Radio Measurement Report, Neighbor Report Request and Neighbor Report Response are carried by Action Frames, the 802.11 extensible Management Frame type. The RCPI Report is carried by a Probe Response, and the AP Channel Report is carried by a Beacon or a Probe Response. Beacons and Probe Responses are also 802.11 Management Frames.

Currently, Access Points and clients cannot share channel information. With 802.11k, an access point could have a client build a "noise histogram," which will display all non-802.11 energy on that channel. An Access Point also can request data about channel load or how long the channel was used during a given time. An access point or WLAN switch can use this information to determine whether there is too much interference or traffic on a channel to use it for WLAN services.

The following section explains how to exchange the radio resource measurement between AP and STA, and then discusses the AP service load and neighbor AP report that are wireless network information provided by AP.

3.2.1 Radio Resource Measurement Exchange

As described in IEEE 802.11k, an AP or a STA can request a STA or an AP to report a measurement. If an AP or a STA receives a Measurement Request message, STA or AP should respond with a measurement report once it successfully conducts the measurements. The measurement request may be sent to a unicast, multicast or broadcast destination address. The source and destination of a measurement request must both be members of the same BSS or members of the same IBSS. Unicast measurement requests take precedence over multicast requests, which take precedence over broadcast requests.

An AP or a STA that issues a measurement request to a STA or an AP to perform a measurement on the serving channel may transmit data frames to that STA during the measurement itself. A STA or an AP that issues a measurement request to a STA to perform a measurement on a non-serving channel is not required to take any special action to suspend traffic to that STA. All stations maintain data services and an association or membership with the BSS or IBSS, respectively, on the serving channel while performing measurements on non-serving channels.

Since measurements on non-serving channels could potentially degrade a station's performance, non-serving channel measurements should be requested sparingly and for short durations. Since measurements on the serving channel execute concurrently with normal traffic processing, measurements of the serving channel may be requested more frequently and for longer durations. If desired, the requesting STA may issue periodic concurrent measurement requests to achieve near-continuous reporting.

3.2.2 Wireless Network Information from AP

The AP service load and Neighbor AP Report are the wireless network information sent from AP to STA. An algorithm using AP service load and Neighbor AP report for AP selection and load balancing will be discussed in Chapter 6.

AP Service Load

The AP Service Load contains information on the current station population and traffic levels in the BSS, which includes the station count, channel utilization and Available Admission Capacity.

The station count indicates the total number of STAs currently associated with this BSS.

The channel utilization field is defined as the percentage of time the QAP senses the medium as being busy, as indicated by either the physical or virtual carrier sense mechanism. This percentage is represented as a moving average of ((channel busy time/(*dot11ChannelUtilizationBeaconIntervals* × *dot11BeaconPeriod* × 1024)) × 255), where 'channel busy time' is defined as the number of microseconds during which the carrier sense mechanism has indicated channel busy, and the MIB attribute *dot11ChannelUtilizationBeaconIntervals* represents the number of consecutive beacon intervals during which the average should be calculated.

The Available Admission Capacity field specifies the remaining amount of medium time available via explicit admission control, in units of 32 microsecond periods per 1 second. The field is helpful for roaming non-AP QSTAs to select a QAP that is likely to accept future admission control requests.

Neighbor AP Report

The Neighbor Report enables the STA to optimize aspects of neighbor BSS transition and ESS operation. A Neighbor Report element contains information on APs which the STA may use as candidates for a BSS transition. A Neighbor Report element contains only entries of neighboring APs that are legitimate members of ESSs identified in the Neighbor Report Request.

A Neighbor Report is sent by an AP, and it contains information about known neighbor APs. A Neighbor Report may not be exhaustive, either by choice, or due to the fact that some neighbor APs might not be known to the AP. The mechanism by which the contents of this table are determined is outside the scope of this amendment, but it may include information from measurement reports received from the STA's within the BSS, information obtained via the management interface, or the DS.

The 802.11k standard defines a Beacon request, in which an access point asks a client to go to a specific channel and report all the access point beacons it hears. The access point collects the data, and it or a WLAN switch will analyze the beacon information, looking at details such as what services and encryption types each access point supports and how strongly the client has heard the access point. The switch or access point uses this information to generate an ordered list of access points, called the Neighbor report.

The Neighbor Report includes the BSSID, Channel number, Channel Band, and PHY type of each AP identified. Since the information in the Neighbor Report may be stale, this information should be considered advisory; information obtained by the report recipient through a scan or other sources may also be considered, possibly overriding information in the Neighbor Report. For example, where information contained within a Neighbor Report is contradicted by information in the Beacon/Probe Response, the Beacon/Probe Response information may be considered.

3.3 Distributed Cooperation

On-going research into the complexity of dynamic, noisy and interfering environments has resulted in distributed, intelligently algorithms. These algorithms give wireless devices the ability to optimize media access control (MAC) configurations and operations, in response to the changes in the environment around a wireless device. Thus, a device's configuration and operations are adjusted to select the "best" access point, to address interference issues, to avoid collision, to optimize WLAN performance, and to improve the mobile user experience. Since the environment around a wireless device is hard to predict, it is almost impossible to define one set of parameters ahead of time that will guarantee optimal performance for all applications. The distributed cooperative mechanism presented in this thesis resolves this problem by allowing a device to make its own adjustments as well as to cooperate with other devices as the wireless environment changes, for example, as interference suddenly drops, or as an access point becomes overloaded.

To achieve this goal, a cooperative MAC architecture is proposed that allows a mobile device to analyze the characteristics of the wireless environment and enable wireless devices to adapt intelligently to a dynamic environment. Figure 21 depicted this distributed cooperative wireless architecture.

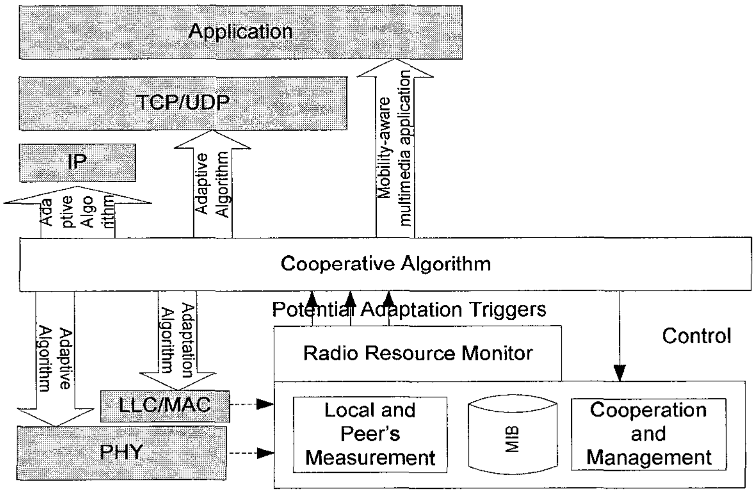


Figure 21: A cooperative MAC architecture

The cooperative MAC architecture includes Radio Resource Measurement, Radio Resource Monitor, Potential Adaptation Triggers, Cooperative Algorithms, and Radio Resource Cooperation.

The function of Medium Sensing and Radio Measurement is to collect radio resource measurement data and wireless network information from PHY and MAC layer, exchange radio resource measurement with its peers, and store the measurement and network information in the MIB. The function of the Radio Resource Monitor is to monitor the measurement data and provide potential adaptation triggers for the cooperative adaptation algorithm.

A potential adaptation Trigger is a set of threshold measurements. It is used to activate one or more cooperative algorithms. The cooperative algorithms are the core functions of this Cooperative MAC framework. Because of the functionalities of MAC layer and its position, the cooperative algorithms in the Cooperative MAC Architecture can be used to optimize multiple protocol layers. For PHY layer, cooperative algorithm can include Adaptive Modulation and Adaptive Transmit Power Control. For MAC layer, Smarts includes Smart Access Point Selection to avoid overloaded APs, Adaptive Fragmentation Size, Adaptive Data Rate, Adaptive RTS/CTS threshold, Adaptive collision avoidance, cooperative power management, cooperative channel allocation and spectral reuse. For application layer, Smarts can influence application for mobility–aware multimedia application.

In order for wireless devices to optimally utilize radio resources, wireless devices need to cooperate with algorithms. To achieve this cooperation, the wireless device use radio source cooperation protocols to exchange cooperative algorithm results and advise other devices to take the recommendation. A distributed cooperative framework is depicted in Figure 22.

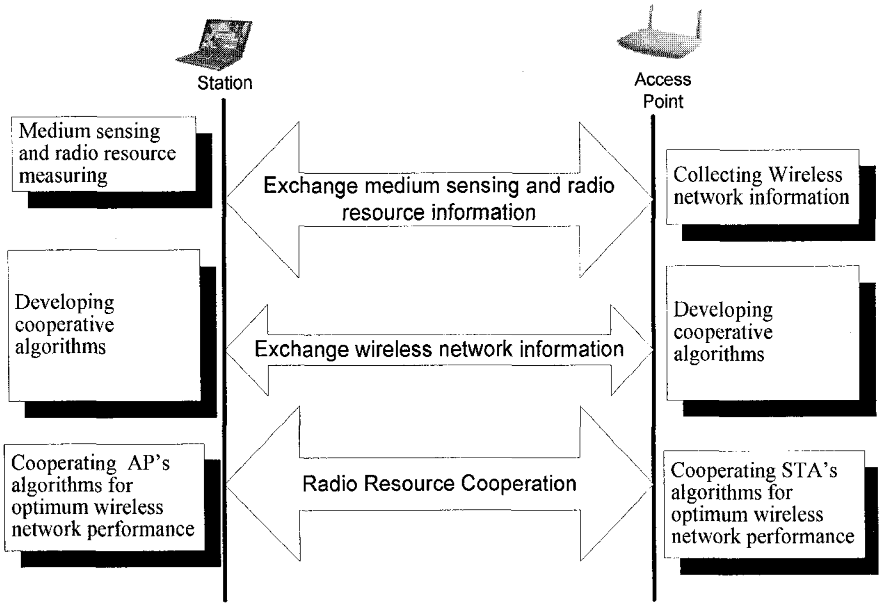

Station
Access Point
Medium sensing and radio resource measuring
Exchange medium sensing and radio resource information
Collecting Wireless network information
Developing cooperative algorithms
Exchange wireless network information
Developing cooperative algorithms
Cooperating AP's algorithms for optimum wireless network performance
Radio Resource Cooperation
Cooperating STA's algorithms for optimum wireless network performance

Figure 22: A distributed cooperative framework

3.4 SUMMARY

This chapter first described channel load condition, receiving channel condition and Transmit QoS metrics. The receiving channel condition includes receiving channel power indication and noise histogram. Then, the IEEE 802.11k was introduced for the radio resource measurement and network information exchange between the mobile stations or mobile station and AP. Finally, a cooperative MAC architecture was proposed to optimize the wireless network performance based on the medium sensing and radio measurement. The cooperative MAC architecture and distributed cooperative framework provide the capability for mobile clients and access point to adapt multiple optimizations simultaneously and cooperatively.

In the following three chapters, the Adaptive Contention Windows Size, the Cooperative Load Balancing and the Cooperative BSS Transition will be proposed and studied, as three examples of distributed cooperative algorithms. A mobile device might conduct these cooperative algorithms sequentially or individually. The Adaptive Contention Windows algorithm is designed to optimize the throughput within BSS. If the mobile device cannot achieve the desired throughput and meet quality of service requirement, the mobile may decide to seek a better AP by using the Cooperative Load Balancing algorithm, After selecting the target AP, the cooperative BSS transition is conducted for the mobile device to securely fast transition to the target AP.

4 ADAPTIVE CONTENTION WINDOW SIZE

As described in previous sections, the DCF method adopts a distributed, contention-based access scheme belonging to the class of CSMA/CA MAC protocols [22][23]. Collision occurs when two or more mobile stations try to simultaneously utilize the resource. The contention reduction, in accessing the shared channel, is obtained with a variable time-spreading of the mobile stations' accesses. Hence, in this system resource wastage is caused both from collision and from the resource idle periods introduced by the time-spreading of the accesses. As the reduction of the idle periods generally produces an increase in the number of collisions, to maximize the resource utilization the protocol should balance these two costs [24][25]. For these protocols, when the number of mobile stations grows, or even in the case of busty traffics arrivals due to high collision rate, low resource utilization is often obtained. This occurrence seems difficult to avoid, because the station's reaction to high contention conditions is based upon collision happenings. Since these costs change dynamically, depending on the network load, it is evident that there is a need for a kind of adaptation of such CSMA protocols with respect to congestion variations in the system [22][25].

This chapter derives an analytical performance model for channel utilization in the ideal channel conditions (i.e., no hidden terminals, and no channel noise). The analytical model shows that the maximum channel utilization that can be reached when an optimum contention window size is selected. By coordinating the medium sensing parameters introduced in Chapter 3, an Adaptive Contention Window Size (ACWS) algorithm is developed. This ACWS reduces the number of collisions as the number of users increases to maintain the maximum channel utilization. The validation of the adaptive contention window size algorithm against simulated performance is given at the end of the chapter.

4.1 Mathematic Model Derivation

The analytical performance evaluation model is based on the assumption of ideal channel conditions (i.e., no hidden terminals, and no channel noise). The channel is occupied by data payload, collision and idle period. The analysis assumes a fixed number of stations, each always having a packet available for transmission. In other words, the operation takes place in saturation conditions, i.e., the transmission queue of each station is assumed to be always nonempty.

The analysis is divided into three distinct parts. First, the collision probability for STA to transmit a packet is studied. Then, using packet transition events that can occur within a generic slot time, the channel utilization is expressed as a function of the computed value. Finally, the optimized contention window size for the maximum channel utilization is derived.

4.1.1 Collision Probability

Assume a fixed number n of contending stations. In saturation conditions, each station immediately has a packet available for transmission, after the completion of each successful transmission. Moreover, since all packets are "consecutive," each packet needs to wait for a random back-off time before transmitting.

A discrete and integer time scale is adopted: *t* and *t+1* correspond to the beginning of two consecutive time slots and the back-off time slot counter of each station decrements at the beginning of each slot time. Note that this discrete time scale does not directly relates to the system time. In fact, as illustrated in Figure 7 in the Chapter 2, the back-off time decrement is stopped when the channel is sensed busy, and thus the time interval between two consecutive time slot beginnings may be much longer than the slot time size, as it may include a packet transmission. Unless ambiguity occurs, the ensuing time slot will refer to either the (constant) value σ , and the (variable) time interval between two consecutive back-off time counter decrements.

The value of the back-off counter of each station depends also on its transmission history (e.g., how many retransmissions the head-of-line packet has suffered). We define convenience $W = CW_{\min}$. Let, m "maximum back-off stage," be the value such that $CW_{\max} = 2^m W$, and let us adopt the notation $W_i = 2^i W$, where $i \in (0, m)$ is called "back-off stage."

Since the back-off is uniformly distributed over 0, 1, ... W–1 for the first attempt, the station's back-off timer is *W/2* (slots), on average. Each transition has a collision probability *p*, and a station transmits a packet multiple times until it received an acknowledgement (thus indicating a successful transmission), so that we can model the number of transmissions per packets as geometrically distributed with a probability of success of *1-p.* Furthermore, each collision causes a dilation of the contention window until the maximum is reached, so that the average back-off slot $Q_{backoff}$ is:

$$Q_{backoff} = W(1-p)/2 + W(1-p)/2\times 2p + \ldots + W(1-p)/2\times(2p)^m$$

$$= W(1-p)/2 \times (1 + 2p + \ldots + (2p)^m)$$

$$= W(1-p)(1 - 2^{m+1}p^{m+1})/(2(1-2p)) \qquad (1)$$

Note that this derivation assumes that $0 \leq p < ½$, as this formula has a singularity at $p = ½$ and is negative for $p > ½$.

The probability τ that a station transmits in a randomly chosen slot time can now be expressed. Any transmission occurs when the back-off time counter is equal to zero, regardless of the back-off stage, average of τ is:

$$\tau = 1/Q_{backoff} = 2(1-2p)/W(1-p)(1-2^{m+1}p^{m+1}) \qquad (2)$$

However, in general, τ depends on the conditional collision probability p , which is still unknown. To find the value of p it is sufficient to note that the probability p that a transmitted packet encounters in a collision is the probability that, in a time slot, at least one of $n-1$ the remaining stations transmit. The fundamental independent assumption given above implies that each transmission "sees" the system in the same state, i.e., in the steady state.

At steady state, each remaining station transmits a packet with probability p. This yields:

$$p = 1-(1-\tau)^{n-1} \qquad (3)$$

Equations (2) and (3) represent a nonlinear system in the two unknowns τ and p , which can be solved using numerical techniques. It is easy to show that this system has a unique solution. In fact, inverting (3), we obtain $\tau^*(p) = 1-(1-p)^{1/(n-1)}$.

This is a continuous and monotone increasing function in the range $p \in (0,1)$, which starts from $\tau^*(0) = 0$ and grows up to $\tau^*(1) = 1$.

4.1.2 Channel Utilization

Let S be the channel utilization, defined as the fraction of time the channel is used to successfully transmit payload bits. To compute S, let us analyze what can happen in a randomly chosen slot time. Let P_{tr} be the probability that there is at least one transmission in the considered slot time. Since n stations contend on the channel, and each transmits with probability τ

$$P_{tr} = 1 - (1 - \tau)^n \qquad (4)$$

The probability P_s that a transmission occurring on the channel is successful is given by the probability that exactly one station transmits on the channel, conditioned on the fact that at least one station transmits, i.e.

$$P_s = n\tau(1 - \tau)^{n-1} \qquad (5)$$

We are now able to express S as the ratio:

$$S = \frac{E[T_s]}{E[T_{slot}]} \qquad (6)$$

$E[T_s]$ is the expected time the channel is sensed busy (i.e., the slot time lasts) because of a successful transmission, and $E[T_{slot}]$ is the expected length of a slot time. Here, T_s is the average time the channel is sensed busy (i.e., the slot time lasts) because of a successful transmission, and T_c is the average time the channel is sensed busy by each station during a collision. The expected channel time $E[T_s]$ of a successful transmit in a slot time is P_sT_s, since a successful transmission occurs with probability P_s. The expected length of a slot time $E[T_{slot}]$ is readily obtained considering that, with probability $1-P_{tr}$, the slot time is empty; with probability P_s it contains a successful transmission, and with probability $P_{tr}-P_s$ it contains a collision. Hence, (6) becomes

$$S = \frac{P_sT_s}{(1-P_{tr})\sigma + P_sT_s + (P_{tr}-P_s)T_c} \qquad (7)$$

Here, σ is the duration of an empty slot time. Of course, the values T_s, T_c and σ must be expressed with the same unit.

Note that the channel utilization expression (7) has been obtained without the need to specify the access mechanism employed. To specifically compute the channel utilization for a given DCF access mechanism, it is now necessary only to specify the corresponding values T_s and T_c.

Let us first consider a system completely managed via the basic access mechanism. Let $H = H_{phy} + H_{mac}$ be the packet header, and δ be the propagation delay. As shown in Figure 9 in Chapter 2, in the basic access case we obtain

$$\begin{aligned} T_s &= H + E[P] + SIFS + \delta + ACK + DIFS + \delta \\ T_c &= H + E[P^*] + DIFS + \delta \end{aligned} \qquad (8)$$

Where, $E[P]$ is the expected length of the longest packet payload that was successfully transmitted. $E[P^*]$ is the expected length of the longest packet payload that was involved in a collision. In this case all packets have the same fixed size, $E[P^*] = E[P] = P$.

4.1.3 Optimum Contention Window

The analytical model given above is very convenient for determining the maximum achievable channel utilization. Let us rearrange (7) to obtain

$$S = \frac{T_s}{T_s - T_c + \frac{(1-P_{tr})\sigma + P_{tr}T_c}{P_s}} \tag{9}$$

As T_s, T_c, and σ, are constants, the channel utilization S is maximized when the following quantity is maximized:

$$\frac{P_s}{(1-P_{tr}) + P_{tr}T_c/\sigma} = \frac{n\tau(1-\tau)^{n-1}}{T_c^* - (1-\tau)^n(T_c^* - 1)} \tag{10}$$

where $T_c^* = T_c/\sigma$ is the duration of a collision measured in slot time units σ. Taking the derivative of (10) with respect to τ, and imposing it equal to 0, we obtain, after some simplifications, the following equation:

$$(1-\tau)^n - T_c^*\{n\tau - [1-(1-\tau)^n]\} = 0 \tag{11}$$

Under the condition $\tau << 1$

$$(1-\tau)^n \approx 1 - n\tau + \frac{n(n-1)}{2}\tau^2 \tag{12}$$

holds, and yields the following approximate solution:

$$\tau = \frac{\sqrt{[n+2(n-1)(T_c^*-1)]/n} - 1}{(n-1)(T_c^*-1)} \approx \frac{1}{n\sqrt{T_c^*/2}} \tag{13}$$

Equation (11) and its approximate solution (13) are of fundamental theoretical importance. In fact, they allow for explicitly computing the optimal transmission probability τ that each station should adopt in order to achieve maximum channel utilization within a considered network scenario (i.e., number of stations). In other words, they show that (within a PHY and an access mechanism, which determine the constant value T_c^*) maximum performance can be, in principle, achieved for every network scenario, through a suitable sizing of the transmission probability τ in relation to the network size.

As n is not a directly controllable variable, the only way to achieve optimal performance is to employ adaptive techniques to tune the values m and W (and consequently τ) on the basis of the estimated value of n. This problem has been specifically considered in [9] in the case of fixed back-off window size (i.e., $m = 0$). In such a case,

$$Q_{backoff} = {W_{opt}}/{2}$$

τ is given by

$$\tau = {1}/{Q_{backoff}} = {2}/{W_{opt}}$$

Therefore the back-off window that maximizes the channel utilization is readily found as:

$$W_{opt} = n\sqrt{2T_c^*} \tag{14}$$

Moreover, the maximum channel utilization is practically independent of the number of stations in the wireless network. This is easily justified by noting that the channel utilization formula can be approximated as follows. Let $K = \sqrt{T_c^* / 2}$, and let us use the approximate Solution $\tau = 1/(nK)$. For n sufficiently large

$$P_{tr} = 1-(1-\tau)^n = 1-(1-\frac{1}{nK})^n \approx 1-e^{-1/K} \qquad (15)$$

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}} \approx \frac{n}{(nK-1)(e^{1/K}-1)} \approx \frac{1}{K(e^{1/K}-1)} \qquad (16)$$

The maximum achievable channel utilization $S_{\max}$ can thus be approximated as

$$S_{\max} = \frac{T_s}{T_s + \sigma K + T_c(K(e^{1/K}-1)-1)} \qquad (17)$$

4.2 ACWS Algorithm and Protocol

The analytical model analysis assumes a fixed number of stations, each always having a packet available for transmission. In real measurement, we assume n is the number of associated stations at Access Point (AP). AP can also obtain the average channel busy fraction b and collision rate P from STAs via the radio resource measurement defined in Chapter 3. Given that:

$$b = (T_c + T_s)/\sigma \qquad (18)$$

And, $P = T_c /(T_c + T_s)$ (19)

Combining (14), (18) and (19), the adaptive contention window size is given:

$$W_{adaptive} = n\sqrt{2bP} \qquad (20)$$

Once AP calculates the adaptive content windows size, the AP can use an Action frame, called Contention Windows Request frame, to notify all its associated STAs about the new content window size. The associated STAs should use this new content windows size as *CWmin*. The Contention Windows Request frame is implemented as an Action frame, and is broadcast by an AP to all of its associated STAs. The format of the Contention Windows Request frame is shown in Table 3.

Table 3: Content Window Request frame format

	Category	Action	Dialog Token	Contention Window Size
Octets:	1	1	1	2

The Category field is set equal to the value indicating the radio resource management. The Action field is set equal to the value indicating Contention Windows Request. The Dialog Token field is set equal to the value in the corresponding Contention Window Request.

The Contention Window Size is set equal to the adaptive contention window size that AP calculates based on equation (20).

4.3 Simulation Results

In order to validate the adaptive contention window size algorithm, a detailed 802.11 simulation was run with a variable number of stations (users) in the system. Of particular interest was the channel occupation, because of data payload, collision and idle period.

A network with 1 to 20 stations was studied in which every station could directly communicate with every other nodes (thereby eliminating the possibility of hidden terminals). A perfect channel was assumed. Each packet consisted of 1K bytes plus MAC and PHY layer headers. Each station generated packets for the MAC layer in accordance with a Poisson process. Every station used the same rate, therefore controlling the offered load through the common packet generation rate. It was also assumed that 802.11 used direct sequence spread spectrum (DSSS) at the physical layer and parameters according to the IEEE 802.11 specification.

Figure 23 and Figure 24 depict the simulation results without ACWS and with ACWS, which show the channel bandwidth occupation versus the total station number. The channel bandwidth occupation can be divided into successful data packets, collision packets, and idle time. Each successful data packet includes a data payload, MAC and PHY header, and associated control packets (ACK, et. al).

Figure 23 shows channel utilization vs. number of stations without ACWS algorithm. In a single station simulation scenario, there is no collision, so channel efficiency is 74% and idle time is 36%. As the number of station increased, collisions increased as well. The idle time decreased as the number of stations increased. The collision time increased from 7% to 30% as the station number increased from 2 stations to 20 stations. The total collision and idle time increased from 25% to 40%. The channel efficiency decreased from 75% to 60%. It is evident that, as the number increases, channel resource wastage is caused by collision. With less active stations or traffic load, channel resource wastage is caused by idle period.

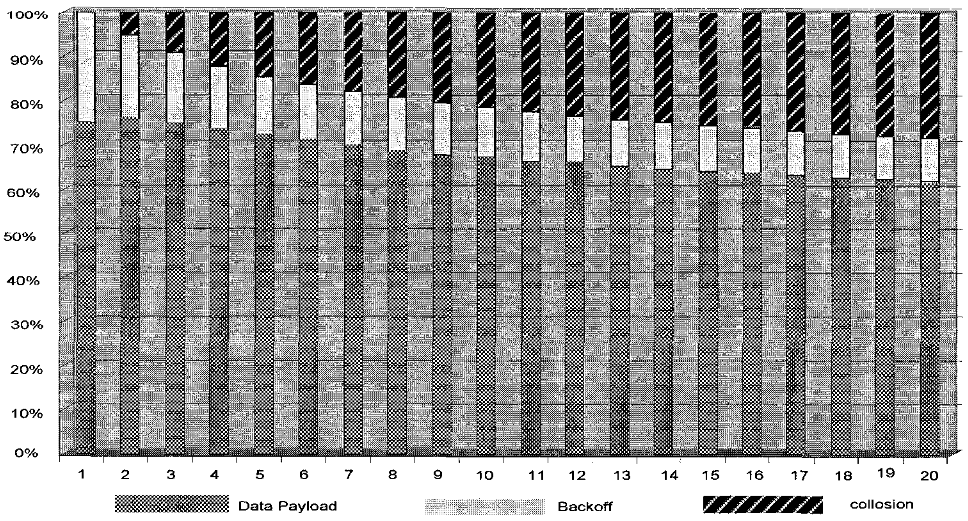


Figure 23: Channel utilization without ACWS algorithm

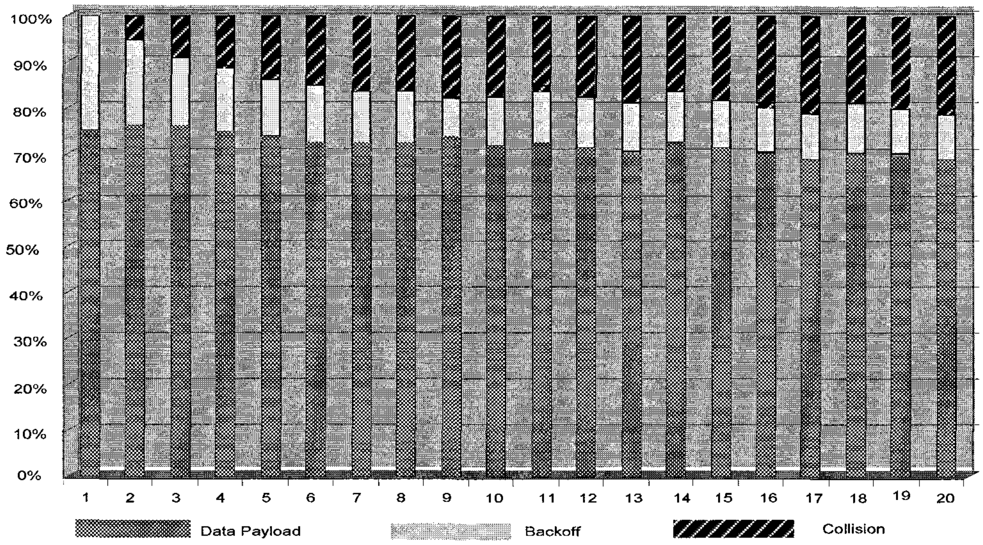


Figure 24: Channel utilization with ACWS algorithm

Figure 24 shows channel utilization vs. number of stations with ACWS algorithm. As shown in Figure 24, the total collision and idle time increased from 25% to 30% and the channel efficiency decreased from 75% to 70% when the number of stations increased from 2 to 20 stations. The channel utilizations as shown in Figure

24 were improved by 10% in 20 stations use case compared with Figure 23. As the station number increased from 2 to 20 stations, the total collision as shown in Figure 24 increased from 7% to 23% compared with 7% to 30% in Figure 23, which is much less compared with Figure 23. It is evident that, as the number increases, the collision is reduced by 7% using ACWS.

4.4 SUMMARY

The fraction of the channel bandwidth used by successfully transmitted messages gives a good indication of overheads required by the MAC protocol to perform its coordination task among stations. This fraction is known as the channel utilization, and the maximum value it can attain is known as the capacity of the MAC protocol. This Chapter derived a performance model and computed the theoretical limit of the IEEE 802.11 MAC protocol capacity in an ideal channel (no channel noise and no hidden nodes). It was shown that if a station has an exact knowledge of the wireless network status and radio environment, it is possible to adjust its configuration parameters to achieve a protocol capacity very close to its theoretical limit, and furthermore, to improve the wireless network performance. An ACWS algorithm was developed for AP dynamically controls *CWmin*, and the algorithm was used by a simulation to demonstrate a 10% channel utilization improvement.

5 COOPERATIVE LOAD BALANCING

In a wireless infrastructure network, each mobile station (STA) is associated with an AP within a Basic Service Set (BSS). Mobile stations periodically receive Beacon frames from multiple APs. Based on the signal strength that it receives, a mobile station can choose which AP to connect to. However, higher receiving signal strength doesn't guarantee that the mobile station will have a higher throughput in the serving BSS. The reason is that in deployed WLANs, the Distributed Coordination Function (DCF) is used as the mechanism to access the medium. It has been shown that the performance of this mechanism strongly depends on channel busy fraction and the number of competing (active) users. The throughput per user decreases, resulting in lower performance, when the number of users competing for the channel increases. Therefore, when a STA selects the AP based only on the receiving signal quality and discards a less loaded AP, it contributes to decreased utilization of the network.

This chapter proposes a cooperative load balancing algorithm, by using the wireless network information and radio source measurements discussed in Chapter 3, which allows the AP and STA to cooperatively balance load and maximize network performance. The remainder of this chapter is organized as follows. The next section gives an overview of load balancing benefits and general approaches. Section 2 introduces a STA and AP cooperative load balancing procedure and AP Selection algorithm to optimize the network performance. Section 3 presents simulation results.

5.1 Overview of Load Balancing

In Wireless LAN based on the IEEE 802.11 protocol, two different topologies can be configured in order to service different communication needs, as described in Chapter 2. These topologies are the Infrastructure mode and the Ad Hoc mode. In the Infrastructure Mode, basic network components are the mobile Stations (STA) and the Access Points (AP) conveying the bridging functions. The STAs are associated to an AP and consequently can communicate through the AP with other wired devices or STAs. The STA can be supported by more than one AP in the same region. The operation functions of the AP (Channel, ESSID, Security, etc) may differ among APs, although they support the same network. The STA usually has the ability to roam from one AP to another AP; when the STA is moving, the roaming function is necessary for best quality of service.

This chapter focuses on the infrastructure mode and deals with the problem of load balancing between AP. In Figure 26, sixteen STAs are associated to the left side AP, while none is associated to the right side AP. This asymmetry causes a high probability of packet loss in the left side AP and consequently an overall network degradation as the station number increases. This situation could be avoided by balancing the number of associated stations among APs. In Figure 26, a better load balance of associated stations to each AP, which can significantly increase the overall system performance and improve QoS effectiveness for VoIP and video streaming [26].

The applicability of a STA distribution algorithm to the AP, whereby a load balance results for each AP, strongly depends on the nature of the wireless LAN [27]. The problem becomes even more difficult due to dynamic network topology changes that cannot be avoided while STAs are moving.

Currently two load balancing approaches are being used in WLAN deployment: client-based load balancing and network-based load balancing. In the client-based load balancing, the client knows its receiving signal strength, the degree of interference that it receives, and the required traffic load and traffic patterns. Thus, the client knows when to transition from the serving AP to another AP and which AP to transition to in order to maximize its performance. However, experience indicates that the client-based load balancing can cause a Ping-Pong effect, and does not reach optimum load balancing from a global network management perspective.

In infrastructure-based load balancing, the network has a global view of channel and BSS load, and can thus avoid global and local non-optimality. As APs are always in service, managing and diagnosing load balancing parameters or inputs can be far easier. Infrastructure-based load balancing can enforce fairness or fight self-serving behavior. However, there is a heavy overhead in infrastructure-based load balancing from the network implementation perspective, and it often requires assumptions about client's scanning and selecting behavior.

The following proposes a STA-AP cooperative load balancing procedure via STA environment learning, as a solution to maximize network capacity and improve overall network performance.

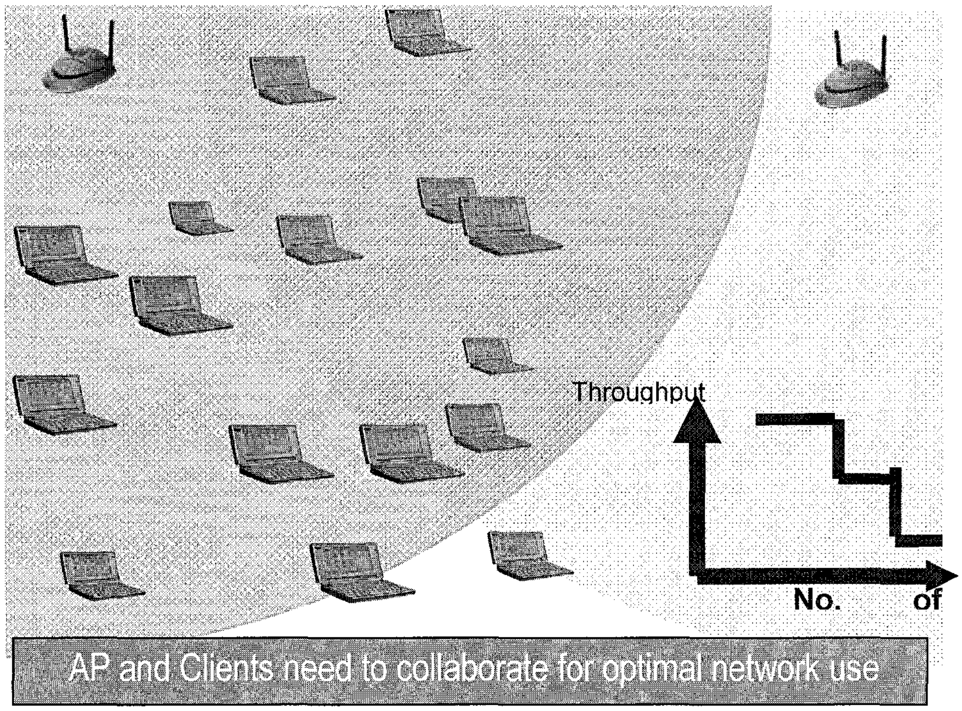


Figure 25: Unbalanced wireless networks

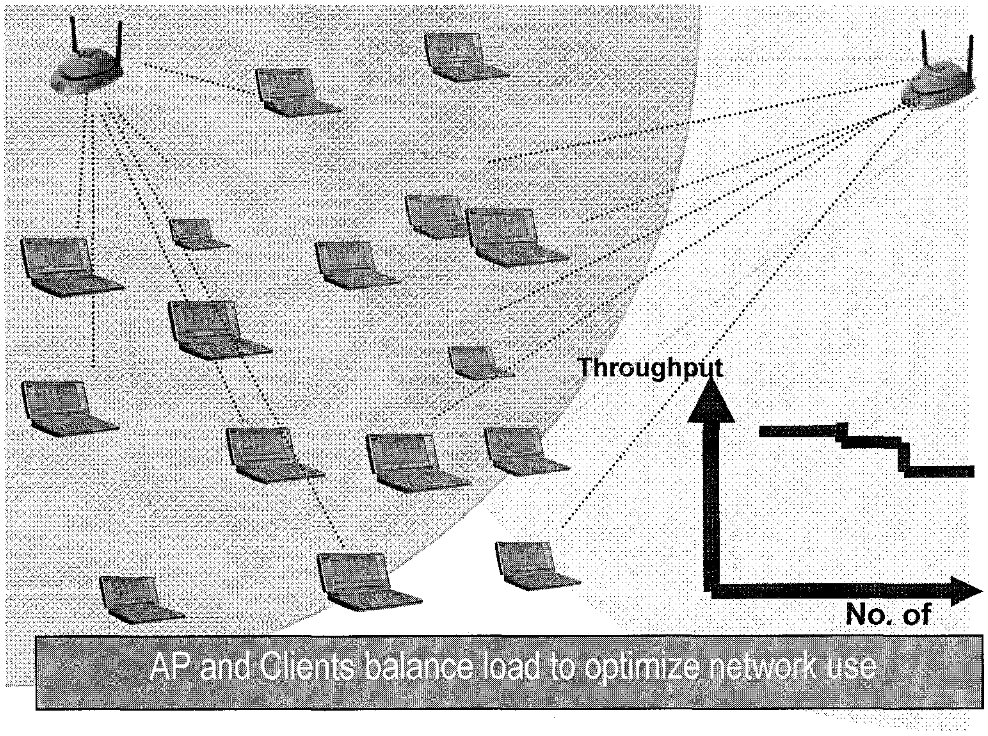


Figure 26: AP Load Balancing

5.2 Cooperative Load Balancing

AP and STA cooperative load balancing combines the two levels of collaboration. AP uses load balancing to suggest lightly loaded APs to one or more associated STAs. A mobile station uses the information from a received load balancing recommendation message to perform selective channel scanning. When using selective channel scanning, the station focuses its Beacon intercept attempts and Probe Requests on the channels specified by the report. The solution provides the means by which the AP service load can be balanced between a congested AP and an under-utilized AP to optimize the use of radio resource and to increase network aggregate throughput and QoS effectiveness. Using the channel information during the scanning, STA will select the "best" target AP from the recommended AP list, and initiate attempts to gain service from the selected AP.

The AP-STA cooperative load balancing includes four major steps as depicted in Figure 27.

Step 1: Network management sends the "lightly loaded" AP list to the overloaded AP

Based on the measurement information that the APs collect, the infrastructure network should conduct load balancing algorithm and send a list of lightly loaded adjacent APs to overloaded APs

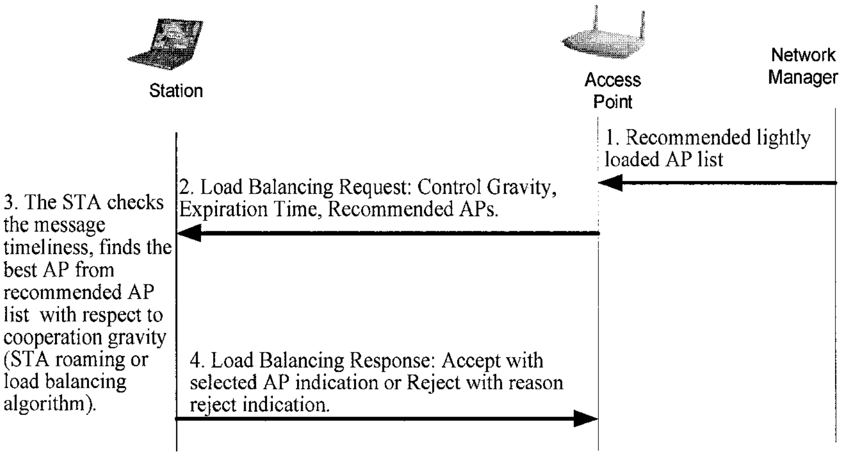

Station
Access Point
Network Manager
1. Recommended lightly loaded AP list
2. Load Balancing Request: Control Gravity, Expiration Time, Recommended APs.
3. The STA checks the message timeliness, finds the best AP from recommended AP list with respect to cooperation gravity (STA roaming or load balancing algorithm).
4. Load Balancing Response: Accept with selected AP indication or Reject with reason reject indication.

Figure 27: Procedure of AP-STA cooperative load balancing

Step 2: The AP sends a list of recommended (e.g. lightly loaded) APs to the STA

When the overloaded AP received the lightly loaded AP list, the overloaded AP sends STA Load Balancing Request message, which includes:

- A list of one or more nearby lightly loaded APs (candidate APs) as recommended. The list should include fields BSSID, Channel, PHY type and BSS Transition information.
- Timeliness indication, which indicates the expiration time of this cooperation request.
- Cooperation gravity: Advisory/Strong-Advisory/Command. "Advisory" indicates that cooperation message is a "hint", and it is up to the STA to make a distributed decision. "Strong-Advisory" indicates that the cooperation message is a strong recommendation and STA should follow. "Command" indicates that the dissociation or interrupted service will be enforced soon.

Step 3: The STA conducts AP Selection

When the STA receives Load Balancing Request message, the STA should:

- Check the cooperation expiration time provided in the Load Balancing Request message. If the time is expired, the STA will send a Load Balancing Response message to its current AP with the time expired rejection indication. Otherwise, the STA continues AP selection as described below.
- The STA receiving the Load Balancing Request message uses this information to perform selective channel scanning. When using selective channel scanning, the station focuses its Beacon intercept attempts and Probe Requests on the channels specified by the report. Using the AP service load information during the scanning, STA can identify the AP or APs from the recommended AP list that will deliver the best service, using the classical AP selection algorithm.
- Attempt to gain service from the selected target AP.

Step 4: The STA sends a confirmation to the currently associated AP

Before the STA switches to the selected target AP, the STA should send Load Balancing Response message to the current AP, indicating to which target AP the STA will switch. If the STA decides to stay with the currently serving AP, the STA should also send Load Balancing Response message with rejection indication.

The details of AP-STA cooperative load balancing are given in the following subsections.

5.2.1 Lightly Loaded Neighbor AP List

As described in chapter 3, an AP can send channel load request to a STA and ask the STA to report the channel load from the STA's perspective. After collecting all of channel load information, AP will report its channel load to the network management entity. After collecting all of the channel load information and available admission capacity, the network will create a matrix of channel load with respect to neighbor APs.

Suppose L0 is the channel load of AP_0, and L1, L2, … Ln are the channel load of AP_0's neighbor AP AP_1, AP_2, … AP_n, respectively.

If L0 ≥ x%, the AP_0 is overloaded. The network should create a lightly loaded neighbor AP list for overloaded AP_0. x% is set to 80% for most applications. The lightly loaded AP list must meet the following requirements:

- Each AP is a neighbor of AP_0.
- Each AP belongs to the same administrative domain.
- The channel load is less than (x-y)%. Note that (x-15)% is set for most applications.

5.2.2 Load Balancing Request

The Load Balancing Request frame is implemented as an Action frame, and is transmitted by an AP to one of associated STAs. The format of the Load Balancing Request frame body is shown in Table 4:

Table 4: Load Balancing Request frame format

	Category	Action	Dialog	Expiration	Cooperation	AP

			Token	Time	Gravity	Elements
Octets	1	1	1	4	1	variable

The Category field is set equal to the value indicating radio resource cooperation. The Action field is set equal to the value indicating Load Balancing Request. The Dialog Token field is set equal to the value in the corresponding Load Balancing response. The Expiration time field should be used for timeliness check at the receiving STA.

The Cooperation Gravity field includes Advisory, Strong-Advisory, and Command. “Advisory” indicates that cooperation message is a “hint”, and it is up to the STA to make a distributed decision. “Strong-Advisory” indicates that the cooperation message is a strong recommendation and STA should follow. “Command” indicates that STA need to move away from the current AP; otherwise, dissociation or interrupted service will be enforced.

The AP Elements field contains the recommended lightly loaded neighbor AP list. The format of the AP element is shown in Table 5.

Table 5: AP element format

	Element ID	Length	Neighbor List
Octets:	1	1	Variable

The Element ID field is equal to the Load Balancing Request value.

The value of Length field is dependent on the number of Neighbor List Entries representing the neighboring APs being reported. Each entry describes an AP and consists of BSSID, BSSID Information, Channel Number, Regulatory Class, and PHY Options, as shown in Table 6.

Table 6: AP list entry format

	BSSID	Channel Number	Regulatory Class	PHY Options
Octets:	6	1	1	1

The BSSID is the BSSID of the BSS being reported. The subsequent fields in the AP List Entry pertain to this BSS. The Channel Number vindicates the current operating channel of the AP represented by the BSSID in this AP list entry. The Regulatory Class field contains an enumerated value specifying the frequency band in which the Current Channel is valid. The PHY Options field contains the Condensed PHY type.

5.2.3 AP Selection Algorithm

Based on the lightly loaded AP list from a Load Balancing Request message, a mobile station analyzes the BSSID match status (ESS match bit, Capability match bit, and Support data rate match bit) of the recommended APs and decides on preferred APs. Since its IP address configuration does not change within the ESS, the matched ESS is always highly preferred. If mobile IP is enabled and the policy manager allows distinct ESS transition, an ESS match is not required. The STA will select APs that have the same capabilities with the STA's and supported rates. In general, the preferred APs should match the STA's ESS, Capabilities, and Supported rate.

If only one preferred AP has been selected, the mobile station will scan that channel and re-associate with the AP. If more than one preferred APs have been recommended, the mobile station will scan these specific channels, collect channel information, estimate the expected performance provided by each AP, and select one of the APs providing the best service as the target AP. If STA's traffic category is data only, STA will estimate the throughput from the preferred APs, and select the best AP that has the highest throughput. If the STA is in or is going to start the VoIP or video stream, the STA should estimate the required QoS bandwidth and select the best AP that can offer adequate QoS requirement. The throughput and QoS bandwidth requirements estimation are given as follows.

Generate the Preferred AP List

As depicted in Figure 28, the Received Channel Power Indicator (RCPI) from the preferred AP must meet the minimal RCPI requirement. The preferred APs are ranked by the following rules: Estimate the throughput for each of candidate APs. The AP that provides the highest achievable throughput will be ranked as the most preferred AP.

If client is in the Constant Bit Rate Traffic (VoIP call), or jump starting to Constant Bit Rate Traffic (VoIP call), the preferred APs will be ranked by the following rules:

- The preferred APs are ranked from category A APs to category B APs. If AP is a QAP and can provide Available Admission Capacity, AP is ranked in category A; otherwise it belongs to category B. Category A APs have higher priority than category B APs. For the category A APs, the Available

Admission Capability of the preferred AP must stratify STA required medium time.

- In each category of APs, the AP that offers the highest throughput for STA will be ranked as the most preferred AP.

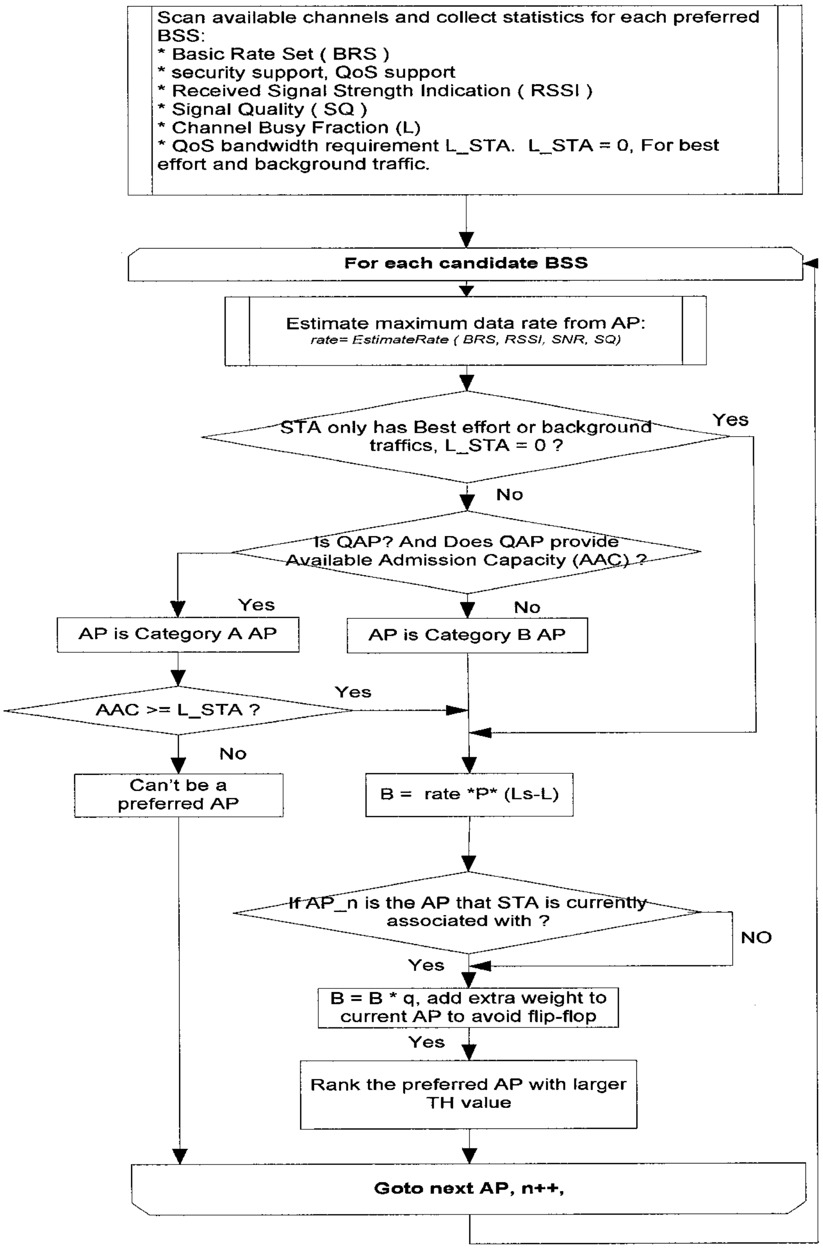

Scan available channels and collect statistics for each preferred BSS:
* Basic Rate Set (BRS)
* security support, QoS support
* Received Signal Strength Indication (RSSI)
* Signal Quality (SQ)
* Channel Busy Fraction (L)
* QoS bandwidth requirement L_STA. L_STA = 0, For best effort and background traffic.
For each candidate BSS
Estimate maximum data rate from AP:
rate= EstimateRate (BRS, RSSI, SNR, SQ)
STA only has Best effort or background traffics, L_STA = 0 ?
Yes
No
Is QAP? And Does QAP provide Available Admission Capacity (AAC) ?
Yes
No
AP is Category A AP
AP is Category B AP
AAC >= L_STA ?
Yes
No
Can't be a preferred AP
B = rate *P* (Ls-L)
If AP_n is the AP that STA is currently associated with ?
NO
Yes
B = B * q, add extra weight to current AP to avoid flip-flop
Yes
Rank the preferred AP with larger TH value
Goto next AP, n++,

Figure 28: Generate preferred AP list

Throughput Estimation

Assume there is only one mobile station in this BSS, with no interference and no collision. The estimated maximum date rate of the mobile station can be represented as a function of the Signal-to-Noise Ratio (SNR), RSSI, Signal Quality (SQ) and supported data rate set (DRS) or

$$rate = f(SNR, RSSI, SQ, DRS)$$

As for the 802.11 standard, a station with a packet to transmit monitors the channel activity until an idle period equal to or larger than the distributed inter-frame space (DIFS) is detected. After sensing an idle DIFS, the station waits for a random back-off interval before transmitting; the back-off time counter is decremented in units of slot time as long as the channel is sensed idle. The counter is frozen when a transmission is detected on the channel and reactivated when the channel is sensed idle again for more than a DIFS. The station transmits its packet when the value of the counter reaches zero. The BSS aggregated throughput is shared by the mobile stations within the BSS. As a result, the throughput per user decreases, resulting in lower performance when the number of users competing for the channel increases. Thus, instead of directly calculating the mobile station throughput, the following algorithm is proposed to estimate the throughput of the mobile station for best Access Point Selection.

The channel load is characterized by two states: Relative idle and Saturated. In the relative idle state, the throughput of the network keeps up with the offered load. The packets are delivered faithfully with nearly no losses and with only a few retransmissions. The throughput of a single mobile station is the same as its traffic load. In the saturated state, the network throughput will be reduced by a collision factor of P proportional to he number of competing users. The packet loss and the fraction of retransmissions increase with network traffic load. The throughput of the individual mobile station drops significantly by sharing the channel with competing users. The throughput at relative idle state is given by:

$$B = rate \times P \times (Ls - L)$$

The throughput at the saturated state is given by:

$$B = rate \times P \times L\_STA \times Ls / L$$

Where,

B is the throughput of the mobile station if it selects Access Point *n* as its AP.

L_STA = B_STA/rate is the mobile station's normalized traffic load

B_STA is the station's bandwidth requirement

P is the collision rate estimated

L is the channel busy fractions on AP *n*, which can be obtain from the current AP. L is between 0 and 1;

Ls is the saturated point of the channel busy fraction;

After estimating the throughput, the mobile station can compare the throughputs and select the Access Point that offers the best throughput. The best throughput has to be at least r% (for example, 15%) better than the existing throughput in order to avoid thrashing between APs.

STA QoS Bandwidth Requirement

Let *L_STA* be the mobile station's normalized traffic load such that the STA can derive from medium time from TSPEC. L_STA will be 0 for best effort and background traffic.

L_STA can be obtained as the following:

$L\_STA = Medium\ Time \times Td\ (ms)/100000\ (ms)$

Where,

$Td = Measurement\ Duration\ of\ Channel\ Busy\ Fraction;$

$Medium\ Time = Surplus\ Bandwidth\ Allowance \times pps \times MPDUExchangeTime;$

There are two requirements to consider: 1) the traffic requirements of the application, and 2) the expected error performance of the medium. The application requirements are captured by two TSPEC parameters: Nominal MSDU Size and Mean Data Rate. The medium requirements are captured by two TSPEC parameters: Surplus Bandwidth Allowance and Minimum PHY Rate. The following formula describes how Medium Time may be calculated (assuming RTS/CTS protection is not used):

$pps = \lceil (Mean\ Data\ Rate / 8) / Nominal\ MSDU\ Size \rceil;$

$MPDUExchangeTime = duration\ (Nominal\ MSDU\ Size,\ Minimum\ PHY\ Rate) + SIFS + ACK\ duration$

Duration () is the *PLME-TXTIME* primitive that returns the duration of a packet based on its payload size and the PHY data rate employed. Surplus Bandwidth Allowance can be set to 1 if it is not available.

5.2.4 Load Balancing Response

The Load Balancing Response frame is implemented by an Action frame, and is transmitted by a STA to its associated STAs. The format of the Load Balancing Response frame body is shown in Table 7.

Table 7: Load balancing response frame format

Field Name	Field Length
Category	1
Action	1
Dialog Token	1
Response Mode	1
Reject Reason	1
BSSID	6
Channel Number	1
Regulatory Class	1

The Category field is set to the value indicating radio resource cooperation. The Action field is set to the value indicating Load Balancing Request. The Dialog Token field is set to the value in the corresponding Load Balancing Request. Response mode is either Accepted (1) or Rejected (0). The Reject Reason indicates the reject reason if Response node is "Rejected".

BSSID, Channel number and Regulatory Class are the selected AP's BSSID, Channel number and Regulatory Class. These fields are only valid when Response node is "Accepted".

5.3 SIMULATION RESULT

This section presents results obtained by using OPNET for modeling Cooperative Load Balancing.

The network topology that demonstrates the Cooperative Load Balancing algorithm is depicted in Figure 29. The simulations were run twice. First results were obtained by generating traffic between the source STA (Station A19) and the destination STA (graph: loadbalanace_bap_12sta). Then, the Cooperative Load Balancing algorithm was enabled, which was implemented by modifying the OPNET 802.11b MAC state machine and gathering results for the same traffic pattern (graph: loadbalance_bap_12sta_rrmc).

The following statements are TRUE for both scenarios:

- Twelve STAs are associated with AP_0. They send traffic to DEST A which is associated with AP_2.
- STA_B17 is associated with AP_1 and sends traffic to DEST A.
- STA_A19 sends traffic to DEST A_0.
- The 12 STAs associated with AP_0 start with light traffic and add more traffic one second (in the simulation) later.

Differences between the two scenarios:

- Scenario loadbalance_bap_12sta: STA_A19 is a "regular" station. STA A19 is always associated with AP_0.
- Scenario loadbalance_bap_12sta_rrmc: STA_A19 is a STA with the Cooperative Load Balancing algorithm enabled. STA A19 is associated with AP_0 at the beginning of the simulation. As the simulation progresses, the load on access point AP_0 increases and A19 reselects a better access point

The following statistics were collected:

- The global throughput of the whole network.
- The throughput of station DEST A_0's in bits/sec (only station STA_A19 sends the traffic to station DEST A_0)

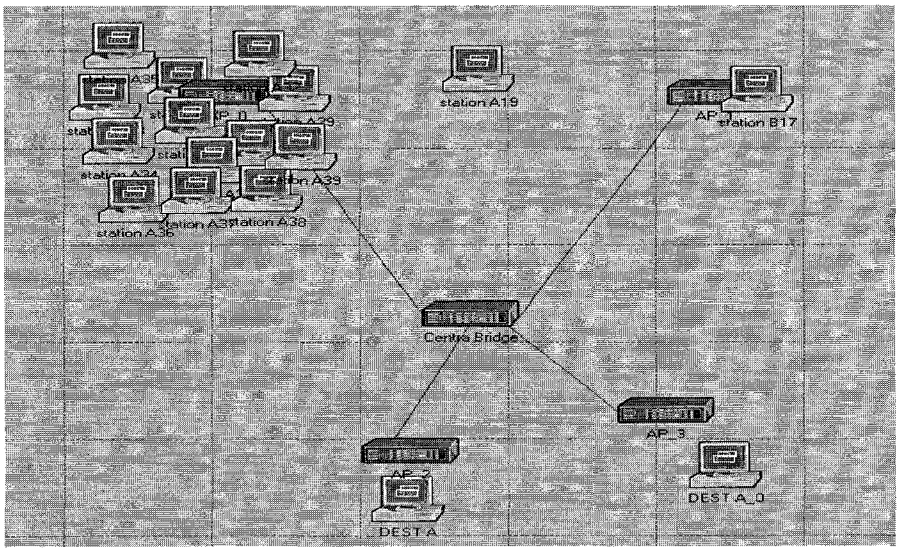


Figure 29: Network Topology for Load Balancing Simulation

Figure 30 and Figure 31 show the results from running the simulations. The following results were observed:

- A 5% improvement in the throughput of the overall network.
- About 40% improvements in DEST A_0's throughput.

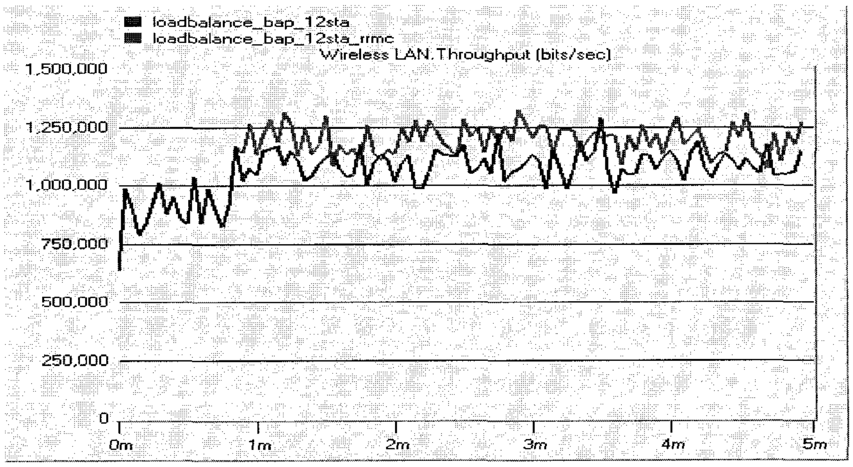


Figure 30: Global throughput comparison

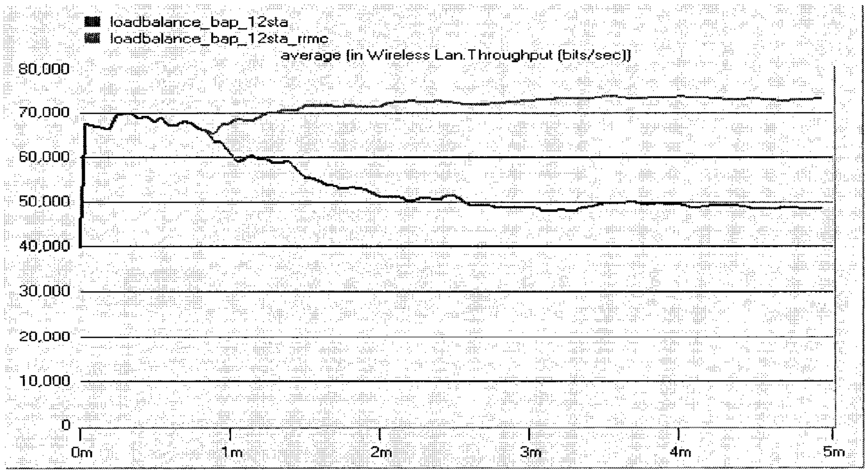


Figure 31: Comparison of the throughput of DEST A_0

5.4 Summary

This chapter explored a STA-AP cooperative load balancing scheme to optimize the network overall system performance and improve the QoS effectiveness of VoIP and video streaming. The AP and STA cooperative load balancing algorithm combines the two levels of collaborations. The AP uses load balancing to suggest lightly loaded APs to one or more associated STAs. The STA uses the information from a received load balancing recommendation message to perform selective channel scanning and radio sensing. Using the channel information gathered from the scanning and sensing, STA selects the "best" target AP from the recommended AP list, and initiates an attempt to gain the service from the selected AP. The simulation results demonstrated a 5% throughput improvement of the overall network and a 40% throughput improvement of the STA with the cooperative load balancing algorithm.

6 COOPERATIVE BSS TRANSITION

The cooperative BSS transition addresses latencies introduced by recent 802.11 amendments, notably the IEEE 802.11i security and 802.11e QoS enhancements while mobile station is transitioning from the serving AP to another. These services introduce several new latencies into AP-to-AP transition, including 802.1X authentication, 802.11i key management, and 802.11e bandwidth allocation. Such new latencies must be reduced to fall within VoIP's real-time constraints, in order to ensure that the voice application can maintain quality of service during device transitions within the WLAN [28].

To deal with these new constraints, the cooperative BSS transition optimization is designed to reduce transition latencies from the following perspectives:

- Reduce latency due to the 802.11e and 802.11i round trip times by piggybacking the messages for these functions over the Re-association exchange.

- Facilitate session key pre-computation, using a new key hierarchy, to ease the computational bottlenecks on resource-constrained devices, such as VoIP handsets.

- Eliminate the need for authentication on every AP-to-AP transition. This step typically removes two or more round trips induced by 802.1X authentication, and usually eliminates several computationally expensive cryptographic operations.

- Introduce mechanisms to afford target APs time to accommodate transitioning STAs, by using indications prior to Reassociation.

This chapter analyzes the optimizations [29][30] that the author and others have proposed to the IEEE 802.11 Working Group to improve transition latency. The BSS transition optimizations are based on the BSS transition optimization in IEEE 802.11r draft [14]. This chapter also identifies security flaws in the current design and proposes simple corrections. Finally, experiment results for transition optimization are explained that demonstrate a significant increase in transition efficiency.

6.1 KEY HIERARCHY

One of the most important components of the cooperative BSS transition optimization is its key hierarchy, as depicted by Figure 32. To understand the motivation for this design, it is useful to recall the 802.11i key hierarchy. The 802.11i key hierarchy has three layers:

- A master session key (MSK), shared between an Authentication Server (AS) and the mobile station,

- A Pairwise Master Key (PMK), which is an authorization token shared between the AP and the station, and finally

- A Pairwise Transient Key (PTK), which is a concatenation of session keys derived from the PMK by the AP and the mobile station.

As the final authentication step, the AS and station derive the MSK, and the AS transports this key to the AP in a (presumably secure) manner that is outside the scope of the standard. The 802.11i then creates the PMK from the MSK; the PMK can be cached across different associations between this AP and station. 802.11i derives a PTK from the PMK on each reassociation, guaranteeing a fresh session key that is cryptographically separated from the key used with any other session.

The design of the 802.11i key hierarchy was inspired by the classical “fat” AP model. In this model each AP is a stand-alone device, itself performing all access point functions. However, cost has made “thin” APs the norm in enterprise deployments, with functionality being split between “light-weight” APs and a central “controller.” In most of these designs, the controller performs the 802.11i PTK derivation for each AP that it controls. Since the controller maintains the PMK within a single cryptographic boundary, there should be no harm in using it to derive a different PTK for each “thin” AP under its control. Indeed, this design potentially could eliminate the performance problem of requiring a full authentication between the mobile station and the authentication server via each access point, which is especially onerous when the authentication server is remote. However, nothing in 802.11i allows the mobile station to distinguish this “thin” AP usage from a compromised key being shared among a set of compromised “fat” APs, so such an implementation would introduce a security flaw. All correct 802.11i implementations within the "thin" AP model still cache a different PMK for each distinct thin AP.

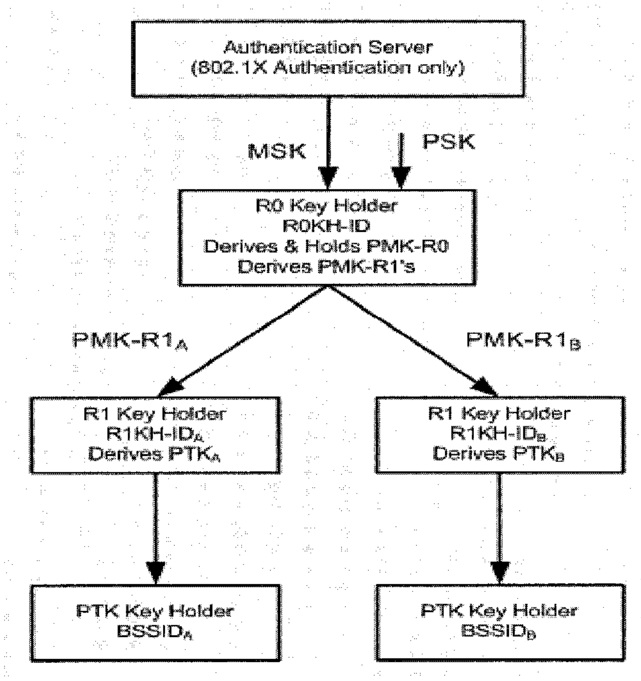


Figure 32: The new key hierarchy

The cooperative BSS transition optimization design attempts to accommodate the "thin" AP model by introducing a new layer of keys. Instead of eliminating the need for a per-AP PMK, it adds a new layer to eliminate frequent authentications between the station and authentication server. The BSS transition optimization splits the PMK into two layers, called the PMK-R0 and PMK-R1.

The BSS transition optimization defines the notion of a mobility domain to utilize the key hierarchy. The mobile station authenticates once with the mobility domain at initial contact. This authentication generates a MSK, which is used to derive a PMK-R0. The initial contact AP or its controller caches the PMK-R0. This party is called the R0 Key Holder. The R0 Key Holder never shares any PMK-R0 with any other party. The R0 Key Holder uses the PMK-R0 only to derive PMK-R1 keys, which it communicates to the R1 Key Holder for each access point within the mobility domain. The PMK-R0 and PMK-R1 lifetimes cannot exceed that of the MSK from which it was derived.

6.2 Handshake Optimizations

The BSS transition optimization optimized handshake, depicted by Figure 33, attempts to reduce latency by overlaying key management and QoS bandwidth allocation on top of the 802.11 reassociation process.

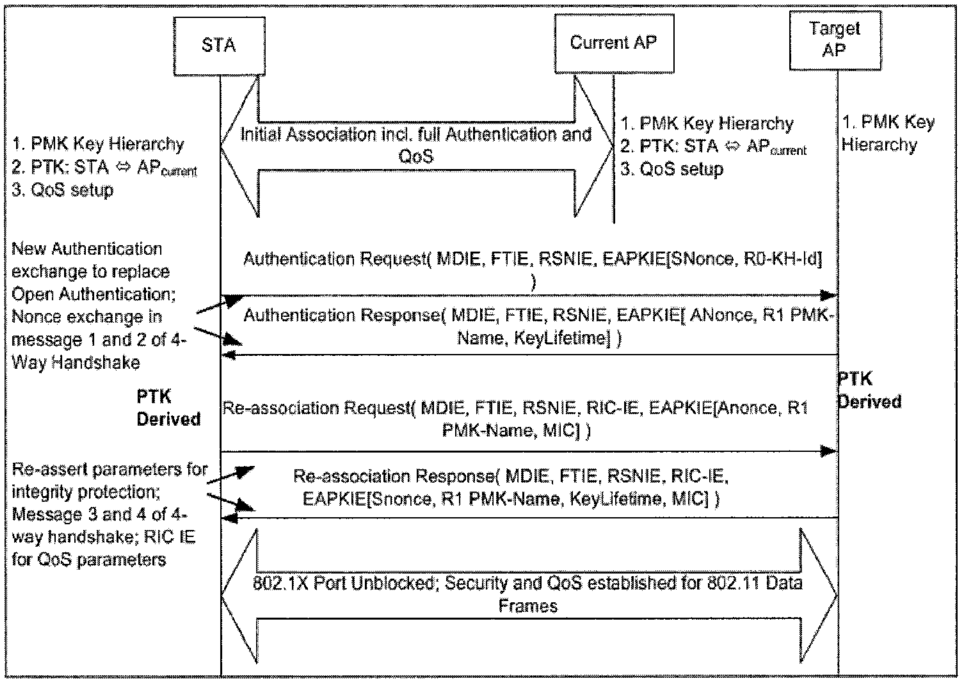

STA
Current AP
Target AP
1. PMK Key Hierarchy
2. PTK: STA ⇔ AP$_{current}$
3. QoS setup
Initial Association incl. full Authentication and QoS
1. PMK Key Hierarchy
2. PTK: STA ⇔ AP$_{current}$
3. QoS setup
1. PMK Key Hierarchy
New Authentication exchange to replace Open Authentication; Nonce exchange in message 1 and 2 of 4-Way Handshake
Authentication Request(MDIE, FTIE, RSNIE, EAPKIE[SNonce, R0-KH-Id])
Authentication Response(MDIE, FTIE, RSNIE, EAPKIE[ANonce, R1 PMK-Name, KeyLifetime])
PTK Derived
Re-association Request(MDIE, FTIE, RSNIE, RIC-IE, EAPKIE[Anonce, R1 PMK-Name, MIC])
PTK Derived
Re-assert parameters for integrity protection; Message 3 and 4 of 4-way handshake; RIC IE for QoS parameters
Re-association Response(MDIE, FTIE, RSNIE, RIC-IE, EAPKIE[Snonce, R1 PMK-Name, KeyLifetime, MIC])
802.1X Port Unblocked; Security and QoS established for 802.11 Data Frames

Figure 33: The Optimized Handshake

The optimized handshake introduces a number of new information elements. The most important of these information elements are the Resource Information Container (RIC) and the EAPOL-Key (EAPK). The mobile station uses a RIC information element in a Reassociation Request to indicate its QoS bandwidth requirements. In a successful Reassociation Response, the RIC information element indicates either that the AP could honor the mobile station's request, or else suggest an alternative amount of bandwidth which the AP can allocate. The EAPK information element piggybacks variants of the 802.11i key exchange over the reassociation.

The original 802.11 standard begins an AP-to-AP transition with an Open System Authentication. The optimized handshake replaces Open System Authentication with a new exchange. The mobile station begins by sending an "authentication" request message that includes an RSN information element, defined by 802.11i, to suggest the security policy the station wishes to use with this association. The first message also contains an EAPK information element conveying a variant of the first message of the 802.11i 4-Way Handshake. The EAPK information element includes a 256 bit random challenge called SNonce and the R0 Key Holder (R0KH) identity with which the STA performed its Initial Association. This allows the targeted AP to identify or acquire the needed PMK-R1.

The AP responds with message 2, which includes corresponding RSN and EAPK information elements. The response EAPK information element includes a 256 bit random challenge ANonce from the AP. The message includes the PMK-R1 name and key lifetime.

When the mobile station is ready to transition, it resumes the optimized handshake with a Reassociation Request as message 3. This includes a RIC information element, to request QoS bandwidth, and an EAPK information, continuing the key exchange. The third message reflects the AP's ANonce from Message 2. It also incorporates a message integrity code (MIC) to detect forgeries.

The AP finishes the exchange with a Reassociation Response. The Reassociation Response either confirms that the reassociation succeeded or failed. The message can also indicate that the reassociation succeeded, but that the AP could not honor the station's bandwidth request. When the reassociation succeeds, the AP and station open their respective 802.1X controlled ports, permitting data to flow over the protected channel between them. In the case where the reassociation succeeded but the AP could not honor the station's bandwidth request, the station may invoke normal 802.11e mechanisms to procure bandwidth beyond that set aside for best effort traffic. A MIC also protects this message from forgery, including the SNonce replicated from Message 1. The Optimized Handshake assumes a secure transfer of the PMK-R1 from the R0KH to the R1KH, but does not define this, as back-end mechanisms in the infrastructure are out of scope of the standard. The design is to enable both push and pull key deliver mechanisms.

6.3 Security Analysis

The key hierarchy and the transition algorithm have their own security defects.

6.3.1 Key Hierarchy Security Analysis

Ironically, 802.11r draft 2.0 includes a security error that prevents the STA from distinguishing correct key sharing from a compromised PMK. It defines

PMK-R1 = kdf(PMK-R0, "R1 Key Derivation", PMKR0Name || R1KH-ID || 0x00 || SPA)

where,

- kdf(., .,.) is the 802.11r key derivation function, with its first parameter being a key derivation key, its second a label, and the third a context delimiter;
- PMKR0Name is the name of the PMK-R0 key;
- R1KH-ID denotes that name of the R1 Key Holder; and
- SPA denotes the MAC address of the mobile station.

It is easy to create an attack against this key share when 802.11r is used with "fat" access points:

- Compromise a "fat" AP
- At a second "fat" AP, provision the R1KH-ID and the PMK-R1 from the compromised AP.

If a mobile station transitions from the compromised AP to the new AP, the 802.11r design allows the station to use the same PMK-R1 to derive its PTK, in countering the desired solution to one of the problems that originally motivated the 802.11r design. This flaw is easily rectified by putting the mobility domain's R1 Key Holders into one-to-one correspondence with APs, and letting the MAC address of each AP be the R1KH-ID. This security flaw is expected to be corrected in future versions of the 802.11r draft.

6.3.2 Transition Security Analysis

Unfortunately, draft 2.0 of the 802.11r introduces three security flaws into the transition process:

- Message 2 of the optimized handshake fails to include any quantity computed from the random value SNonce from Message 1 of the optimized handshake. Thus, it is impossible for the mobile station to match the appropriate Message 2 to reply with its request in Message 1, i.e., Message 2 does not belong to a well-defined session. The 802.11i 4-Way Handshake solved this problem by deriving the PTK from SNonce and including a MIC computed over the message under part of this key. Since the PMK-R1 is not necessarily available for computing Message 2, the only viable solution seems to be to include SNonce in Message 2 instead.
- A flooding attack exists against Message 3, the Reassociation Request. To mount this attack, the adversary floods a large number of first messages to the AP, each giving a distinct SNonce value. The AP will not know that these messages are invalid, and must cache a <SNonce, ANonce> pair for each (or the PTK that results from each of them). Then the adversary sends one Reassociation Request for each SNonce value, each with a random value in place of the Message 3 MIC. While these

Reassociation Requests will be invalid, the AP can determine this only by trying to verify MIC using each and every <SNonce, ANonce> pair. Thus, the protocol requires the AP to undertake Message 2 MIC verifications for n forged Message 1's. It may be argued that using distinct ANonce values for each ANonce alleviates this problem directly, but the third security flaw below rules out this option. 802.11i solved this problem by requiring the mobile station to commit to its SNonce value and by including SNonce instead of ANonce in its equivalent of Message 3. 802.11r could defend against the flooding attack with the same precaution.

- A similar flooding attack exists against Message 4, the Reassociation Response. This defect can be rectified by requiring the AP to commit to its ANonce value for this session and by including ANonce instead of SNonce in Message 4. Committing to an ANonce value rules out the naïve solution to the second security flaw.

6.4 PERFORMANCE EVALUATION

How does this design measure up against its objectives? Qualitatively, it seems to improve performance noticeably. The design removes the expensive scanning step from the direct transition process. It reduces the number of round trips at transition from five or more to two, and allows pre-computation of the PTK. It permits data to continue to flow until the moment of transition, resulting in fewer lost data frames. This section presents experimental evidence supporting this qualitative analysis.

A Linux-based prototype [31] is implemented to characterize the transition times of the optimized handshake. The Linux hostAP daemon is modified to support cooperative BSS transition. The mobile station prototype was based on the public domain Linux WPA Supplicant, and used the IPW 2200 b/g wireless card and its associated Linux device driver. The WPA Supplicant software is modified to use the BSS transition algorithms. The implementations support both first contact and subsequent transitions. Figure 34 depicts the experimental framework used in our test bed for the "fat AP" model. The authors are planning to conduct a future research experiment for the controller-based "thin" AP.

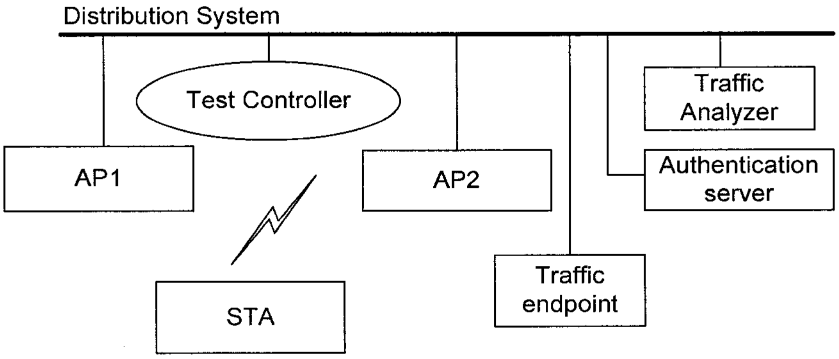


Figure 34: Test-bed setup for BSS Transition

The test bed consists of two Linux-based APs labeled as AP1 and AP2, a traffic endpoint server, a traffic analyzer, and an 802.1X authentication server. The APs are set to different channels to avoid co-channel interference. The traffic analyzer aggregates data about the system's behavior during the experiment. The authentication server provides the PMK-R0 during initial contact. The Distribution System is provided by an Ethernet. We used an SSL channel to transfer the appropriate PMK-R1 from AP1 to AP2.

In our experiment, a STA performs an initial association with AP1 to populate the key hierarchy. Once this association is complete, the STA establishes a two-way, constant bit rate data stream with the traffic endpoint server, with one message in each direction every 20 milliseconds. This process simulates VoIP over WLAN RTP traffic using a 20 milliseconds codec. The STA is moved from AP1 towards AP2. Since it is difficult to know when a STA actually decides to move, the station itself has to record when it enters and exits transition. The time taken for transition is measured during the baseline 802.11i and prototype of BSS transition optimization scenarios and compared. The results are collated and presented in Figure 35 and Figure 36. The 802.11r optimized handshake shows that there is significantly less transition time compared to the methods based on IEEE 802.11i, resulting in fewer packet loss and hence much better user experience through improved voice quality [32].

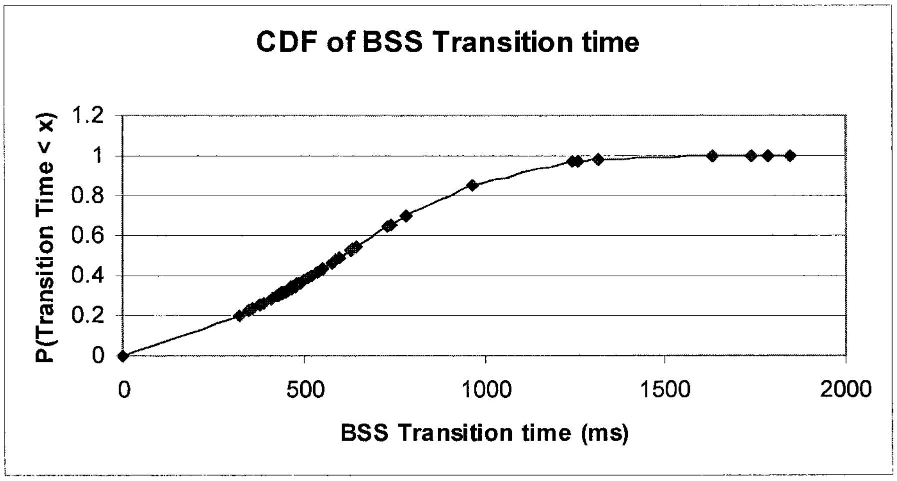


Figure 35: CDF of BSS Transition Time

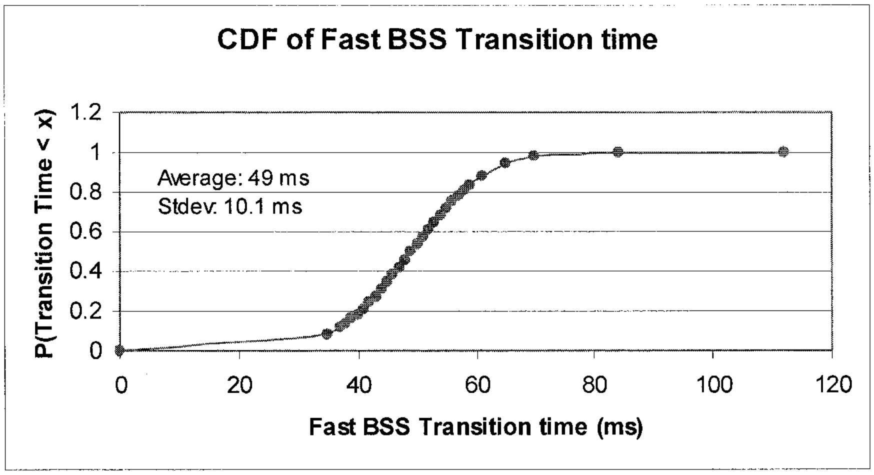


Figure 36: CDF of BSS Transition Optimization Time

Table 8 shows the detailed results of average roaming time and packet loss using baseline 802.11i authentication and BSS transition optimization. The experiment result was achieved by running several sessions of a 2-way G.711u codec VoIP stream (20 ms interval) between the STA and endpoint and performing BSS transition during a particular instance of the testing. The transition time and the packet losses recorded during each session were averaged to obtain the following results.

Table 8: Effect of Roaming on Packet Loss

Authentication method	Average Roaming time (millisecond)	Average Packet loss %	Maximum consecutive lost datagram (Average)
Baseline – Full 802.1X EAP authentication	525	1.8	53
Cooperative BSS Transition using 802.11r	42	0.2	6

6.5 SUMMARY

Improved AP-to-AP transition performance is a critical mobility feature for Voice over IP over WLAN. It is important to maintain acceptable voice quality [9] during the roaming process while also maintaining acceptable security and quality of service. The research prototype has shown that, with the proposed cooperative transition optimizations described in this chapter, typical transition times range from 25 – milliseconds, a 90% reduction compared with 200 – 500 milliseconds that was measured without transition optimization.

7 SECURITY REQUIREMENTS AND PROTECTIONS

As discussed in the previous chapters, radio resource measurement and cooperation are expected to collect wireless environment data and facilitate better wireless performance and management. This environmental data can be analyzed by wireless devices to estimate signal quality, channel load, and wireless network information. Therefore, based on the medium sensing and radio resource information, cooperative algorithms can be developed to facilitate wireless network configuration and operation to improve network performance.

Action frames were described to communicate cooperative information such as radio resource measurements and cooperation between Medium Access Control (MAC) instances across a WLAN in Chapter 3, 4, 5 and Chapter 6. An Action Frame is a class 3 IEEE 802.11 Management Frame. Action Frames are subject to forgery, because the existing 802.11 security mechanisms protect only Data Frames. Forged Action Frame messages can lead to poorer performance (or worse), as well as misleading topology of the WLAN configuration.

This chapter analyzes security threats to distributed cooperative wireless networks, and proposes a mechanism to protect Action frames by extending the existing 802.11 security scheme. In doing so, it is first necessary to review major components and features of radio resource measurements and cooperation that are proposed in the previous chapters. The goal of the proposed security techniques is to protect Action Frames without introducing a new encryption and key management scheme. The protection scheme proposed in this dissertation is backward compatible with the existing Action Frames used widely in other 802.11 amendments, so it extends automatically to protect many other management messages besides those of radio resource measurement and cooperation.

7.1 Security Assessment

Three classes of threats are considered: data exposure threats, data forgery threats, and denial-of-service (DoS) threats. There is a fourth class of threats not considered: mis-measurement and mis-controlled threats. A complete threat analysis requires understanding of how the measurements are implemented, how the measurements are used, and how the cooperation is responded to by the peers. Therefore, each of the radio resource measurement and cooperation frames is analyzed against this model, and conclusions are drawn about the security requirements implied by each threat.

7.1.1 Radio Measurement Requests

Radio Measurement Request messages are used to solicit measurements. Measurement Requests include requests for the Beacon Request, Frame Request, Channel Load Request, Noise Histogram Request, Location Configuration Information Request and STA Statistics Request. 802.11k permits a station to send a Measurement Report frame in response to a Measurement Request, or without any solicitation.

All of the information conveyed by these requests is public information. That is, all of it can be learned by other channels, such as by passive monitoring. What is not otherwise public is the fact that a specific station is requesting specific measurements. Hence, these requests introduce a new data exposure threat. Defending against this threat, however, would require a mechanism to defeat traffic analysis, which is a threat usually not addressed by commercial systems. Thus, this threat will be ignored in this thesis.

Every request message can be forged. The consequences of a forged request vary, depending on the implementation of the receiver. If the receiver rejects additional requests while performing a measurement, then allowing forgeries presents a new denial of service opportunity. Similarly, if the receiver queues pending requests while performing a measurement, then any physical realization will have a finite queue length, and an attacker might be able to create a denial-of-service by filling the queue with forged requests. As another example, the attacker can alter the destination of the recipient of a genuine request, targeting a STA that the actual requestor did not intend to poll; this scenario results in denial-of-service against both the requestor and the targeted STA. While denial-of-service opportunities are inherent in the design of IEEE 802.11, these denial-of-service attacks are new and specific to the IEEE 802.11k measurement system. This list of possible forgeries suggests that each of these requests should be protected from forgery whenever feasible.

A stream of requests can be used directly to deny service by consuming all of the channel bandwidth, not just the measurement resources of the intended receiver. This, however, is no different than any other type of message or jamming, and so is not a new denial-of-service threat. Similarly, an attacker can jam any particular measurement message, but can do the same for any other message. Accordingly, it makes little sense to protect against these sorts of denial-of-service attacks.

7.1.2 Radio Measurement Reports

In response to a Measurement Request, a station may send a Measurement Report frame. The information in the report seems benign, so the natural first reaction is to say confidentiality is unnecessary. This reaction would be incorrect, however. The specific measurements taken by individual stations relate to the particular conditions at the location of the station and cannot be obtained by simple monitoring. Therefore, there may be circumstances where confidentiality of measurements is desirable.

There are many ways to create a forged Measurement Report. A forgery could omit information from a genuine Measurement Report, insert new information into a genuine Measurement Report, or alter information in a genuine Measurement Report. An attacker could also create a Measurement Report that the real sender never sent. Omission of information could be used to hide the presence of rogue stations and access points. Insertion of information into a Measurement Report could be used in an attempt to cover up the fact that devices have been taken off-line in an unauthorized fashion. Changed Measurement Reports or entirely forged Measurement Reports can be used to mask the true WLAN topology. All of these attacks undermine the utility of the Measurement Report, so when security is an issue, Measurement Reports should be protected from forgery.

In addition, all of the measurement reports are stored in the MIB table, and the MIB can be accessed by SNMP. Some versions of SNMP permit data confidentiality.

If it is the policy of the local domain to apply a data confidentiality mechanism when this data is transported by SNMP, then allowing the STA statistics report to be sent without confidentiality would undermine the organization's SNMP policy. This means that the basic 802.11k security design must provide a mechanism to preserve the confidentiality of this Report message.

7.1.3 Neighbor Report Requests

Mobile stations use Neighbor Report Request messages to solicit Neighbor Reports. An Access Point will send a Neighbor Report Response in response to a Neighbor Report Request, or without any solicitation.

All of the information conveyed by Neighbor Report Requests is public information. A Neighbor Report Request message can be forged. As an example, the attacker can alter the destination of the recipient of a genuine request, targeting an Access Point that the actual requestor did not intend to poll; this is a novel denial-of-service against both the requestor and the targeted STA. The Neighbor Report Request therefore requires protection from forgery.

7.1.4 Neighbor Report Responses

The Neighbor Report identifies a collection of access points that are candidates to which a station can transition (i.e., "roam"). A Neighbor report is sent by an AP; it contains information on known neighbor APs. The Neighbor Report is compiled by the AP. It may include information from measurement reports received from the STA's within the BSS, information obtained via the management interface, or the Distribution System (DS).

The information in the Neighbor Report is public, so it ought to introduce no new confidentiality requirement. However, Neighbor Report contents are derived from the Management Information Base (MIB) table, and the MIB can be accessed by SNMP. Some versions of SNMP permit data confidentiality. If it is the policy of the local domain to apply a data confidentiality mechanism when this data is transported by SNMP, then allowing the Neighbor Report to be sent without confidentiality would undermine the organization's SNMP policy. This means that in some cases it is desirable to preserve the confidentiality of the Neighbor Report.

An attacker can eliminate elements from the Neighbor Report to create a novel denial-of-service, insert elements into the report to mask unauthorized shutdown of another AP, or change the information reported about a particular BSSID to create novel denial-of-service attacks and effect service theft. The Neighbor Report therefore requires protection from forgery.

7.1.5 Radio Resource Cooperation Request and Response

Radio Resource Cooperation (RRC) Request and Response messages are used to cooperate and facilitate wireless network cooperation, which includes configuration cooperation, performance and resource cooperation, operations cooperation, and location service. Load Balancing Cooperation Request and Response, discussed in a previous chapter, is one the Wireless Network Cooperation messages.

Radio Resource Cooperation messages are subjected to forgery attacks. An unauthorized entity could set or change any unprotected cooperation parameter, including those related to configuration, operations, and accounting. RRC message could be reordered or replayed to effect unauthorized cooperation operations. For example, an unauthorized entity could modify load balancing message to direct the STA to another AP. An unauthorized entity may also attempt some cooperation operations by assuming the identity of an authorized entity. For example, an unauthorized AP who is not authorized for STA's contention window size update may attempt to update STA's contention window size.

RRC messages also need to be protected from disclosure to an authorized entity. An entity could observe data exchanged between an AP and a STA and thereby learn the values of managed objects and learn of notify-able events. For example, the observation of a set of location information and cooperation (for Location Service) would enable an attacker to learn asset tracking. RRC policy needs to be consistent with SNMP v3 policy for confidentiality, because all WNM objects are in the SNMP MIB. If SNMP policy requires confidentiality, so does RRC.

Denial-of-service is another attack on RRC messages. An attacker may prevent exchange between AP and STA and cause Wireless Network Connection Failure and disruption of all types of exchanges. Cooperation message forgery can create novel denial-of-service attacks

A complete Security assessment for the above Radio Resource Measurement and Cooperation (RRMC) is depicted in Table 9.

Table 9: Security Assessment of RRMC Message

Measurement Messages	Security Assessment		
	Confidentiality	Forgery	DoS
Radio Measurement Request	Yes	Yes	Yes
Radio Measurement Response	Yes	Yes	Yes
Neighbor Report Request	No	Yes	Yes
Neighbor Report Response	Yes	Yes	Yes
Radio Resource Cooperation Request	Yes	Yes	Yes
Radio Resource cooperation Response	Yes	Yes	Yes

7.2 Action Frame Protection

The security assessment section (section 7.1) concluded that all of Radio Measurement Request, Radio Measurement Response, Neighbor Report messages and Radio Resource Cooperation messages require protection from forgery, and many require confidentiality protection, as well. Since all measurement messages are carried by Action Frames, these security requirements can be addressed by providing a way to protect Action Frames. This section gives an overview of the Data Frame protection scheme from 802.11i, and then builds an Action Frame protection scheme on the 802.11i scheme.

7.2.1 Data Frame Protection

The original IEEE 802.11 standard [3] defined a data confidentiality mechanism called Wired Equivalent Privacy (WEP). The security goal of WEP is data confidentiality equivalent to that of a wired LAN. It is well known that WEP falls short of its goal. IEEE created the 802.11i standard [11] to address WEP's shortcomings. This standard does so by introducing more comprehensive security architecture and more robust security algorithms.

The 802.11i introduces a security capabilities discovery and negotiation mechanism. The 802.11 uses a data structure called an Information Element (IE) for this function, and the 802.11i defines a new IE, called the Robust Security Network IE (RSN IE). In Beacon and Probe Responses, the RSN IE advertises the authentication and cipher suites used, as well as other security-related characteristics used in the WLAN. In (Re-) Association Requests, the IE selects the security capabilities for the session from those advertised.

The 802.11i uses the IEEE 802.1X, a separate stand-alone standard, to effect authentication. The 802.11i defines a new key management scheme to deliver fresh session keys to protect the link between the station and the access point and to synchronize replay counters. The key management scheme is also used to defend against downgrade attacks, where one of the RSN IEs has been forged by an attacker.

The 802.11i defines a new cipher suite called AES-CCMP. AES-CCMP is based on AES with 128 bit keys in CCM mode. CCMP protects Data Frame confidentiality and prevents data forgeries and replays. The 802.11i also defines a new cipher suite called TKIP, meant to be implemented as a software patch to WEP-capable equipment already in the field. TKIP provides relatively weak but real security guarantees, addressing all known WEP deficiencies.

7.2.2 PROTECTION FOR ACTION FRAME

This thesis describes how to extend 802.11i to protect Action Frames. The scheme defines a new Action Frame attribute, which determines whether or not the Action Frame can be protected. This approach results in two classes of Action Frames: a Protection-capable Frame and a Non-Protection-capable Frame. By default, all Action Frames are Non-Protection-Capable for backward compatibility. Non-Protection-capable Action Frames will be "normal" Action Frames – protection never applied. Protection-capable Action Frames can be protected and will be protected whenever local policy requires. If the WLAN policy does not require protected Action Frames, then stations and access points send all Action Frames without protection, including all Protection-capable Action Frames. If the WLAN policy requires protected Action Frames, then a station protects all Protection-capable Action Frames. A station does not send Protection-capable Action Frames at all if the peer has not agreed to protection. If the WLAN policy requires protected Action Frames, then a station discards any unprotected Protection-capable Action Frame it receives, including those received before the IEEE 802.11i key management has established. A station never protects a Non-Protection-capable Action Frame it sends and discards any it receives that are protected.

This scheme allocates a bit in the RSN IE to negotiate Action Frame protection. The access point advertises the WLAN policy of protecting Action Frames by asserting this bit in Beacon/Probe Responses, and clears the bit to advertise that Protected Action Frames are not protected. Responding stations set the bit as the

Beacon/Probe Response source sets the bit if they support Protected Actions Frames and clear the bit otherwise. Figure 37 details this negotiation.

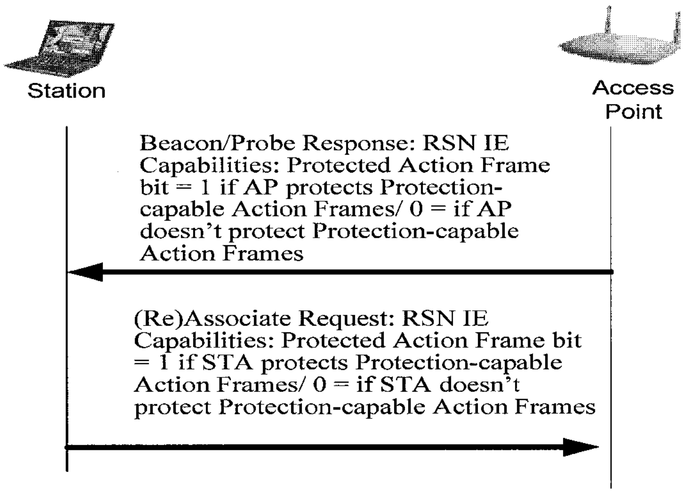

Station
Access Point
Beacon/Probe Response: RSN IE Capabilities: Protected Action Frame bit = 1 if AP protects Protection-capable Action Frames/ 0 = if AP doesn't protect Protection-capable Action Frames
(Re)Associate Request: RSN IE Capabilities: Protected Action Frame bit = 1 if STA protects Protection-capable Action Frames/ 0 = if STA doesn't protect Protection-capable Action Frames

Figure 37: Action Frame Protection Negotiation

If Action Frame protection is negotiated, the access point and station will apply the negotiated 802.11i unicast cipher suite and key to all unicast Protection-capable Action Frames, while sending all Non-Protection-Capable Actions Frames without any protection. The negotiated 802.11i Group key will be used by the Access Point and Station to protect the broadcast and multicast Action Frames. Since both the AES-CCMP and TKIP cipher suites provide protection against confidentiality, forgery, and replay threats, this scheme extends the same protection to Action Frames. The protected Action Frame format is depicted Figure 38.

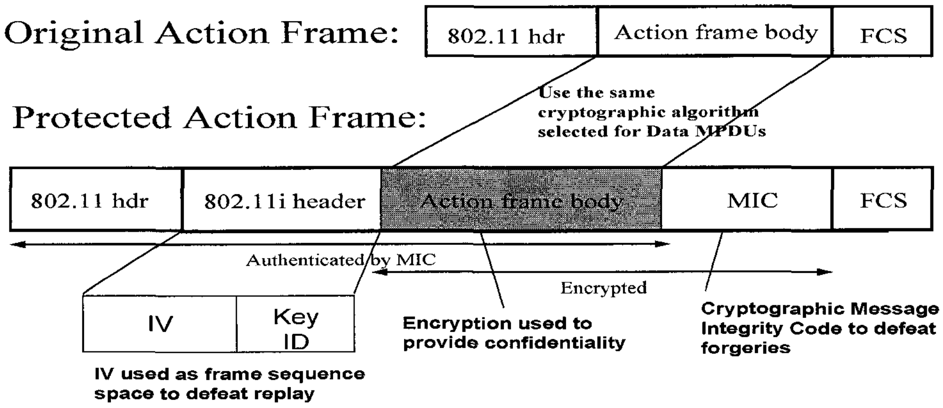

Original Action Frame:
802.11 hdr
Action frame body
FCS
Protected Action Frame:
Use the same cryptographic algorithm selected for Data MPDUs
802.11 hdr
802.11i header
Action frame body
MIC
FCS
Authenticated by MIC
Encrypted
IV
Key ID
IV used as frame sequence space to defeat replay
Encryption used to provide confidentiality
Cryptographic Message Integrity Code to defeat forgeries

Figure 38: Protected Action Frame Format

There is an interesting issue around replay protection. Since 802.11 sends Action Frames at a priority independent from all data priorities, the 802.11i data replay counters cannot be used. Hence the scheme introduces a new replay counter that each receiver must implement. If a Protected Action Frame is received in sequence with respect to this new counter, it can be considered for decryption and message authenticity processing; if it is out of sequence, then the frame must represent a replay or an out-right forgery. This technique allows the scheme to safely reuse the same session keys that are used to protect data, meaning no new key management scheme is required.

Figure 39 and Figure 40 detail the processing required at the transmitter and the receiver, respectively.

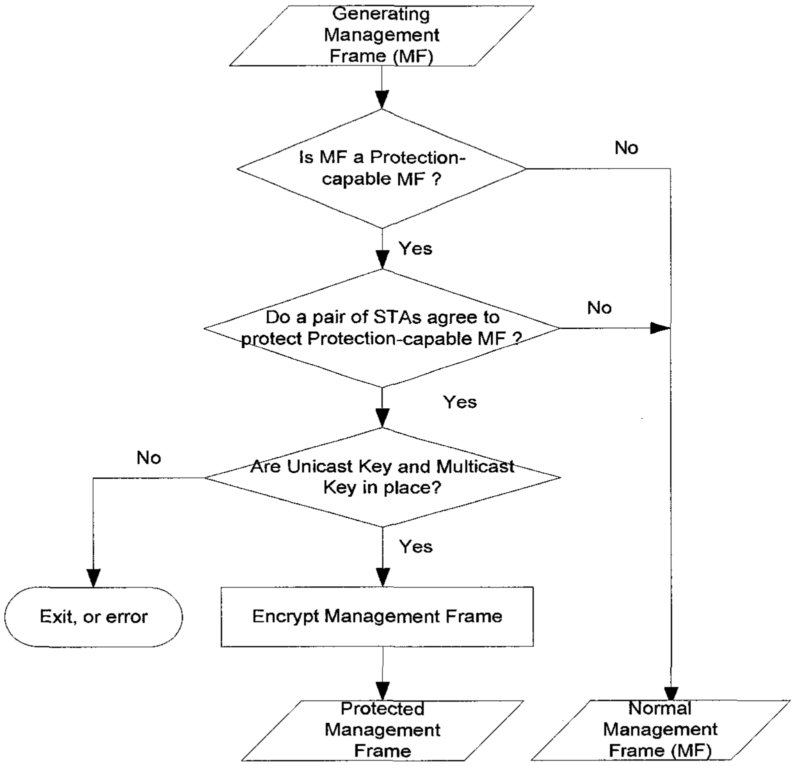

Generating Management Frame (MF)
Is MF a Protection-capable MF ?
No
Yes
Do a pair of STAs agree to protect Protection-capable MF ?
No
Yes
Are Unicast Key and Multicast Key in place?
No
Yes
Exit, or error
Encrypt Management Frame
Protected Management Frame
Normal Management Frame (MF)

Figure 39: Procedure of the transmitting STA

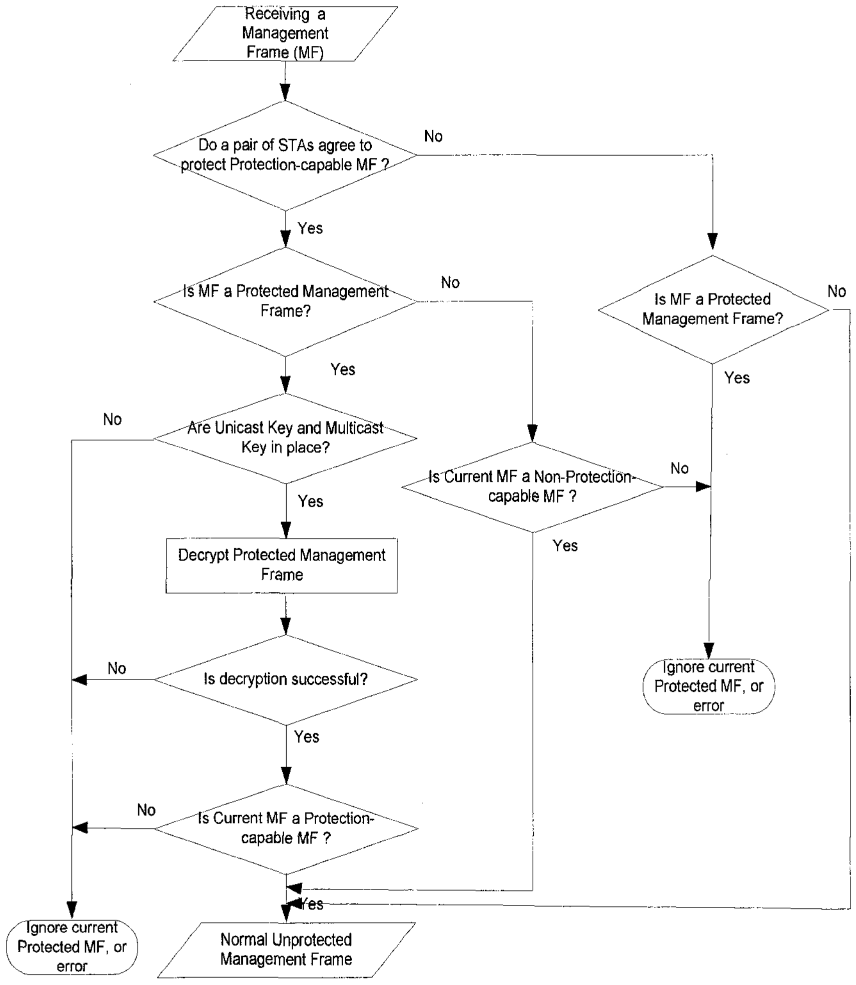

Receiving a Management Frame (MF)
Do a pair of STAs agree to protect Protection-capable MF ?
No
Yes
Is MF a Protected Management Frame?
No
Yes
Are Unicast Key and Multicast Key in place?
No
Yes
Decrypt Protected Management Frame
Is decryption successful?
No
Yes
Is Current MF a Protection-capable MF ?
No
Yes
Ignore current Protected MF, or error
Normal Unprotected Management Frame
Is Current MF a Non-Protection-capable MF ?
No
Yes
Is MF a Protected Management Frame?
No
Yes
Ignore current Protected MF, or error

Figure 40: Procedure of the receiving STA

7.2.3 Protect Broadcast Frames against Insider Forgery

The broadcast and multicast Action Frame, one of the IEEE 802.11 management frames, is used to request radio resource measurement, network information, and network optimization cooperation in this thesis as well in IEEE 802.11 amendments (11k, and 11h). These management frames contain valuable radio resource requirement and network information and are subject to forgery. The consequences of a forged broadcast request vary depending on the implementation of the receiver. For instance, if the receiver rejects additional requests for measurement while performing a measurement, then allowing forgeries presents a new denial of service opportunity. Similarly, if the receiver queues pending requests while performing a measurement, then any physical realization will have a finite queue length, and an attacker might be able to create a denial-of-service by filling the queue with forged requests. As another example, the forged network cooperation messages can lead to poorer performance (or worse), than by ignoring valid messages.

In the previous section, the 802.11i security mechanisms were extended to broadcast and multicast. However, this scheme is subject to an insider forgery. The forged radio resource measurement request and network management cooperation messages can lead to poor performance for wireless device and wireless network.

This section defines a new protection scheme to protect broadcast and multicast data and management frames from outsider forgery, as well as insider forgery. Major steps in this mechanism are described below. The message flow of this protocol to IEEE 802.11 WLANs is depicted in Figure 41.

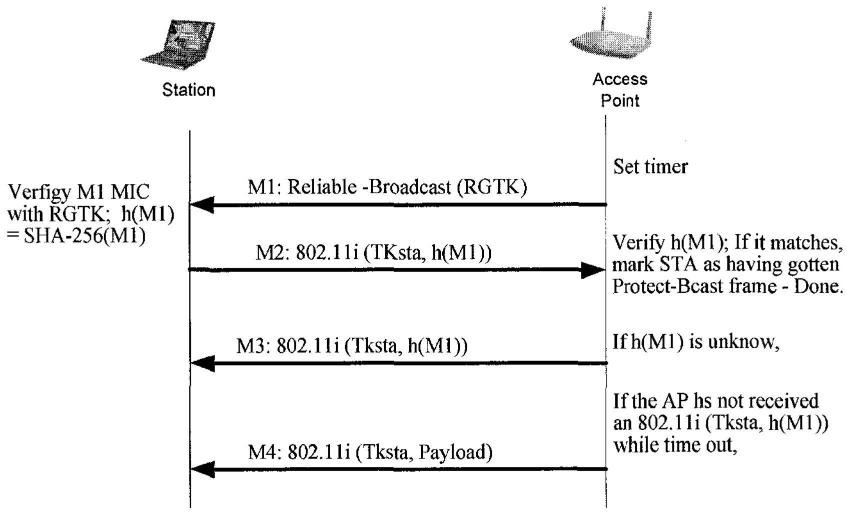

Station
Access Point
Set timer
Verfigy M1 MIC with RGTK; h(M1) = SHA-256(M1)
M1: Reliable -Broadcast (RGTK)
M2: 802.11i (TKsta, h(M1))
Verify h(M1); If it matches, mark STA as having gotten Protect-Bcast frame - Done.
M3: 802.11i (Tksta, h(M1))
If h(M1) is unknow,
If the AP hs not received an 802.11i (Tksta, h(M1)) while time out,
M4: 802.11i (Tksta, Payload)

Figure 41: Protection mechanism against insider forgery

Key Provisioning

The AP generates a new key called RGTK (Reliable GTK). The AP distributes RGTK with the GTK. A new key is necessary, because the scheme below relies on a non-802.11i cipher suite to protect broadcasts.

Messages

Define a new message: Protected Broadcast MPDU. This message is defined as follows:

Reliable-Broadcast(K) ::= Dot11Hdr | Payload | MicIE

MicIE ::= keyId | seqNum | Mic(K)

Mic(K) ::= AES-CMAC(K, Mute(Dot11Hdr) | Payload | keyId | seqNum)

Here,

Dot11Hdr is the 802.11 MPDU header.

keyId | seqNum is the 802.11i header, with one of the reserved key id bits set to 1, to indicate this is a reliable broadcast.

Payload is the “data” is sent in the MPDU.

MicIE is an information element consisting of an 802.11i header and a MIC.

Mic is a message integrity code based on AES-128 in CMAC mode.

Mute(.) means to apply the 802.11i header muting rules to its argument.

AES-CMAC(.,.) means computing the AES-128-CMAC of its second argument, using its first argument as a 128 bit key.

The Reliable-Broadcast MPDU format puts the 802.11i “header” and MIC in a non-standard location, after the payload, to preserve backward compatibility with broadcast receivers that do not support this protection.

Protocol

```
AP:  set a timeout
AP → STA: Reliable-Broadcast(RGTK)              // M1
STA: Verify M1 MIC with RGTK
STA: Calculate h(M1) = SHA-256(M1)
STA → AP: 802.11i (TKSTA, h(M1))                // M2
STA: Process payload
AP: if h(M1) is unknown,
        AP → STA: 802.11i (TKSTA, h(M1))        // Countermand (M3)
else
                Mark the STA as having gotten the Protected-Broadcast
On timeout:
```

If the AP has not received an 802.11i (TKSTA, h(M1)) from the STA, and the STA has negotiated to use the protection

AP → STA: 802.11i (TKSTA, Payload) to the STA // Resend Payload (M4)

Here,

"A → B:m" means A sends message m to B

Reliable-Broadcast(.) was defined in the previous section.

RGTK is the Reliable GTK defined above.

h(M1) is the SHA-256 hash of message M1 = Reliable-Broadcast(RGTK).

802.11i(.,.) is an 802.11i protected message. The first argument is the temporal key used in this scheme, and the second is the data payload. TKSTA is the 802.11i pairwise temporal key used to protect this session.

The AP advises the STA with message M3 that it received a forged message from an insider. An adversary can corrupt messages M2, M3, and M4, thus preventing the reliability mechanism from working. However, this is no worse than any other acknowledgement mechanism h(M1) that identifies the transaction the AP has initiated. The scheme uses a unicast message M4 to deliver the payload to those STAs that have not responded by the timeout. There is no data confidentiality with this scheme.

7.3 SUMMARY

This Chapter described a threat analysis of the protocol for radio resource measurement and cooperation messages. The chapter also proposed a straight-forward extension to 802.11i that can be used to address these threats for action frames. To protect broadcast action frames from insider forget, a new forgery protection was also discussed. The proposed solution immediately addresses similar threats to many other 802.11 management messages as well. Adoption of this scheme within the IEEE 802.11 working group is currently underway.

8 CONCLUSION AND FUTURE WORK

The dissertation analyzed radio resource environment, and presented a distributed cooperative framework and algorithms to optimize wireless network performance, along with security enhancements.

The cooperative MAC architecture is based on medium sensing and radio measurement. The cooperative MAC architecture and distributed framework provide the capability for mobile clients and Access Point to adapt multiple optimizations simultaneously and cooperatively. As examples of the distributed cooperation, three cooperative algorithms have been explored:

- The Adaptive Contention Window Size to maximize channel utilization.
- The Cooperative Load Balancing to maximally balance network loads and utilize the network capacity.
- The Cooperative BSS Transition to optimize BSS transition time while retaining good security.

With encouraging simulation or experiment results that demonstrate a significant improvement of wireless network performance, the distributed cooperative architecture and algorithms provide excellent starting points for development of distributed cooperative wireless networks. IEEE 802.11 standards committee is working on a new standard IEEE 802.11v which is suitable for distributed cooperative wireless networks and wireless network management. The author envisions that a wide range of sophisticated cooperative algorithms will be developed and deployed to improve wireless network and mobile device performance. One particularly interesting research area that the author wants to pursue is cooperative power saving for energy-constrained battery-powered mobile devices as energy efficiency is becoming the predominant issue in the mobile industry.

REFERENCE

[1]. IEEE, "IEEE Standard for Information Technology – Telecommunications and Information Exchange between systems – Specific Requirements – Part 11: Wireless LAN MAC and PHY specifications", IEEE Std. 802.11-1997.

[2]. IEEE, "IEEE Standard for Information Technology – Telecommunications and Information Exchange between systems – Specific Requirements – Part 11: Wireless LAN MAC and PHY specifications", IEEE Std. 802.11-1999.

[3]. IEEE, "IEEE Standard for Information Technology – Telecommunications and Information Exchange between systems – Specific Requirements – Part 11: Wireless LAN MAC and PHY specifications", IEEE Std. 802.11-2003.

[4]. IEEE, "Draft supplement to Standard for Telecommunications and Information Exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Enhancements for Higher Throughput", IEEE Std 802.1n draft 1.09, February 2007.

[5]. Y. C. Tay, K. C. Chua, "A Capacity Analysis for the IEEE 802.11 MAC Protocol", Wireless Networks 7, pp159-171, 2001 Kluwer Academic Publishers, 2001.

[6]. Y. Xiao, J. Rosdahl, "Throughput and Delay Limits of IEEE 802.11", IEEE Communications Letters, Vol. 6. No. 8, pp355-357, August 2002.

[7]. D. Qiao, S. Choi, K. G. Shin, "Goodput Analysis and Link Adaptation for IEEE 802.11a Wireless LANS", IEEE Transactions on Mobile Computing, Vol. 1, No. 4, pp278-292, October-December 2002.

[8]. L. Bononi, M. Conti, L. Donatiello, "Design and Performance Evaluation of a Distributed Contention Control (DCC) Mechanism for IEEE 802.11 Wireless Local Area Networks", Wowmom 98 Dallas Texas USA, pp59-67, ACM 1-58113-093-7, 1998.

[9]. G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", IEEE Journal on Selected Areas in Communications, Vol. 18, No 3. pp535-547, March 2000.

[10]. F. Cali, C. Marco, G. Gregoti, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism", IEEE Journal on Selected Areas in Communications, Vol. 18, No 39 pp1774-1786, March 2000.

[11]. IEEE, "Draft supplement to Standard for Telecommunications and Information Exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Amendment i: Medium Access Control (MAC) Security Enhancements", IEEE Std 802.11i draft, July 2004.

[12]. IEEE, "Draft supplement to Standard for Telecommunications and Information Exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Amendment e: Quality of Service (QoS) Enhancements", IEEE Std 802.11e draft D13.0, January 2005.

[13]. IEEE, "Draft supplement to Standard for Telecommunications and Information Exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Specification for Radio Resource Measurement", IEEE Std 802.1k draft 7.0, March 2007.

[14]. IEEE, "Draft supplement to Standard for Telecommunications and Information Exchange between systems – LAN/MAN Specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Amendment r: Fast BSS Transition", IEEE Std 802.11r draft D2.0, March 2006.

[15]. N. Cam-Winget, R. Housley, D. Wagner, and J. Walker, "Security Flaws in 802.11 Data Link Protocols," Communications of the ACM, Vol 46, No 5, pp 35-39, May 2003.

[16]. N. Borisov, I. Goldberg, and D. Wagner, "Intercepting mobile communications: The Insecurity of 802.11," Proceedings International Conf. on Mobile Computing and Networking, pp 180-189, July 2001.

[17]. J. Walker, "Unsafe key size: an analysis of WEP," IEEE 802.11 doc 00-362, October 2000.

[18]. Arbaugh, W, A. Mishra, and M. Shin,"An empirical analysis of the IEEE 802.11 MAC layer handoff process." ACM SIGCOMM CCR Vol 33(2), pp 93-102, April 2003.

[19]. Andrew S. Tanenbaum, Computer Networks, 3rd edition, ISBN 0-13-349945-6, 1994.

[20]. Jean-Poerre Ebert and Adam Wolisz, "Combined Tuning of RF Power and Medium Access Control for WLANs", IEEE 0-78-3-5904-6/99, 1999.

[21]. Emily Qi, Ravi Murty, Larson Dylan *"An Adaptive Approach to Wireless Network Performance Optimization"*, Technology@Intel Magazine, February/March 2004.

[22]. K.C. Chen, "Medium Access Control of Wireless LANs for Mobile Computing", IEEE Networks, 9-10, 1994.

[23]. J. Hastas, T. Leighton, B. Rogoff, "Analysis of Backoff Protocols for Multiple Access Channels", Siam J. Computing. Vol. 25, No. 4, pp740-774, August 1996.

[24]. T. S. Ho, K. C. Chen, "Performance evaluation and enhancement of then CSMA/CA MAC protocol for 802.11 wireless LANs" , IEEE PIMRC, Taipei, Taiwan, pp392-296, October 1996.

[25]. G. Bianchi, "IEEE 802.11 – Saturation Throughput Analysis", IEEE Communication Letter, Vol. 2, pp318-320, December 1998.

[26]. Emily Qi, Kapil Sood, Vivek Gupta *"Seamless Platform Mobility Across Wireless Networks",* Technology@Intel Magazine, August 2005

[27]. A. Veres, A. T. Campbell, M. Barry, "Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control", IEEE Journal on

Selected Areas in Communications, Vol. 19, No 10, pp2081-2093, October 2001.

[28]. C.Yap, Emily Qi, K. Sood, S. Bangolae and C. Bell, "Issues with real-time streaming applications roaming in QoS-based secure IEEE 802.11 WLANs", IEE Mobility Conference, October 2005.

[29]. Emily Qi, Kapil Sood, Jesse Walker, et al, "802.11 TGr Just-In-Time Transition Acceleration Proposal (JIT-TAP) Proposal", Submission to IEEE 802.11 Task Group r: 11-05-0362-01-000r-jit-tap-proposal-text.doc, January 2005.

[30]. Dorothy Stanley, Pat Calhoun, et al, "Transition Acceleration Protocol (TAP)", Submission to IEEE 802.11 Task Group r: 11-04-1542-00-000r-transition-acceleration-protocol-draft-text.doc, November 2004.

[31]. Emily Qi, Sangeetha Bangolae, Kapil Sood, and Jesse Walker, "BSS Transition Optimizations and Analysis for VoIP over WLAN", Springer's International Journal on Wireless Personal Communication, July 2007.

[32]. Intel White Paper, "Overcoming Barriers to High-Quality Voice over IP Deployments".