Voice Quality Framework for VoIP over WLANs

Md. Atiur Rahman Siddique

A thesis submitted for the degree of
Doctor of Philosophy at
Monash University

December 2010

Declaration

This thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Md. Atiur Rahman Siddique
2 December 2010

Dedicated to the one and only creator of the universe.
Dedicated to my parents for all their love and care.
Dedicated to my wife and son for their inspirations.

Copyright notices

Acknowledgments

Praise be to Allah, the most merciful, the most gracious, for blessing me with the courage, opportunity and intellect to undertake this research. I am profoundly indebted to my supervisor Dr. Joarder Kamruzzaman for his constant guidance, insightful advices, helpful criticisms and valuable suggestions. He has given me sufficient freedom to explore research challenges of my choice and guided me when I felt lost. Without his insights, encouragements and endless patience, this research would not have been completed.

My sincere gratitude goes to my parents Md. Mahbubar Rahman Siddiqui and Mrs. Fatema Siddiqui, my wife Ishrat Jahan Sumana and my elder brother Md. Matiur Rahman Siddiqui for their endless love and inspirations throughout my life. The excellent thirst for knowledge of my father and my wife was the primary driving factor behind my doctoral studies. They believed in myself more than I ever did. Writing this dissertation would not be possible without the endless patience and devotion of my wife, especially in caring for my son while I was busy writing. My mother has been my courage all my life, and every success in my life is attributed to her. I still remember calling her before every major exam just to hear her voice and regain confidence. My brother has always been there for me since before a time, when I started to realize it. Most importantly, somehow he understands everything I say. My special thanks go to my parents-in-law Prof. Md. Abdur Rahman Sarkar and Mrs. Nargis Rahman and my siblings-in-law Samina Yesmin and Abdullah Al Farooq for their endless inspirations and support in my times of need. I would like to thank my son Ayaan Raaid Siddique for smiling enough to make me forget all my worries. He has been my special source of divine joy and inspiration since his birth.

An important acknowledgment goes to Patrick Leegel for proofreading my whole thesis in a short time and Ishrat Jahan Sumana, Shaila Pervin, Md. Kamrul Islam, Md. Tanvir Hossain and Md. Golam Rabbani for their valuable comments on individual chapters despite their busy schedules. I thank Monash Research Graduate School for providing the generous scholarships and Gippsland School of Information Technology (GSIT) for providing an excellent competitive environment and resources necessary to undertake this research. Finally, I thank all staffs of GSIT and Gippsland Library and all post-graduate students of Monash University, Gippsland campus for their inspirations during the last few years.

Abstract

Voice over IP (VoIP) has quickly been adopted by commercial and private users due to its low cost of service. To attain an even greater consumer acceptance by offering the convenience of mobility, we have to use a wireless carrier to deliver VoIP packets at mobile phones. The low cost of service is the ultimate bargaining chip of VoIP and using a high cost cellular network, like GSM or CDMA, for the last mile coverage compromises its appeal. The IEEE 802.11 WLANs, in this regard, offer a low cost wireless coverage which intrigued many researchers and device manufacturers to investigate the voice performance in WLANs using theoretical analysis and simulation. Simulation based studies offer only restricted insight since the call capacity after changing any network aspect can not be determined easily and, therefore, analytical capacity models are more helpful to network designers and administrators. But a number of over-simplified assumptions used in the existing voice capacity models make these studies impractical. Standard voice quality measures were ignored and unrealistic assumptions were used in modeling the IEEE 802.11 medium access mechanisms, which make the results unreliable.

In this thesis, we model the voice capacity of the DCF and PCF based WLANs with a special attention to the voice quality to assist in network design and planning. We use the ITU-T E-model to determine the voice quality and formulate a quality impairment budget for medium and high quality calls. We develop two novel Markov models to estimate the delay and loss in the DCF and PCF based medium access mechanisms. To model real world environment closely, we consider the impacts of imperfect channel and capture effect. Since most WLANs operate under the unsaturated condition for a considerable amount of time, we consider both saturated and unsaturated traffic conditions. In conjunction to network performance, we also incorporate the characteristics of voice codecs and dejitter buffer. We identify that the delay and loss in the queue degrade voice quality significantly and utilize standard queuing analyses to determine its impact. Extensive simulation studies show a close match with our analytical results.

In the current literature, no analytical model exists for the voice capacity of multi-channel, multihop WLANs with multi-interface nodes. Especially, the formation of critical zones due to accumulated packet arrivals in multihop networks creating a bottleneck was never considered. We model multi-interface WLAN nodes operating in multichannel, multihop WLANs and determine the voice capacity of such networks. We also identify and analyze the benefit of utilizing multihop networks in order to exercise spatial reuse and improve voice capacity.

Considering that DCF offers wider coverage and spatial reuse while PCF offers lower delay by the use of time synchronized medium access, we present a novel medium access mechanism that combines the best features of the above two and yields performance benefits including VoIP call capacity increase through the introduction of "virtual access point" concept. The protocols for different network scenarios are explained and performance comparisons to the DCF and PCF based medium access are illustrated. The capacity models proposed in this thesis will be very useful in designing WLANs to maintain acceptable voice quality using off-the-shelf WLAN devices supporting DCF and PCF while the incorporation of our proposed medium access mechanism will improve the voice performance in future WLANs using virtual access point enabled mobile devices.

Acronyms

ACELP	Algebraic Code Excited Linear Prediction
ACK	Acknowledgment
AC	Access Category
ADC	Analog to Digital Converter
ADPCM	Adaptive Differential PCM
AES	Advanced Encryption Standard
AODV	Ad hoc On Demand Distance Vector
AP	Access Point
ARF	Auto Rate Fallback
AVVID	Architecture for Voice, Video and Integrated Data
AWGN	Additive White Gaussian Noise
BC-PQ	Backoff Control Priority Queuing
BEB	Binary Exponential Backoff
BER	Bit Error Rate
BSS	Basic Service Set
CAC	Call Admission Control
CAP	Controlled Access Phase
CCK	Complementary Code Keying
CDMA	Code Division Multiple Access
CELP	Code-Excited Linear Prediction
CFP	Contention Free Period
CNG	Comfort Noise Generator
CP	Contention Period
CS-ACELP	Conjugate Structure Algebraic Code-Excited Linear Prediction model
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTS	Clear To Send
DAC	Digital-to-Analog Converter
DCF	Distributed Coordination Function
DIDD	Double Increment Double Decrement
DRR	Deficit Round Robin
DSDV	Digital Simultaneous Voice and Data
DSP	Digital Signal Processing
DSSS	Direct-Sequence Spread Spectrum

DS	Distribution System
DTMF	Dual Tone Multi Frequency
DTX	Discontinuous Transmission
EDCA	Enhanced Distributed Channel Access
EILD	Exponential Increase Linear Decrease
ESP-IP	Encapsulating Security Payload with IP
FEC	Frame Error Correction
FHSS	Frequency-Hopping Spread Spectrum
GFSK	Gaussian Frequency Shift Keying
GPS	Global Positioning System
GSM	Global System for Mobile Communications or Groupe Spécial Mobile
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
HMM	Hidden Markov Model
HoQ	Head of Queue
IBSS	Independent Basic Service Set
iLBC	Internet Low Bitrate Codec
IMBE	Improved Multi-Band Excitation
IR	Infrared
ISDN	Integrated Services Digital Network
ISM	Industrial, Scientific and Medical
ISP	Internet Service Provider
ITU-T	International Telecommunication Union, Telecommunication standardization sector
LLC	Logical Link Control
LOS	Line of Sight
LPC	Linear Predictive Coding
MACAW	Multiple Access with Collision Avoidance for Wireless medium
MAC	Medium Access Control (layer)
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MOS	Mean Opinion Score
MPC-MLQ	Multipulse LPC with Maximum Likelihood Quantization
NAV	Network Allocation Vector
NI	Network Interface
OFDM	Orthogonal Frequency Division Multiplexing
OLSR	Optimized Link State Routing
PAMS	Perceptual Analysis Measurement System
PAQM	Perceptual Audio Quality Measure
PBX	Private Branch Exchange
PCF	Point Coordination Function
PCM	Pulse-Code Modulation

PC	Point Coordinator
PESQ	Perceptual Evaluation of Speech Quality
PHY	Physical (layer)
PIFS	PCF Inter Frame Spacing
PLCP	Physical Layer Convergence Protocol
PLC	Packet Loss Concealment
PLR	Packet Loss Ratio
PMD	Physical Medium Dependent sublayer
PRN	Pseudo Random Number
PSQM	Perceptual Speech Quality Measure
ROHC	RObust Header Compression
RPE-LTP	Regular Pulse Excited Long Term Prediction Linear Predictive
RTP	Real-time Transport Protocol
RTS	Request To Send
SID	Silence Insertion Descriptor
SIFS	Short Inter Frame Space
SIP	Session Initiation Protocol
SNIR	Signal to Noise and Interference Ratio
TXOP	Transmission Opportunity
UDP	User Datagram Protocol
UPQ	User Perceived Quality
UWB	Ultra-wideband
VAD	Voice Activity Detection
vAP	Virtual Access Point
VSELP	Vector Sum Excited Linear Prediction
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WMN	Wireless Mesh Networks
WPAN	Wireless Personal Area Network

Nomenclature

α	Shape parameter of a truncated Pareto distribution
Δ_i	Cumulative delay faced by a packet generated by a node in the outermost ring ϕ as the packet hops through the rings $\phi, \phi-1, \ldots, i$
η_i	Number of transactions required in a CFP in the ring i of a multihop WLAN using the vAP based medium access mechanism
η_i	Number of transactions to be executed in a CFP by a vAP in the i-th ring of a multihop WLAN
Γ_i	Cumulative packet loss faced by a packet generated by a node in the outermost ring ϕ as the packet hops through the rings $\phi, \phi - 1, \ldots, i$
λ	Packet arrival rate (packet/sec)
λ_i	Effective packet arrival rate of a node at the i-th hop in a multihop WLAN (packet/sec)
Λ_i	Total area under the i-th ring in a multihop WLAN
ν_i	Expected number of contenders in a collision domain at the i-th hop in a multihop WLAN
Ω_i	Node density in the i-th ring of a multihop WLAN
ϕ	Maximum number of hops in multihop WLAN
π_i	Probability of observing exactly i interfering packets in a time slot
ψ	Throughput (bps)
ρ	Queue utilization ratio
τ	Probability that a given node transmits in the channel at a given time slot, i.e., transmission probability
θ	Density parameter for the Exponential node distribution in a multihop WLAN
ζ	Prioritization parameter for the bidding mechanism
A	Advantage factor
b_E	Steady state probability of the system being in the empty queue state
$b_{i,k}$	Steady state probability distribution of any state (i, k)
d_c	Channel access delay in a single hop WLAN
$d_c(\nu_i, \lambda_i)$	Channel access delay in the i-th hop of a multihop WLAN using the DCF based medium access mechanism where there are ν_i nodes in the collision domain and λ_i packet arrival rate per node

$d_c(i)$	Channel access delay in the i-th hop of a multihop WLAN using the vAP based medium access mechanism
d_e	End-to-end delay faced by a voice packet
d_f	Length of a voice frame (sec)
d_i	Delay in the Internet
d_q	Queuing delay in a single hop WLAN
$d_q(\nu_i, \lambda_i)$	Queuing in the i-th hop of a multihop WLAN using the DCF based medium access mechanism where there are ν_i nodes in the collision domain and λ_i packet arrival rate per node
$d_q(i)$	Queuing delay in the i-th hop of a multihop WLAN using the vAP based medium access mechanism
e_c	Channel access loss in a single hop WLAN
$e_c(\nu_i, \lambda_i)$	Channel access loss in the i-th hop of a multihop WLAN using the DCF based medium access mechanism where there are ν_i nodes in the collision domain and λ_i packet arrival rate per node
$e_c(i)$	Channel access loss in the i-th hop of a multihop WLAN using the vAP based medium access mechanism
e_e	End-to-end packet loss faced by a voice packet
e_q	Queuing loss in a single hop WLAN
$e_q(\nu_i, \lambda_i)$	Queuing loss in the i-th hop of a multihop WLAN using the DCF based medium access mechanism where there are ν_i nodes in the collision domain and λ_i packet arrival rate per node
$e_q(i)$	Queuing loss in the i-th hop of a multihop WLAN using the vAP based medium access mechanism
G	Processing gain of a correlation receiver
H	Upper limit of a truncated Pareto distribution
$I_d(d_e)$	Delay impairment caused by an end-to-end delay of d_e
I_{dd}	Impairment caused by too long absolute one way delay
I_{dle}	Voice quality impairments due to listener echo
$I_{e\_eff}(e_e)$	Effective equipment impairment factor for an end-to-end packet loss e_e
l	Expected delay jitter
L	Lower limit of a truncated Pareto distribution
l_d	Payload of a data packet
l_f	Payload of a voice frame (byte)
m	Retry limit of the MAC layer
m'	Number of retry stages in the BEB algorithm
MOS	Voice quality level using the Mean Opinion Score standard
n_a	Aggregation level n_a
n_c	Number of channels in a WLAN
n_s	Number of nodes in the WLAN
n_t	Number of network interfaces per node
n_v	Number of VoIP applications or connections initiated by each node

N_{or}	Room noise at the receiver side (represented as equivalent circuit noise)
N_{os}	Room noise at the sender side (represented as equivalent circuit noise)
OLR	Overall Loudness Raging (OLR)
$P\{i', k'\|i, k\}$	One step transition probability (in our Markov model) from state (i, k) to (i', k')
$P\{X\}$	Probability of an Event X
p_b	Probability of the channel being busy
p_c	Collision probability
p_f	Transmission failure probability
P_i	Probability of exactly i packets being in the queue
p_p	Probability of a capture given that there is a collision
p_p	Capture probability
p_s	Transmission success probability
R	Voice quality rating using the ITU-T E-model
s	Spreading factor of sender's code spreader
s_q	Queue size in number of packets
SLR	Send Loudness Rating
T_a	Absolute one-way delay between the sender and receiver
t_a	Time to transmit an acknowledgment frame (sec)
t_t	Length of a PCF transaction (sec)
t_{bid}	Time taken in a bidding period in a vAP enabled WLAN (sec)
t_{dummy}	Time to transmit a Bidding frame (sec)
t_{idle}	Length of an idle slot (sec)
t_{pause}	Length of a pause time for real human mobility (sec)
t_{recce}	Length of the reconnoiter period (sec)
W	Initial contention window size
$WEPL$	Weighted Echo Path Loss

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Voice over IP (VoIP) allows voice communication using IP (Internet Protocol) networks and has become the most widely used real-time application [6] of the Internet. This technology has received tremendous acceptance from both corporate and home users due to its low cost of service. VoIP based communications do not need a dedicated infrastructure, can utilize an existing IP network and share the same network with different types of other applications. It reduces the costs and human labor associated with deployment and regular maintenance of a dedicated infrastructure. The only cost is associated with the bandwidth usage which is cheap with the ubiquitous availability of the broadband Internet. Due to the low cost of service, giant corporations like Bank of America, Boeing, Ford and Vonage have adopted VoIP [7]. A similar interest among home users already caused a 30% decrease in the US landline home phones [8], and 32.3 million VoIP subscribers are expected by 2011 [9]. Due to the recent global recession, the adoption of VoIP has become a lucrative option for cost cutting [10, 11] in business and private communications.

The use of VoIP in business applications with high call volume can reduce communication costs by a great magnitude. This is why VoIP is widely used in call centers, help desks, corporate Private Branch Exchanges (PBX) and over-the-phone sales. But the traditional means of using VoIP involve initiating the call using a personal computer or a landline phone, where the user needs to be statically located during the call. However, mobility is the key single-most convenience that people are used to now-a-days in voice conversation. This is why, despite the high service cost associated with cellular networks (GSM or CDMA), mobile phones are more popular than landline or VoIP services due to the convenience of mobility. To attain an even greater consumer acceptance than the current market, VoIP service needs to be offered in mobile devices while maintaining the low cost of service.

To offer VoIP services in mobile phones (i.e., in order to support mobility of users), we need to use a low cost wireless carrier to provide the last mile coverage[1]. A number of service providers have attempted to provide VoIP services in mobile phones using GSM or CDMA networks as the carrier, but did not receive substantial customer acceptance due to the high cost of service incurred by the associated cellular networks. The low service cost is the ultimate bargaining chip for VoIP, and a costly solution for the wireless carrier simply compromises the primary appeal to the customers. In this respect, wireless coverage using the IEEE 802.15, 802.11, 802.16 based networks should also be considered. IEEE 802.15 WPAN uses the Bluetooth technology and offers a low data rate ($1 \sim 3$ Mbps) at a short transmission range. IEEE 802.16 WMAN or WiMAX offers a large coverage (50 km) and high data rate (up to 134 Mbps), but the service cost is high due to the use of costly devices and licensed spectrum bands. Therefore, neither the IEEE 802.15 nor 802.16 standards can be used to provide VoIP service in mobile phones with sufficient bandwidth at a low cost. Only the IEEE 802.11 based WLANs (also known as Wi-Fi) offer a moderate data transmission rate (up to 54 Mbps) and coverage (100 m) at a low cost by operating in unlicensed bands and using low cost devices. The ad hoc nature of the medium access mechanism makes the deployment and maintenance of WLANs very easy and keeps operational costs to a minimum. Therefore, the IEEE 802.11 standards hold the true potential for providing VoIP service in mobile devices at a low cost. Moreover, the IEEE 802.11 is the only viable networking solution for temporary emergency wireless networks at places like road-crash sites, public fairs or war-fields. Due to the wide popularity of the Wi-Fi technology, hand held device vendors like Google and Apple included Wi-Fi compatible radio interfaces in their most popular mobile devices including Android, iPhone, iPad and iPod-Touch, etc. The wide spread availability of 802.11 interfaces in mobile devices, PDAs and laptops make the IEEE 802.11 WLANs the most lucrative option for providing wireless last mile coverage for voice calls.

But the available bandwidth in the IEEE 802.11 standards is lower compared to the wired networks, and in the absence of a proper quality assurance mechanism, voice calls suffer from unacceptable quality degradation or jitter. While the Ethernet based LANs offer a 1 Gbps of data rate, the IEEE 802.11 standards offer a theoretical maximum of 54 Mbps. The IEEE 802.11 standards define two medium access mechanisms, namely, Distributed Coordination Function (DCF) and Point Coordination Function (PCF). Public Wi-Fi networks mostly use DCF due to its wider availability than PCF in mobile devices. The backoff mechanism (explained in Section 2.3.4) used in the DCF based medium access control incurs high delay and loss in the medium access that degrade the

[1]The network coverage providing connection between the Internet and the handsets or wireless devices.

voice quality. Under a high traffic load, voice calls get dropped or disconnected which can lead to severe customer dissatisfaction. In this dissertation, we highly emphasize on voice quality since customers will not use VoIP if the call quality is unacceptable [12]. To avoid voice quality degradation in real networks, network administrators need to carefully consider the expected call volume in conjunction with real network aspects during network design and planning. A call capacity estimation model will greatly assist network designers in this regard. Instigated by the growing interest in VoIP over Wi-Fi, many authors investigated techniques to estimate and increase the call capacity of the 802.11 WLANs.

A number of works in the current literature estimated the voice capacity of WLANs using simulation and analyses (a complete survey of such works is presented in Chapter 2). Simulation based investigations are useful in validating analytical models or providing general guidelines, but they offer only limited insight since a simulation based investigation fails to address a network condition for a changed set of parameters. On the other hand, analytical capacity models can address a wide variety of network aspects and can be more useful than simulation based studies in network design and planning since any change in the network conditions can be investigated with great ease and without repeating all simulations for the changed conditions.

But the existing analytical call capacity models for VoIP ove Wi-Fi used oversimplified assumptions which make their findings unreliable. The channel access delay and loss in the DCF based medium access mechanism exhibit non-linear relationships with the offered traffic load and network size. But the analytical models in [13, 14, the most cited works on call capacity models] used unrealistic assumptions like no collisions and no channel error. While approaching the capacity limit, the IEEE WLANs operate in saturated or near-saturated conditions and suffer from frequent collisions. But the models in [13, 14] assumed the nodes to be operating with their initial contention window size and approximated the length of the idle backoff time with the minimum backoff length [14] or a small value derived from simulations in a 2-node WLAN [13]. Moreover, real world environment factors like capture effect and imperfect channel are ignored in the analytical models for DCF based medium access. Power capture can improve the voice performance while an imperfect channel has a severe degrading effect on it; and therefore, both should be incorporated in a call capacity model. Most importantly, standard voice quality measures were ignored in all existing capacity models and either the delay [14–17] or loss [15, a simulation based study] were used to determine the upper limit of call capacity. Voice quality is impaired by the combined effect of delay and loss [18] which was not considered in any of these studies. Real networks operate under the unsaturated condition for a considerable amount of time [19], but the voice quality under the unsaturated condition was also

never investigated. The existing analytical call capacity models for the PCF based medium access mechanism also suffer from similar pitfalls as they ignored standard voice quality measures and either considered ideal channel only [17, 20] or inadequately analyzed the effect of imperfect channels using unrealistic assumptions [3].

As mentioned before, the available bandwidth in the IEEE 802.11 WLANs is low which limits the voice call capacity. Therefore, alternative methods should be investigated to increase the call capacity in such networks. The IEEE 802.11 standards allow the simultaneous use of multiple channels which can increase call capacity substantially. The use of multihop WLANs offers a wide wireless coverage, and additionally allows spatial reuse which can improve the performance of voice traffic. Future mobile devices are expected to bear multiple network interfaces [21] which will also affect the voice quality. But no existing analytical model considered the impacts of multiple channels, network interfaces or multihop routes on the voice call capacity.

Existing investigations on the performance analysis of the IEEE 802.11 WLANs revealed that the high delay and loss in channel access associated with the DCF based medium access can be reduced significantly by using the PCF based medium access. But the PCF based medium access is strictly limited to single hop WLANs whereas DCF can be used over multiple hops offering a wider coverage. The best features of these two mechanisms, i.e., the wider coverage of DCF and the lower delay and loss of PCF, can be achieved concomitantly if some of the client nodes volunteer to coordinate the time synchronized medium access mechanism of PCF over multiple hops. In the DCF mechanism also, the intermediate client nodes relay packets generated by or addressed to nodes which are located far from the access point. A similar approach can be utilized to enable the use of PCF over multiple hops. But the existing IEEE 802.11 standards do not offer such a mechanism which has the potential to increase voice capacity significantly. Despite the considerable number of approaches (outlined in Section. 2.6) in the current literature to enhance the voice performance in WLANs, the time synchronized medium access is still not operable over multiple hops which must be investigated.

In order to offer VoIP service in mobile devices at a low cost while maintaining an acceptable voice quality, we need to overcome a number of research challenges. In the following, these challenging issues are identified and the problem statement for this research is formulated.

1.1 Problem Statement and Motivations

In order to ensure voice quality in the IEEE 802.11 WLANs, voice quality guidelines need to be carefully consulted in network design and planning. In this regard, an

analytical call capacity model will be of great assistance if it considers voice quality requirements and all realistic network aspects that impair or improve the voice quality. The existing analytical models (including [13, 14, 22]) ignored standard voice quality measures and used either delay or loss only to define the upper limit of voice quality, although voice quality is impaired by the combined effect of both delay and loss. Ensuring voice quality is of utmost importance since customers will not use VoIP over Wi-Fi if the voice qulality is unsatisfactory. The most popular voice quality estimation methods are the Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ) and ITU-T E-model. Both MOS and PESQ require an existing network to be used while the E-model uses codec characteristics and various delay and loss elements experienced by voice packets to estimate voice quality. The E-model can be utilized to consider network performance measures like the delay and loss in medium access and queue to formulate a call capacity model for IEEE 802.11 WLANs.

To estimate the call capacity of a WLAN accurately, it is crucial to determine the exact delay and loss in the medium access and queue. Despite the complex nature of the 802.11 DCF mechanism, over-simplified assumptions were used in the existing works which make their findings unreliable. To provide precise guidelines on voice capacity, the inherent characteristics of the medium access mechanism and their impact on voice quality should be investigated carefully. An imperfect channel causes transmission errors and incurs severe degradation in the voice quality. On the other hand, power capture can reduce transmission failures from frequent collisions and can bring a positive impact on the voice performance. To model a real scenario closely, both an imperfect channel and the capture effect should be incorporated in a capacity model. Real world WLANs operate in both saturated and unsaturated conditions and voice quality under both conditions need to be investigated. It is identified from simulations in our study that the delay and loss in the queue affect the voice quality severely which should also be taken into account. An analytical call capacity model can be formulated using the E-model to ensure acceptable voice quality while considering all the above factors concomitantly to determine the delay and loss in the medium access and the queue.

Except for a few simulation based approaches that investigated the voice performance over multiple hops, there is no analytical call capacity model in the current literature for multihop, multi-channel WLANs. The intermediate nodes in a multihop WLAN forward (or relay) data packets to and from the nodes located far from the AP. As a result, these intermediate nodes experience a higher traffic arrival rate than nodes located in the outer regions and suffer from higher delay and loss in the channel access. The high channel access delay incurs a high delay and loss also in the queue, and should be considered in the capacity modeling. Future mobile devices

are expected to bear multiple network interfaces which allow a higher service rate for the queue of a WLAN node at the cost of increased contention in the network. A call capacity model can be formulated addressing the changed traffic condition in a multihop, multi-channel WLAN with multi-interface nodes. Additionally, public networks suffer from the formation of hotspots since the client nodes try to be nearer to the access point. As a result, the node distribution becomes non-uniform and the contention for channel access increases at the nearest vicinity of the access point. The formation of such hotspots degrades the network performance and, as a result, the call capacity decreases. The analytical model for a multihop WLAN should determine the impact of such hotspots in fine details so that real networks can avoid any call quality degradation in these circumstances.

The DCF based medium access uses a backoff mechanism to reduce collisions and adapt to changes in the traffic load, but incurs a high delay in the medium access. On the other hand, in the IEEE PCF based medium access mechanism, collisions do not occur since access to the channel is coordinated by an access point (AP). Therefore, backoff is no longer needed, and the delay in channel access decreases. Additionally, the high packet loss due to collisions in a dense network can also be reduced if the PCF based medium access is used. In the DCF based medium access, the AP needs to contend for channel access with the same priority as the client nodes. But since the AP is the only node with Internet connectivity, all voice packets must pass through it. Therefore, the AP experiences a much higher traffic load than the client nodes and an unbalanced traffic condition occurs. The PCF based medium access ensures that the same number of uplink (from the client nodes to the AP) and downlink (from the AP to the client nodes) frames are transmitted in a given time. Due to these reasons, a higher call capacity is expected with the PCF based medium access. Most of the existing call capacity models for PCF mechanism [17, 20] considered ideal channels only and ignored standard voice quality measures. Although the effect of imperfect channels was considered in [3], the delay in channel access was incorrectly estimated by ignoring the difference in the channel time usage for uplink and downlink transmissions. The delay and loss in the queue and the voice quality requirements were also ignored in [3]. Therefore, an analytical model for the call capacity with PCF based medium access can be designed by considering standard voice quality requirements and specific impacts of channel errors in the uplink and downlink transmissions. The delay and loss in the queue should also be incorporated to allow realistic call capacity estimation.

Despite the high delay and loss in the channel access, the use of DCF is beneficial for Wi-Fi networks in public places since it offers a wide coverage through the use of multihop WLANs, and allows spatial reuse. On the other hand, PCF offers a lower delay in the channel access than DCF and reduces packet loss due to collisions. But

the use of PCF is strictly limited to single hop WLANs (i.e., low coverage) since an AP is required to coordinate channel access. DCF is operable over multiple hops because its ad hoc nature of operation allows the nodes to access the channel without any coordination by an AP, and intermediate nodes relay packets to and from the nodes located far from the AP. A similar approach can be used for PCF-like time synchronized medium access where the client nodes coordinate channel access in a multihop WLAN whenever a regular AP is unavailable. Since the IEEE 802.11 standards do not offer such a mechanism, a novel approach is required to design the protocols needed to establish one node among all client nodes as the virtual access point (vAP) at each hop. Therefore, a hybrid mechanism needs to be designed so that the vAP selection mechanism can be executed on an ad hoc basis (like DCF); and once a client node is designated as the vAP, it can coordinate the channel access using the PCF mechanism. If two such vAPs are located within each other's transmission range, frequent collisions can occur since carrier sensing mechanisms are not used in PCF based medium access mechanism and therefore, additional mechanisms should be carefully designed to avoid such scenarios. The client nodes are expected to be mobile but if the vAP moves away from its current location, the ongoing transmissions are affected which will have a drastic impact on the voice quality. Therefore, the impact of mobility on such an access mechanism should be carefully analyzed considering the real world human mobility patterns. Finally, the possible call capacity increase by employing such a mechanism should be investigated considering the voice quality requirements.

1.2 Research Objectives

The aim of this thesis is to assist in network design and planning so that an acceptable voice quality can be maintained in the IEEE 802.11 WLANs. Inspired by the research opportunities discussed in the previous section, the following key research objectives are formulated in order to identify the network factors which degrade voice quality, incorporate standard quality measures in call capacity estimation, determine the potential call capacity of a given WLAN with a choice of network aspects and formulate a scheme to increase the voice capacity of real world WLANs. Specifically, the primary objectives are as follows.

1. Formulate an analytical call capacity estimation model for a single hop, single channel WLAN employing the DCF based medium access mechanism using standard voice quality assessment method and considering realistic traffic model and real-world environment factors including imperfect channel and capture effect.

2. Extend the above call capacity model to a multihop, multi-channel scenario considering the cumulative packet arrivals over multiple hops and the collision domain size. Additionally, consider the impacts of multiple network interfaces and non-uniform node distribution on the voice capacity.

3. Develop a call capacity model for PCF based WLANs considering the impacts of imperfect channels and the performance of the queue on the voice capacity, and determine the possible capacity increase using the PCF based medium access mechanism compared to DCF.

4. Formulate a novel medium access scheme to allow the PCF based time synchronized medium access mechanism over multiple hops so that the wider coverage and spatial reuse of DCF as well as the lower delay and loss of PCF can be attained. Determine the possible voice capacity increase, compared to the DCF based medium access mechanism, in multihop WLANs.

1.3 Overview of Contributions

In order to accomplish the above research objectives, this dissertation presents a number of original contributions. The contributions are illustrated through the schematic diagram shown in Fig. 1.1. The contributory works are presented as blocks (labeled from A to D) and an arrow from Block X to Block Y indicates that the contributory work in Block Y is inspired by the findings in Block X and some of its core concepts are also used in Block Y. Block A in Fig. 1.1 represents a call capacity estimation model for DCF based single hop, single channel WLAN considering the voice quality guidelines recommended by ITU-T, and addresses Objective 1. The voice quality requirement, determined in Block A, is also used in every other work (Blocks B $\sim$ D) in this dissertation. Block B uses the model of DCF based medium access for a single hop WLAN introduced in Block A and incorporates additional considerations to model the call capacity of a DCF based multihop, multi-channel WLAN and thereby, addresses Objective 2. The high delay and loss suffered by the voice packets in the DCF based medium access inspired the investigation of the call capacity with PCF based medium access which is denoted by Block C and accomplishes Objective 3. The capacity increase using the PCF based medium access (in Block C) inspired Block D where a medium access control mechanism is proposed to allow the time synchronized medium access mechanism over multiple hops and addresses Objective 4. The key contributions of this thesis are as follows.

1. A novel call capacity estimation model is introduced for single channel, single hop WLANs considering the voice quality impairments as dictated by the ITU-T

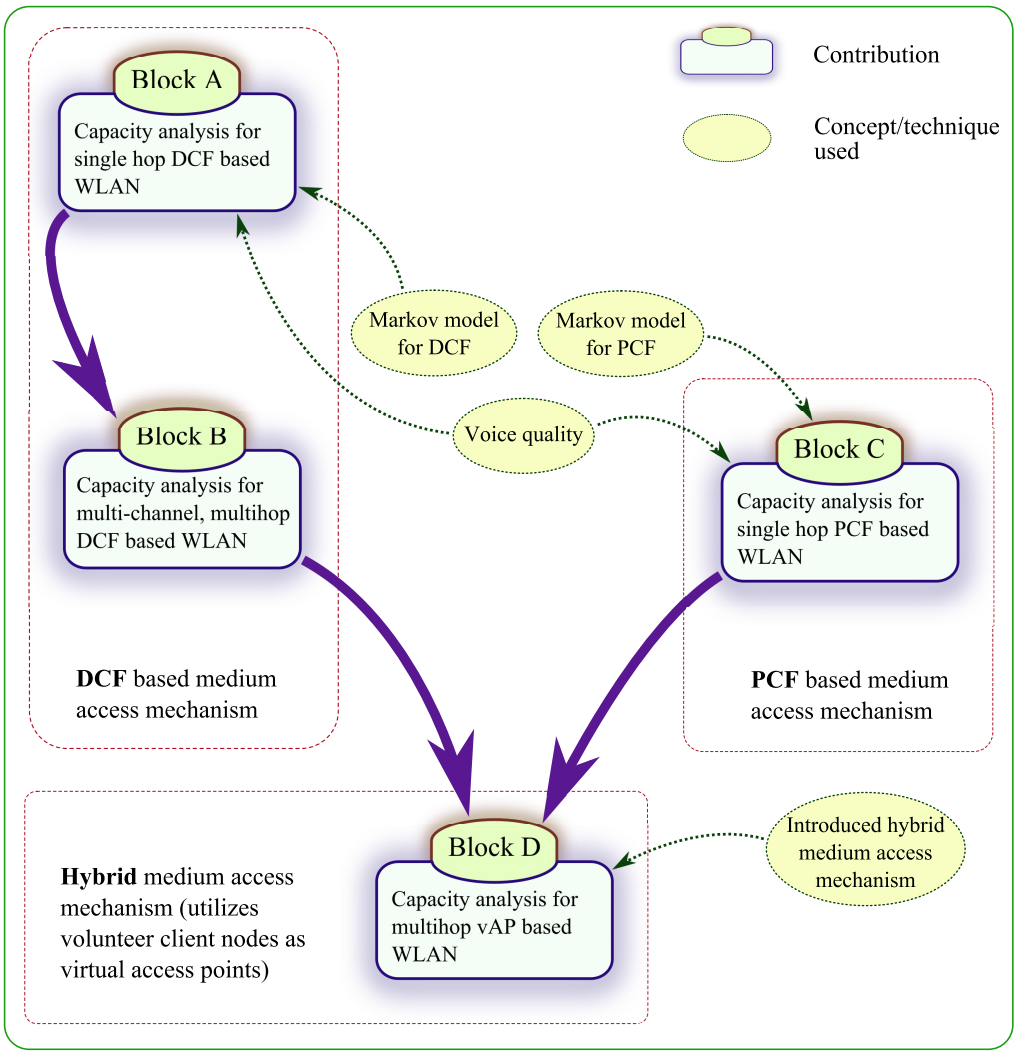


Figure 1.1: A schematic diagram illustrating the major contributions (Blocks A $\sim$ D) in this dissertation.

E-model. The E-model is employed to determine the voice quality impairments due to the end-to-end delay and loss, and an impairment budget for voice quality is formulated. The estimated call capacity is presented in the form of an optimization problem which uses the impairment budget as a limiting condition so that the WLANs (designed using our model) are sufficiently provisioned to serve the expected call volume while maintaining an acceptable voice quality. We identify from simulations in our study that the delay and loss in the queue degrade the voice quality severely, which is ignored in the existing call capacity models. We use a $M/M/1/s_q$ system to model a finite capacity queue to reflect its effect on the voice performance. This work has been previously published by Siddique and Kamruzzaman [23].

2. The IEEE 802.11 DCF mechanism suffers from frequent collisions, especially under a high traffic load. Such frequent collisions result in high delay and loss in the medium access which degrade the voice quality severely. To accurately assess the delay and loss in a practical environment, all real-world network aspects must be incorporated in the MAC layer model, which the current literature lacks. To address this, a Markov model is introduced to determine the delay and loss in the channel access considering real world factors including imperfect channel and capture effect. Real networks operate under the unsaturated condition for a considerable amount of time [19] which is also incorporated in our Markov model. The use of the proposed Markov model in the call capacity estimation models allows a precise determination of voice quality by estimating the exact delay and loss in channel access. Extensive simulations under a wide variety of conditions and network size validate the proposed Markov model. This Markov model has been previously published by Siddique and Kamruzzaman [24].

3. The IEEE 802.11 standards allow simultaneous use of multiple channels which can increase call capacity significantly. Future mobile devices are expected to bear multiple network interfaces and reduce the network outage problem. The impact of multiple channels and network interfaces should be carefully considered. We use a multi-server, finite capacity queuing system to model the queue of a multi-interface node and formulate a call capacity estimation model for multi-channel WLANs with multi-interface nodes. This model has been published by Siddique and Kamruzzaman [25, won best paper award in IEEE WCNC 2010]. Multihop WLANs are now commonplace and can improve the voice performance through spatial reuse. Moreover, a wide coverage can be attained by a single AP in a multihop scenario where the inner client nodes forward packets to and from the nodes located in the outer region. But the cumulative packet arrivals give rise to the formation of a critical zone. Additionally, as client nodes try to be nearer to the AP, a non-uniform node distribution can be expected. Both of these factors can create a bottleneck and degrade the voice capacity. In this dissertation, a call capacity model is proposed considering the effects of multiple channels and interfaces in a multihop WLAN considering all the above factors concomitantly. This work has been previously published by Siddique and Kamruzzaman [26].

4. The existing call capacity models of the PCF based medium access mechanism either considered an ideal channel only or incorrectly estimated the delay and loss in the channel access due to the inappropriate assumptions on channel time wastage in an imperfect channel (i.e., the channel time wastage is lower if errors

occur in a downlink frame instead of an uplink frame which was ignored). The call capacity estimation becomes unusable due to such inadequate analyses. Additionally, no analytical call capacity model of PCF based WLANs considered the voice quality requirements using an adequate method. Such a method requires an accurate estimation of the delay and loss in the PCF based medium access. A Markov model is proposed in this dissertation to model the PCF based medium access mechanism in order to determine the delay and loss in channel access. This Markov model has been previously published by Siddique and Kamruzzaman [27].

5. DCF based medium access incurs a high delay and loss in channel access which can be reduced considerably by the use of PCF based medium access. A call capacity model for the PCF based medium access mechanism is developed to determine the VoIP call capacity in a PCF based WLAN with consideration for imperfect channels. Theoretical analyses as well as extensive simulation results demonstrate that a higher call capacity can be attained by using the PCF based medium access mechanism than DCF. This work has been published by Siddique and Kamruzzaman [28].

6. DCF based multihop WLANs offer a wide coverage and allow spatial reuse but suffer from a high delay and loss in the medium access and queue. On the other hand, PCF offers a low delay and loss in the medium access and queue but its use is strictly limited to single hop WLANs only. In order to increase the voice capacity while providing a wide wireless coverage, a novel hybrid medium access scheme is proposed to allow the time synchronized medium access mechanism of PCF over multiple hops through the introduction of the "virtual access point" concept. The effect of human mobility on the proposed scheme is analyzed and the call capacity increase by the use of the proposed scheme, compared to DCF, in multihop WLANs is also demonstrated. The proposed scheme has been published by Siddique and Kamruzzaman [29].

1.4 Structure of this Thesis

This dissertation is organized as follows—

Chapter 2 presents an overview of the Voice over Internet Protocol (VoIP) mechanism and its applications, and establishes its importance in corporate and home communications. The voice network architecture is explained and the most popular voice codecs are reviewed. The voice quality is the primary focus of this

thesis and therefore, voice quality estimation methods are discussed and their advantages and limitations are identified. Among the various voice quality estimation models, the use of E-model is found to be the most appropriate for network design and planning and some of its real world applications are reviewed. To offer VoIP service in mobile devices, we need to use a low-cost carrier to provide the last mile coverage. IEEE 802.11 WLANs are identified as the most potential alternative to offer such wireless coverage. The medium access mechanisms defined in the IEEE 802.11 standards are reviewed and the carrier sensing mechanisms are explained. A precise determination of the delay and loss in the 802.11 medium access mechanisms is of utmost importance in determining voice quality. The most cited performance analysis works are discussed followed by a review of the existing call capacity estimation models. The challenges in voice over WLANs are identified and the existing works which attempted to increase the voice capacity in 802.11 WLANs are reviewed.

Channel 3 uses a top down approach to design a call capacity estimation model for a DCF based single hop WLAN by determining the voice quality impairments using the ITU-T E-model. An impairment budget is defined in terms of the network performance measures, i.e., the end-to-end delay and loss, codec characteristics and dejitter buffer performance. To determine the delay and loss in the medium access, a Markov model is developed which considers the real world factors including an imperfect channel and the capture effect. Both saturated and unsaturated traffic conditions are considered and, to model the delay and loss in the queue, a $M/M/1/s_q$ system is used. The call capacity is presented as an optimization problem and is analyzed for codec parameters and different network conditions. Extensive simulations are used which validate the proposed model.

Channel 4 extends the single hop WLAN call capacity model of Chapter 3 by considering the effects of multiple channels, network interfaces and multihop routes. In a multihop WLAN, the intermediate nodes relay traffic on behalf of the nodes in the outer regions, and suffer from an accumulated traffic load. The collision domain size and the effective traffic arrival rate at each intermediate hop are determined. Additionally, a $M/M/n_t/s_q$ system is used to model the shared queue of a multi-interface node. The end-to-end delay and loss are defined and the necessary conditions to avoid the formation of bottlenecks are derived. The call capacity of such a network is presented as an optimization problem. The node distribution affects the voice quality and network performance in a multihop WLAN which is also considered.

Channel 5 illustrates the limitations of the existing models of PCF based communication using imperfect channels and establishes the ground to model the PCF based medium access from the scratch using a novel Markov model. The delay and loss in the medium access are determined using the proposed Markov model and the queuing system proposed in Chapter 3 is used to estimate the delay and loss in the queue. The call capacity with the PCF based medium access is presented and compared to that with the DCF based medium access.

Channel 6 combines the best features of the DCF and PCF based medium access mechanisms and proposes a hybrid medium access scheme in order to improve the voice capacity of multihop WLANs. The proposed medium access scheme allows time synchronized medium access mechanism to be used over multiple hops. Necessary protocols to address different network conditions are described in detail and the effect of node mobility on the proposed scheme is analyzed. Using the logical argument and analytical methods used in Chapter 4, the call capacity using the proposed mechanism in a multihop WLAN is determined and compared to that in an equivalent DCF based WLAN.

Chapter 7 presents concluding remarks of this dissertation and outlines possible future research directions.

Chapter 2

VoIP over Wireless Local Area Networks

2.1 VoIP: Voice over Internet Protocol

VoIP is the most widely used real-time application [6] in the Internet and is becoming increasingly more popular every day. Both corporate and home users have embraced this technology leading to its quick growth. VoIP can offer voice communication at a very low price which has become a lucrative option for private users due to the recent global recession [10, 11, 30]. Furthermore, the globalization of businesses and industrial markets in the last decade has heightened competition among companies and led to rigorous cost cutting in firms of all sizes. But the communication requirement is always increasing. Due to the global nature of many businesses, international calls have become an expensive everyday necessity. This is why companies are increasingly adopting VoIP to reduce the cost of local and international voice communication. Some examples of VoIP adoption in businesses are given here [7]—

- Bank of America[1] is deploying more than 180,000 Cisco VoIP phones;
- Boeing[2] is planning to equip its 150,000 workers with VoIP;
- Ford[3] is going to deploy 50,000 VoIP phones;
- Vonage[4] has 600,000 customers and getting new subscribers at the rate of 15,000 per week;

[1] www.bankofamerica.com
[2] www.boeing.com
[3] www.ford.com
[4] www.vonage.com

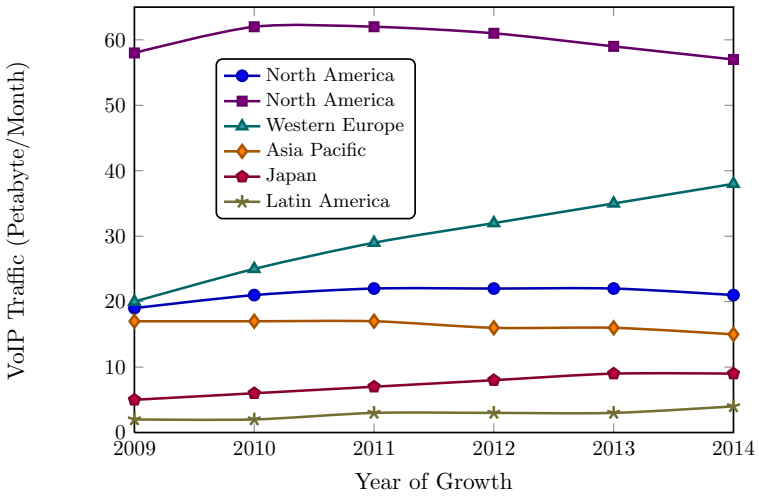


Figure 2.1: Expected growth of VoIP market [1].

- British Telecommunications plc (BT)[5] announced that it will convert its infrastructure to VoIP;
- Monash University[6] converted all PSTN phones to VoIP.

The expected growth of the subscriber base, a large portion of which are private users, is also astonishing which can be found in the following forecasts—

- The VoIP market is expected to reach 32.3 million subscribers by 2011 [9];
- Mobile VoIP users are expected to reach 107 million by 2012 [31];
- Only mobile HD voice market will reach 487 million users by 2015 [32];
- VoIP has already caused a 30% decrease in U.S. landline home phone use [8];
- VoIP penetration in U.S. businesses is expected to reach 79% by 2013 [33];
- VoIP market profit is expected to reach $3.3 billion by 2010 [34].

Skype's acquisition by Ebay at $4.1 billion [35] following Microsoft's purchase of Teleo at an undisclosed amount [36] also clearly indicate the market's expectation from this technology. A forecast [1] by Cisco Systems, Inc. suggests that by 2014 VoIP traffic in the Internet will reach 21, 57 and 38 Peta byte per month in North America, Western Europe and Asia Pacific, respectively. The expected growth of VoIP traffic volume in the Internet according to the forecast in [1] is shown in Fig. 2.1.

[5]www.bt.com
[6]www.monash.edu

2.1.1 Importance of VoIP

Voice service cost is primarily associated with fixed deployment and regular maintenance. VoIP does not require any dedicated infrastructure and a major advantage of using VoIP is the cost savings by hardware reduction and the elimination of the service charges and toll costs associated with regular fixed-line telephony like PBX and PSTN. VoIP can be used over any existing IP network and can share the same network with other applications. Deployment usually only requires installation of VoIP phones, if they are used, although personal computers can be used instead to further reduce the installation cost. Sometimes central directory and call monitoring services are employed in order to improve voice performance and the users' convenience. Such services are usually very cheap and can be co-located with other services and hosted on any existing server. Corporate networks are usually maintained for data services anyway and no separate maintenance is required for voice calls. In this way, VoIP can minimize the cost of deployment, operations and maintenance for intra-company and inter-company (external) communications.

VoIP is a key factor in IP convergence where a single high speed Internet connection can be used for all kinds of business communication services including voice calls, teleconferencing, email, pager, etc. Convergence is another reason behind corporate interest in VoIP. A converged network eliminates the need for duplicate hardware and special vendors to manage separate voice networks and Private Branch Exchanges (PBX) [37]. Convergence allows voice, video and data to be used in a combined fashion using a single device and a single network infrastructure. Thus, a high user interactivity can be achieved at a much lower cost through convergence. A recent survey showed that 36% of small business organizations use VoIP over Wi-Fi for their internal communications [38].

Household Internet users have embraced VoIP with equal enthusiasm. With a growth rate of 88%, residential broadband users are expected to reach 121 million in western Europe only [39]. About 73% of these users will use VoIP giving an estimate of 87.8 million home users who will start using VoIP by 2012. A growth rate of 240% for household VoIP users with 50% household penetration is also reported in [39]. A VoIP penetration of 20% in U.S. household users is expected in [40]. The number of Internet users is a good indication of IP telephony use since most broadband users also use the VoIP services that comes without incurring any additional cost. Latest statistics in [41] shows that 28.7% of world population (6.845 billion) is using the Internet. Although, most users are from Asia (42%) and Europe (24.2%), the user growth rate is highest in North America (77.4%), Oceania (61.3%) and Europe (58.4%).

2.1.2 Applications of VoIP

The ability of cheap verbal communication and the widespread penetration of the broadband Internet are the major attractions of VoIP. A number of potential VoIP applications are listed in this section.

Call Center and Help desk Many companies offer a help desk to assist their customers. For instance, computer and laptop services, mobile phone operators and Internet Service Providers (ISP) all offer some "Helpline" telephone numbers where customers can call to seek assistance. The volume of the calls to these numbers is very high and the use of VoIP for these services can reduce production/service cost.

Sale over phone Today producers of consumer goods are adopting over the phone sales where operators call potential customers and offer them products such as life insurance policies, utility connections, holiday accommodations or fitness equipments. This sales method reduces the costs involved in renting an office, paying salary to a high number of employees and the maintenance of office equipments. Moreover, semi-interested customers who might not visit a shop or are not looking seriously for a product can be reached. The problem with this method is that it requires a large number of phone calls where VoIP can be used to reduce the operational costs and save a fortune.

Corporate PBX Most organizations provide phones to their employees to increase interactivity and lessen process delay so that a high productivity can be achieved. But the use of a PSTN phone for every employee is expensive and traditionally a Private Branch Exchange (PBX) is used to minimize the PSTN line requirement. But PBX needs a separate, specifically built infrastructure which incurs an additional cost. VoIP based PBX can utilize a corporate LAN and keep the communication cost to a minimum.

Home Phone Applications The traditional PSTN connections used in home applications are now largely being overtaken by VoIP connections. The overwhelming interest of home users in VoIP based packages is primarily due to the low service cost. Moreover, VoIP based connections also support the simultaneous use of multiple devices, e.g., PDA, laptop, mobile phone and wireless handset, etc., with a single connection. With an appropriate configuration, all such connected devices can start ringing on an incoming call which offers the convenience of choosing the appropriate device.

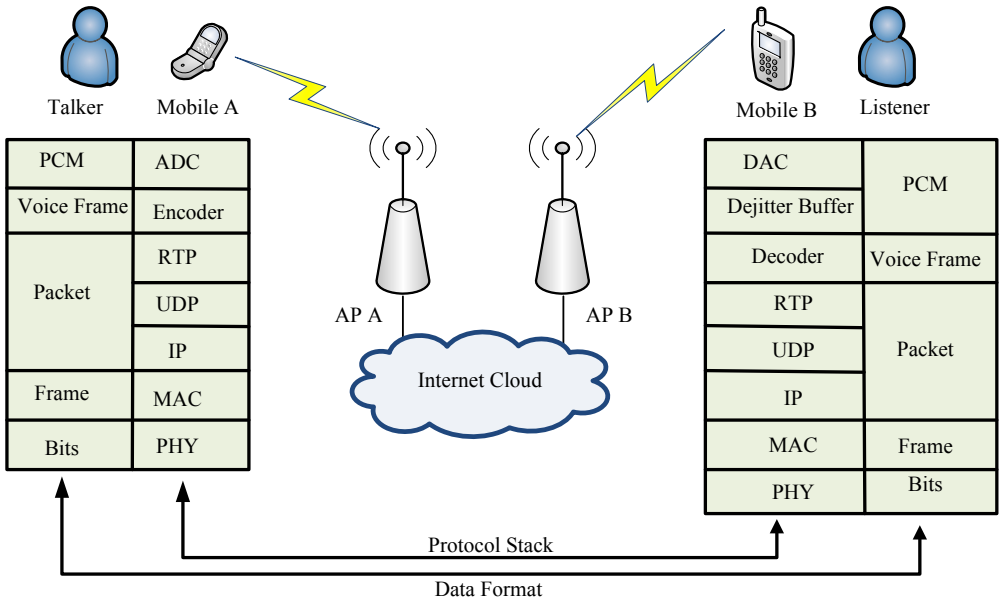


Figure 2.2: A typical VoIP system with different protocol layers and data units.

On-call Paging On-call staffs and patients at hospitals and nursing homes frequently need to contact each other. Considering the volume of such communications, a low-cost solution would be highly beneficial. VoIP can be used in such scenarios to provide a reliable voice communication system at a low cost of service.

2.1.3 Voice Network Architecture

To analyze the capacity and performance of VoIP over wireless networks, we must understand the network architecture and different components involved in the end-to-end communication. Each voice call can be modeled as a set of voice streams or connections. A traditional two way VoIP call comprises two voice streams while a multi-peer communication like teleconferencing usually involves a number of such voice streams. Each stream uses a one way IP connection where a talker uses a sender module to send voice packets and the listener uses a receiver module. VoIP uses the IP network to communicate, and hence a TCP/IP protocol hierarchy is used at both the sender and receiver. The components and protocol hierarchy in a VoIP system are shown in Fig 2.2.

The sender uses an Analog to Digital Converter (ADC) to capture an analog voice signal as Pulse-Code Modulation (PCM) data or digital data. The PCM data is stored in chunks which are called voice frames. The sender application uses a voice encoder or codec which reads the voice frames and encodes them. The encoding process involves data compression while maintaining key voice features and can also include

encryption if security protocols are implemented. There are a number of commercial codecs which use different compression mechanisms and differ in their compression efficiency and bandwidth requirements. Most of these codecs are designed by the International Telecommunication Union, Telecommunication Standardization Sector (ITU-T). We discuss a few of the most popular commercial codecs in Section 2.1.4. A packetizer puts a number of encoded voice frames in a User Datagram Protocol (UDP) packet that is to be transmitted to the receiver. The number of voice frames put into a single UDP packet is called the aggregation level. Packetization is viewed as a separate process from encoding by some authors but, for clarity, we use the term encoder to denote the module that performs both the encoding and packetization.

The sender module maintains two sessions with the receiver for the duration of the call. A session management protocol is used to communicate control information and negotiate codec parameters while a voice session is maintained to send the UDP packets. Session creation, modification and termination are the responsibilities of the session management system. Session Initiation Protocol (SIP)[42] is most widely used to manage both voice and video sessions. SIP is an application layer protocol designed to be independent of the underlying transport layer protocols, e.g., TCP or UDP. TCP provides more reliable communication than UDP but each TCP packet must be acknowledged which incurs an additional transmission overhead. Reliable communication is crucial for the session management while the voice session can tolerate low packet loss. Therefore, TCP is used for the session management while UDP is used for the voice traffic. However, UDP does not provide any packet sequencing which is needed to put the voice frames in their original sequence at the receiving end. Therefore, Real-time Transport Protocol (RTP) is used on top of UDP to provide packet sequencing. VoIP is indifferent to any of the lower layer protocols, for example, Routing, Logical Link Control (LLC), Medium Access Control (MAC) or Physical layers. This indifference allows VoIP to be used over any type of network that supports IP packets. With the ubiquity of IP networks, VoIP can be used in almost any existing network which leads to a ubiquity of VoIP as well.

Voice packets are needed at regular intervals at the receiving end to play out an uninterrupted voice signal. But the timely arrival of voice packets can not be guaranteed. There is always some variation or jitter in the end-to-end delay of packet delivery. To comfort this, the received packets are put in a dejitter buffer which helps in reducing the delay jitter of received packets. A dejitter buffer can either have a static or dynamic length which is discussed in detail in Section 3.5. The decoder picks the voice packets from the dejitter buffer and reconstructs the voice as PCM data. A Digital to Analog Converter (DAC) then plays out the voice. The decoder may implement Packet Loss Concealment (PLC) mechanism which produces dummy voice

frames in replacement of the lost packets. Most simple PLC mechanisms create silence or noise frames which are either predefined or background sounds received from the sender. More sophisticated PLC techniques create sound forms that are convenient to the listener. For example, frame interpolation or decoder state information can be used to create signals that match the sounds surrounding the lost portion. However, the packet loss concealment technique can only cover a small amount of random (not bursty[7]) packet loss.

A number of commercial solutions defined proprietary standards for VoIP network architecture. Among them Cisco's Architecture for Voice, Video and Integrated Data (AVVID) [43] and AT&T's Common VoIP Architecture are the most popular in the industry. These standards define the different components in the network including the sender, receiver, router, bridge and repeater, etc. The standards define the particular type and performance requirements for all devices in order to maintain voice quality. However, due to the indifference of VoIP to the protocol layers lower to IP, the key parameters defining the performance of VoIP are primarily associated with the configuration of the codec and underlying network. We discuss some of the most widely used codecs in the next section.

2.1.4 Voice Compression Codecs

Codecs are used to compress and sometimes encrypt the voice signal. Each codec uses a different compression technique which results in a different voice quality level, computational complexity, robustness to packet loss and bandwidth requirements. Voice codecs can be divided into two broad classes which are the "waveform coders" and model based "speech coders". A waveform encoder encodes the voice data frame by frame, thus the encoded voice and the original analog signal can be mapped to each other in a deterministic fashion. At the decoder, the original speech is reproduced frame by frame in a similar manner. The simple Pulse-Code Modulation accomplishes these tasks by quantizing each speech sample to one of a fixed number of levels. The rate of sampling the voice and the number of quantization levels define the bandwidth requirement. Assuming 256 levels of sampling a signal at 8 kHz, the resulting encoded data stream is generated at 64 Kbps. Adaptive Differential PCM (ADPCM) uses predictions with differential quantization to reduce the bandwidth requirement. The quantization process incurs distortion in the decoded signal which can be reduced by sampling at a higher rate, but then it requires a higher bandwidth to be transmitted to the receiver module. The computational complexity of the waveform coders are usually low. Early Digital Signal Processing (DSP) devices mostly adopted waveform

[7]Packet loss occurred in bursts.

coders for their low computational complexity and their ability to maintain a high level of voice quality.

On the other hand, model based speech coders use a parametric model to approximate short segments of speech. Model parameters are determined for each speech segment which are used at the receiving end to synthesize a speech signal that is perceptually close to the original one. Instead of recreating the original signal, only the perceptual contents of the speech are approximated. The approximations allow the coder to operate at much lower data rate than waveform coders. Despite the low bandwidth requirements of the model based coders, the waveform coders are more popular for their better quality voice.

Voice codecs can be categorized into three categories according to their bandwidth requirements [44]. Narrowband codecs operate on an audio signal filtered to a frequency range of 300–3400 Hz and sampled at 8 kHz. These include G.711, G.729, G.729A–E, G.723.1, GSM-FR, GSM-HR, GSM-EFR, iLBC, etc. Broadband codecs operate on voice signals filtered to 50–7000 Hz and sampled at 16 kHz. G.722, G.722.1, G.722.2 and G.729.1 are examples of broadband codecs. Multimode codecs can operate on either the narrowband mode or broadband mode. Speex and BroadVoice are examples of such codecs. Due to the low bandwidth availability in IEEE 802.11 WLANs, most works in the literature considered narrowband codecs only. In fact, most authors used G.711 which provides a moderately high voice quality while some works additionally considered G.729, G.723.1 or G.726.

A number of works focused on reducing the data rate of waveform coders while maintaining an adequate voice quality. One such technique is Code-Excited Linear Prediction (CELP) where vector quantization is combined with linear prediction. This technique uses some concepts from the model based coders. CELP uses a vector quantization dictionary as a codebook which can be algebraic (e.g., ACELP) or explicit (e.g., SPEEX). For instance, an all-pole model can be used to approximate the speech spectrum and a long-term predictor can be used to represent the pitch or local periodicity of the speech signal. An all-pole filter is used because it is a good representation of the human vocal tract and is easy to compute [45].

We discuss some of the most popular commercial codecs in this section.

2.1.4.1 G.729

The G.729 [46] codec is a Conjugate Structure Algebraic Code-Excited Linear Prediction model (CS-ACELP) that was designed by ITU-T for low bit rate communication. It is based on the popular Code-Excited Linear Prediction coding model. G.729 offers

toll quality[8] calls while consuming only 8 Kbps which is one fourth of the 32 Kbps acquired by the ADPCM technique although both yield the same call quality level. G.729 uses voice frames of 10 ms length only, which ensures a low packetization interval. Each voice frame is encoded into 10 bytes of payload and therefore, the bandwidth consumption is also very low. The encoder reads the following voice frame to compress the current frame with a higher efficiency which incurs a look ahead delay of 5 ms. The computational complexity of the encoding process is 20-25 MIPS and the memory consumption is less than 4 Kbps.

Despite the good voice quality, the computation complexity was deemed too high for some early devices. That is why the G.729 Annex A (G.729A) was developed which uses the Digital Simultaneous Voice and Data (DSVD) technique and is a low-complexity version of the G.729 standard with a slightly poorer voice quality and inter-operable methods, i.e., the voice frames encoded using G.729 can be decoded with G.729A and vice versa. Although G.729A has a slightly lower voice quality, it offers the best computational complexity to voice quality ratio in the industry.

G.729 Annex B (G.729B) was designed to further reduce the bandwidth usage by utilizing silence suppression. Silence suppression means suppressing transmissions during a silent period of the talker. To accomplish this, the G.729B utilizes three techniques which are Voice Activity Detection (VAD), Comfort Noise Generation (CNG) and Discontinuous Transmission (DTX). The VAD module determines the active and silent states of the talker and transmits packets only when the talker is saying something. Whenever the talker is silent, no packets are generated or transmitted. Therefore, at the receiver side no packet arrival indicates a silence of the talker. But this silence incurs inconvenience to the listener since a sudden silence can also mean that the connection is dead or disconnected. Therefore, the CNG module generates some background noise so that the connection seems alive to the listener. The DTX module determines when to send the background noise in non-speech frames. The DTX module in the G.729B codec uses a two byte Silence Insertion Descriptor (SID) frame to initiate the CNG.

G.729 Annex D (G.729D) was further introduced that reduces the bandwidth consumption by operating at 6.4 Kbps. During a network congestion, the conversation can be encoded using G.729D which results in a lower voice quality but with a very low bandwidth usage. It uses Frame Error Correction (FEC) to compensate for any packet loss and can reconstruct lost voice frames under low packet loss. G.729 Annex E (G.729E), on the other hand, uses a higher bit rate of 11.8 Kbps with a much higher voice quality. It can be used when sufficient bandwidth is available and the

[8]Voice quality equivalent to that of long distance public switched telephony network.

conversation has background noise or music. G.729E uses either CS-ACELP or Linear Predictive Coding (LPC) depending on the highest fidelity achieved for each frame.

2.1.4.2 G.711

G.711 [47] was the first codec designed by ITU-T for audio companding[9] (compression and expanding). It uses Pulse-Code Modulation scheme to digitize the voice signal. G.711 generates 8-bit frames after scaling the voice frames of 125 μs logarithmically resulting in a 64 Kbps data stream. G.711 was aimed at ensuring the interoperability between international carriers and is the standard codec used in H.323 and Integrated Services Digital Networks (ISDN). It has two variants: G.711 μ-law uses 14 bit voice samples and gives a greater resolution to higher range signals while G.711 A-law uses 13 bit voice samples and provides more quantization levels for lower signal levels. The A-law version is used in Europe while the μ-law version is used in the U.S. and Japan.

G.711 Appendix I includes a Packet Loss Concealment algorithm that hides the packet loss in the end-to-end path. Like G.729B, G.711 also supports silence suppression by utilizing DTX, VAD and CNG. This technique is a part of G.711 Appendix II. To reduce bandwidth consumption, G.711.0 was designed which can reduce bandwidth use by 50%.

2.1.4.3 G.723.1

G.723.1 [48] was originally designed for delivering video and speech over regular PSTN phone lines and allows very low bit rate coding of a voice signal. It is a dual-rate speech codec designed for the H.323 and H.324 audio and video conferencing standards for compressing voice at toll quality. G.723.1 uses 30 ms long voice frames and encodes them at either 5.3 Kbps using Multipulse LPC with Maximum Likelihood Quantization (MPC-MLQ) algorithm or 6.3 Kbps using Algebraic Code-Excited Linear Prediction (ACELP) algorithm. Including the look ahead delay of 7.5 ms, the minimum algorithm delay is 37.5 ms. A silence suppression technique is incorporated in G.723.1 Annex A which uses 4 bytes of silence insertion descriptor (compared to the 2-byte SID of G.729). Due to the high level of compression, the voice quality of G.723.1 is not high and the Dual-Tone Multi-Frequency (DTMF) or Fax tones can not be reliably transmitted using it.

2.1.4.4 G.726

G.726 [49] is another ITU-T standard for voice encoding that uses Adaptive Differential Pulse-Code Modulation (ADPCM) and supersedes both G.721 and G.723. It was

[9] http://en.wikipedia.org/wiki/Companding

initially targeted to use half of the bit rate of G.711 so that the network capacity is increased considerably. The ADPCM algorithm is used to map a series of 8 bit μ-law or A-law PCM samples into a series of 4 bit ADPCM samples, thereby halving the bit rate. G.726 can operate at bit rates of 16, 24, 32 and 40 Kbps. Encoding at 16 and 24 Kbps does not provide toll quality speech, and it is recommended that the 16 and 24 Kbps rates be alternated with higher data rates so that average sample size remains within 3.5–3.7 bits per sample. Due to the use of very low bit rate coding, G.726 is ideal for speech transmission over digital networks, video conferencing, multimedia, speech archiving and ISDN communications. G.726 was primarily targeted for and mostly used in international voice trunks. It is also used in the DECT wireless phone systems and in some Canon cameras [50].

2.1.4.5 GSM-FR

The ITU-T GSM 06.10 Full Rate (FR) was introduced in 1987 and was the first speech codec used in GSM cellular networks. It uses Regular Pulse Excitation Long Term Prediction Linear Predictive (RPE-LTP) algorithm. The encoder uses a 13 bit PCM signal and generates voice streams at 13 Kbps. Although the voice quality is not very high, the computational complexity is very low which incited its popularity in the early GSM networks.

2.1.4.6 GSM-HR

The GSM 06.10 Half Rate (HR) was introduced in 1994 for GSM networks. It uses Vector Sum Excited Linear Prediction (VSELP) algorithm and operates at 5.6 Kbps. The bandwidth requirement is therefore half of that needed by GSM-FR and the voice capacity is nearly doubled. However, it provides a slightly lower voice quality and requires higher computational power.

2.1.4.7 GSM-EFR

The GSM 06.10 Enhanced Full Rate (EFR) was introduced in 1997 as an extension of GSM-FR. It uses ACELP based algorithm and operates at 12.2 Kbps. Although it uses less bandwidth than GSM-FR, the resulting voice quality is higher.

2.1.4.8 iLBC

The Internet Low Bit rate Codec (iLBC) is a free codec that uses a block independent Linear-Predictive Coding algorithm. Due to the independence between different voice frames, in its coding, it is considerably robust to packet loss compared to G.729A or

G.723.1. It is very popular due to its royalty free availability, and therefore, many open source and free applications such as Skype, Gizmo, OpenWengo and GoogleTalk use it. iLBC can operate at either 15.2 Kbps using 20 ms voice frames or at 13.33 Kbps using 30 ms voice frames. Although its computational complexity is close to G.729A, iLBC provides a higher voice quality.

Codecs define the length of each voice frame and the number of bytes it is compressed into. This information can be used to determine the payload size and arrival rate of voice packets. Therefore, the codecs define the traffic load in a network along with the network size, i.e., number of nodes in a network. Additionally, the packetization delay, look ahead delay and any other algorithmic delay information can be used to determine the end-to-end delay in the voice stream. This information can be used in determining the user perceived voice quality in a network under investigation. A detailed analysis of each codec is out of scope of this thesis, however, a comparison of the most popular codecs is presented in Table 2.1 for reference. We discuss some voice quality estimation methods in the following section.

2.1.5 Voice Quality Assessment Methods

VoIP call quality is measured in terms of the User Perceived Quality (UPQ) using several standard techniques in the industry. These techniques can be categorized as either intrusive or non-intrusive. The intrusive techniques require a reference signal to be transmitted through the network. At the other end of the network, the received signal is compared to the original reference signal. On the other hand, the non-intrusive techniques either use the network configuration or observe network traffic to determine expected voice quality. In this section, we discuss the three most popular voice quality estimation methods, which are Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ) and E-model.

2.1.5.1 Mean Opinion Score (MOS)

The Mean Opinion Score (MOS) [51] is a subjective voice quality estimation method. The MOS defines a method of determining the user satisfaction in a telephone call in terms of a rating called the MOS level. In this method, a number of human experimenters listen to a voice conversation that has passed through a network and rate the voice quality on an scale of 1 to 5. The voice quality can degrade due to packet loss, circuit noise, ambient noise or echo, etc., but MOS only considers their

Table 2.1: Comparison of voice compression codecs.

Codec	**Type**	**Bit Rate (Kbps)**	**Frame length (ms)**	**Look Ahead Delay (ms)**	**Complexity (MIPS)**	**Memory (word)**	**Intrinsic MOS**
G.729	CS-ACELP	8 Kbps	10	5	20–25	5K	3.92
G.729A	CS-ACELP	8	10	5	10	2K	3.7
G.729D	CS-ACELP	6.4	10	5	20	4K	3.8
G.729E	CS-ACELP	11.8	10	5	25–30	6K	4
G.711	Companded PCM	64	10	0	1	1	4.4
G.723.1	MP-MLQ/ACELP	5.3/6.3	30	7.5	20–25	3K	3.6
G.726	ADPCM	16/24/32/40	0.125	0	1.25	50	3.5
G.722	Sub-band ADPCM	64/56/48	0.125	1.5	10	1K	4.1
G.722.1	MLT	24/32	20	40	15	14.5K	4
G.728	LD-CELP	16	0.625	0	30–40	2K	3.61
GSM-EFR	ACELP	12.2	20	0	15–20	4K	4.1
IS-733	QCELP	13.3/6.2/2.7	20	5	20–25	4K	–
IS-127	RCELP/ACELP	8.5/4/0.8	20	5	25–30	4K	–
AMR	ACELP	4.75/6.7/12.2	20	5	15–20	4K	–
MS-GSM	RPE-LTP	13	20	0	4.5	1K	4.1
GSM-FR	RPE-LTP	13	20	0	4.5	1K	3.6
GSM-HR	VSELP	6.5	20	5	30	4K	3.5
iLBC	LPC	13.33/15.2	20/30	25/40	18	15K	3.9
BV16	TSNFC	16	5	5	12	12K	4
BV32	TSNFC	32	5	5	17.5	12K	4.1

Table 2.2: User perceived quality of voice and corresponding mean opinion score.

User Perceived Voice Quality	MOS **level**
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

cumulative effect on the voice quality. The voice quality levels and their corresponding MOS levels are presented in Table 2.2.

2.1.5.2 Perceptual Evaluation of Speech Quality (PESQ)

The Perceptual Speech Quality Measure (PSQM) is an intrusive mechanism developed by Koninklijke PTT, Netherlands which is an adaptation of a more general method called the Perceptual Audio Quality Measure (PAQM). The human brain recalls familiar voices more accurately than music and other sounds. This phenomenon was found from experiments leading to the introduction of PSQM which is specially optimized for telephony speech signals. In essence, PSQM converts the physical-domain signal characteristics into perceptually meaningful psychoacoustic-domain quality levels. The degraded voice quality is presented on a scale of 0 to 6.5 which can then be mapped to the MOS levels. However, the original PSQM was designed to evaluate codecs in cellular networks and does not consider the QoS interruptions due to packet loss, delay jitter and non-sequential packet arrival. These factors, all of which are common with VoIP, can make the PSQM perform inaccurately in IP networks. Therefore, PSQM+ was developed to more accurately determine the voice quality degradation under realistic network conditions.

The Perceptual Analysis Measurement System (PAMS) is another intrusive mechanism developed by British Telecom. Both PSQM and PAMS inject a reference signal at one end of the network. At the other end, the propagated signal is received and then compared to the reference signal using a mathematical description of the psychophysical properties of the human hearing process. The mathematical description considers the subjectivity level of the errors that was incurred in the signal while it was being transmitted through the network. The voice quality is represented with a MOS like scale of 1 to 5.

The ITU-T developed Perceptual Evaluation of Speech Quality by combining the best features of PSQM and PAMS. PESQ considers the filtering, variable delay, coding distortion and channel errors which are all ignored in PSQM. It also adopted the excel-

lent psycho-acoustic and cognitive model of PSQM and the time alignment algorithm of PAMS that can handle varying delays accurately [44]. The resulting voice quality is presented on an scale of 1 to 4.5. PESQ is primarily used for codec evaluation and the testing of live or prototype networks. It does not require any knowledge of the network and uses only the reference and degraded signals to determine the user perceived quality of voice.

2.1.5.3 E-model

The E-model [18] was first introduced in 2000 by ITU-T to determine subjective UPQ using objective methods. This method assumes that the network configuration and performance can be mapped to a users' psychological satisfaction levels. The E-model is non-intrusive and uses a computational model to assess the network performance and device properties to determine the user perceived quality of voice in terms of a transmission rating factor called R. The E-model considers voice quality impairments caused by any network delay, packet loss and a low bit rate codec, etc., by computing a series of input parameters. Assuming that "the psychological factors on the psychological scale are additive" [18], the rating factor R is computed as,

$$R = R_0 - I_s - I_d - I_{e\_eff} + A.$$

Here,

- R_0 is the basic signal to noise ratio including any noise source such as circuit noise and room noise.
- I_s is the simultaneous impairment factor that combines all the impairments which occur more or less simultaneously with the voice signal. Voice quality impairments due to too loud speech levels, non-optimum sidetone and quantization noise that occur more or less simultaneously with the voice signal are considered in I_s.
- I_d is the delay impairment factor that considers any delay and echo in the end-to-end path of the voice stream.
- I_{e_eff} is the effective equipment impairment factor due to a low bit-rate voice compression codecs. It also includes the impairment due to random distribution of packet loss.
- A is the advantage factor that represents the advantages of access, for instance, a little degradation in the voice quality is more likely to be accepted by a mobile phone user than a fixed line user.

Table 2.3: Relationship between MOS level, R score and user satisfaction.

Minimum R score	**Minimum MOS value**	**User Satisfaction Level**
90	4.34	Very satisfied
80	4.03	Satisfied
70	3.60	Some users dissatisfied
60	3.10	Many users dissatisfied
50	2.58	Nearly all users dissatisfied

The computation of the rating factor R involves a large number of parameters. but network designers usually look for changes in the voice quality with respect to only a few of these parameters. That is why ITU-T additionally proposed complementary recommendations [52–56] which advise default values for these parameters for predetermined conditions, i.e., when these parameters are not changing and a prescribed value can be used. For instance, to determine the equipment impairment factor for a particular codec, the associated parameters can be used from ITU-T G.113 [55] which was designed to assess transmission impairments in international telephone calls. To assess the impairment incurred by a waveform type codec, [55] used the quantization distortion method which was further enhanced by incorporating a new planning method called the Equipment Impairment Factor. Equipment impairment factor considers the impacts of non-waveform codecs, non-optimum overall loudness rating, talker echo and the speech difficulty associated with long one-way delay.

Since MOS is the mostly used method for voice quality measurement, all the other methods usually use a mapping to determine the voice quality on the corresponding MOS level. Table 2.3 presents R-scores and MOS-levels for different user satisfaction levels. For a given R score, the corresponding MOS score can be determined as,

$$MOS = \begin{cases} 1 & \text{for } R < 0, \\ 1 + 0.0035 \times R + 7 \times 10^{-6} \times R(R-60)(100-R) & \text{for } 0 \leqslant 100, \\ 4.5 & \text{for } R > 100. \end{cases} \quad (2.1)$$

MOS requires human experimenters and can not be used in any automated testing or mathematical analysis. The PESQ and E-model are, on the other hand, objective methods that can determine the subjective perception of voice quality. But PESQ is intrusive and requires a reference signal to be injected into the network under investigation. Both MOS and PESQ require an existing network to be used, and therefore, they are useless in the design and planning phase. During network design and planning, the network administrators need to determine the possible range of devices, their parametric configuration, codec, packetization interval, number of access points, transmission range, number of channels, etc. Only the E-model can take the network configuration

and device parameters into account in determining the resulting voice quality. We discuss some real world applications of the E-model in the following section.

2.1.6 Utilization of the E-model

The E-model can be used by network designers and administrators in order to maintain a targeted voice quality level. In this respect, the E-model can be utilized in any of the following ways.

- **Network Design and Planning:** For a targeted number of voice calls with an acceptable level of voice quality, the network configuration can be determined using the E-model in terms of the number of channels, the number of access points and their placements, the network interface cards with a predefined transmission and interference range, the channel noise, capture threshold, etc. When the network is designed with these design decisions as constraints, the resulting traffic can be served while maintaining the desired level of voice quality.

- **Call Admission Control:** Alternatively, the E-model can be utilized in conjunction with a Call Admission Control (CAC) system that usually resides in the access point. By considering the current network configuration and traffic load, the CAC module can determine whether to accept a new voice connection request. When the available resources are insufficient to accommodate more calls, any new requests will be denied so that the existing connections can maintain an acceptable level of UPQ.

- **Adaptive Network Management:** It is also possible to modify the network resources (generally increase resources only) in order to serve a sudden increase in traffic volume. For example, if a temporary increase in the traffic demand occurs that is crucial to accommodate, some additional access points or mesh routers can be placed at appropriate locations in order to extend the coverage or bandwidth availability. This kind of situations might occur at an emergency road crash site, public open fair or football match with an unknown traffic volume and where the participants can join any time without reservation. If a sudden rush of incoming crowd appears, the growing demand of bandwidth can cause the ongoing calls to suffer from call jitter or call drop. Therefore, the network administrators need to observe the traffic growth rate (usually done using traffic monitor applications), then estimate the achievable voice quality using the E-model and take appropriate measures.

Adaptive network management can not be automated, as it requires manual work in observation and maintenance. On the other hand, call quality assurance by network planning and by employing a CAC module are commonplace. In this respect, we target to design voice capacity models that can assist in both the network designing and call admission control. A call admission control system needs to monitor the performance of all existing calls and observe the changing network configuration since the WLAN client nodes can be mobile. Therefore, the CAC module incurs additional overheads in terms of memory, computational time and power supply. There is no CAC module specification in the 802.11 standards and the performance of a CAC system is heavily dependent on the specific implementation. This is the reason why we have put special focus on the network design and planning approach. Since only the E-model can be used to predict the voice quality in a network before it exists, we choose the E-model in the current study. We discuss a few practical implementations of the E-model in this section.

Maintaining voice quality in an existing network requires monitoring of the ongoing voice calls so that a change in traffic load and its effect on the voice quality are detected early enough to allow appropriate measures to be taken. PROGNOSIS VoIP Monitor by Avaya[10], OpManager by ZOHO Corporation[11] and Ditechs Experience Intelligence (EXi)[12] are examples of commercial VoIP quality monitors. But measuring the voice quality for every voice packet is computationally costly and can hamper the network performance if it is not done carefully. Although the time complexity for calculating the R score from the network parameters using the E-model is $O(1)$, the computational complexity was deemed too high, specially for early voice quality monitors. The E-model uses a set of 27 equations using 18 parameters to determine the voice quality. To measure these parameters and calculate the R score in real time was considered too costly in terms of the computational power for network monitors or admission control modules. For this reason, early researchers tried to approximate the E-model in a number of ways as discussed below.

Cole and Rosenbluth [57] presented a transport level voice quality monitoring technique. To utilize the E-model at the transport level, the existing E-model was reduced and represented in terms of the low level transport level metrics which impact the voice performance, for example, delay, delay jitter, loss, etc. The packet loss, delay

[10] www.avaya.com
[11] www.zohocorp.com.cn
[12] www.ditechnetworks.com

jitter, dejitter buffer performance and packet size were combined with a frame erasure mask which characterizes the salient features of the loss distribution at the decoder.

A packet Loss Concealment algorithm was used in conjunction with a look-up table to determine the approximated equipment impairment factor in a simplistic manner. In fact, in the older version of the E-model standard [58] the equipment impairment factor was presented as a tabulated parameter that can be used as a lookup table. A parameterized logarithmic function was used in [57] to approximate the actual value of the equipment impairment factor. But in the latest version of the E-model standard [18], the equipment impairment factor is defined using another equation which we shall discuss in Section 3.4.

Similarly, the delay impairment was computed in [57] for a range of end-to-end delay. Then a piecewise linear curve was used to approximate it within this region. However, it was shown in [23] that the approximation in [57] is not accurate enough for the required range of the end-to-end delay and a piecewise quadratic curve was proposed to give a more accurate approximation. Finally, the approximations of the delay and loss impairments were combined in [57] to give the R score. A closed form solution for the dejitter loss was also proposed to estimate the end-to-end loss for a static dejitter buffer. The system computes the voice quality in terms of the R score in a very simplistic manner so that it can be used in monitors with a relatively low computational power in order to assist in fault detection and quality management.

Carvalho *et al.* [59] presented a measurement tool based on the E-model and used this tool in the Brazilian National Education and Research Network (RNP) backbone. Some corrections in the use of the E-model were suggested, for instance, according to [51] the MOS level becomes less than 1 for $0 < R < 6.5$. But MOS should be measured on a scale of 1 to 5, and [59] suggested a MOS level of 1 to be used in (2.1) for $R < 6.5$, instead of $R < 0$. Corrections of the equipment impairment function and packet loss probability for bursty traffic were also suggested. The tool uses the G.711 μ-Law codec with an aggregation level of three and employs OpenH323[13] which is an open source implementation of the H.323 [60] mechanism. H.323 defines protocols to manage the audio-visual communication sessions for a packet based network. This standard defines the call signaling and control, multimedia transport and control and bandwidth control for point-to-point and multi-point conferences. The tool in [59] can determine the voice quality by analyzing a trace from the dejitter buffer.

The backbone networks are usually sufficiently provisioned and well maintained. However, if the voice quality impairments are not carefully considered, the backbone networks can become bottlenecks leading to a degradation in VoIP performance. Markopoulou *et al.* [61] investigated whether Internet backbones are ready for VoIP

[13] http://openh323.sourceforge.net

based telephony. Different types of Internet carriers in five U.S. cities were used to measure the delay and loss jitter over 43 different paths in order to determine any voice quality impairments. The delay and loss were measured by means of probe packets which were used in an implementation of the E-model to determine the voice quality variations. A significant number of Internet Service Providers were found to lack sufficient capabilities to ensure voice quality. Moreover, when stringent quality requirements like the interactivity level for business conversations were applied, these Internet paths became completely unacceptable for Internet telephony. Packet loss and high delay jitter were found to be a ruling factor in degrading the voice quality. Most of the problems were found to be related to link failures, routing reconfiguration and router operations rather than to the traffic load. More effort in maintaining the reliability was suggested to improve the voice quality.

2.2 VoIP over Wireless LANs

Despite the wide acceptance and ubiquitous popularity of VoIP, mobile phones are still more popular due to their convenience of mobility. The ability to communicate while the user can walk, talk or do other works simultaneously is very attractive. Mobile phone services achieve this by delivering voice services using a GSM (Global System for Mobile Communications, originally derived from Groupe Spécial Mobile) or CDMA (Code Division Multiple Access) cellular network. The cellular network uses a huge number of base stations to provide wireless coverage to almost all areas where their subscribers can be. But the deployment and maintenance of such a vast cellular network incur a high cost which is ultimately passed down to the subscribers. This is why the mobile phone services are still very costly. VoIP service can be provided in mobile phones using a wireless carrier like GSM, CDMA, WPAN, WiMAX or Wi-Fi.

There have been a number of attempts to deliver VoIP calls using cellular network (for instance, Gotalk VoIP mobile phones[14]). Such solutions are not very popular because the subscribers still have to pay the service charges of the cellular network provider. The only benefit is achieved with international calls and mobile-to-PSTN calls where the intermediate carriers are replaced with VoIP connections. But for mobile to mobile calls, the service still remains costly.

IEEE 802.15 defined the Wireless Personal Area Network (WPAN) based on Bluetooth v1.1 and v1.2 technology. It defined a media access control and physical layer specification and was targeted to short range, low bandwidth communication. Initially, only a low transmission rate of 1 Mbps was supported using the GFSK modulation. In later versions, 2 and 3 Mbps data rates were included using the DQPSK and 8DPSK

[14] http://www.gotalk.com.au

based modulation mechanisms. Bluetooth communicates in a master-slave fashion where a master can communicate with up to seven slave devices. Later versions of WPAN adopted Ultra Wide Band (UWB), WiMedia and ZigBee to provide a higher data rate and greater transmission range. Despite the low cost of the devices, WPAN technology is only suitable for personal gears like headphones, PDAs, wireless keyboards and mouses, etc.

IEEE 802.11 defined Wireless Local Area Network (WLAN) whose industry counterpart is known as Wireless Fidelity (Wi-Fi), maintained by the Wi-Fi Alliance[15]. WLAN offers a moderate coverage of 100 m and a data rate of 1∼54 Mbps. Wi-Fi uses low cost devices and operates in unlicensed bands at no additional cost. The ad hoc nature of operation makes deployment and maintenance very easy keeping the operational costs to a minimum. Moreover, Wi-Fi is the only viable networking solution for temporary emergency wireless networks like road-crash sites or war-field networks. The ease and low-cost of using Wi-Fi devices is the main reason for their ubiquitous use in home, public and corporate networks. Another important advantage of using Wi-Fi technology is its widespread availability. Today every laptop, PDA and many mobile phones are equipped with Wi-Fi modules. The wide popularity of Wi-Fi technology led to its inclusion in the most popular hand-held devices like Apples iPhone, iPad and iPod-Touch and Google's Android. A recent report by Coda Research Consultancy [62] states that by 2015 Wi-Fi enabled devices in use will reach 2.6 billion units with a US household Wi-Fi penetration of 71%.

WiMAX or Wireless Metropolitan Area Network (WMAN) was introduced in IEEE 802.16. It was designed to provide wireless broadband Internet to home users. A transmission range of up to 50 km allows only a few access points to cover a whole city. WiMAX supports a high data rate of 134 Mbps which is able to support most day to day application needs. Initially, WiMAX supported only Line of Sight (LOS) communication but in recent revisions non-LOS communication is also supported. Despite the availability of a high data rate and long transmission range, this standard is seldom used in mobile devices due to their high cost of manufacture and operation. Only rooftop devices are popular with Internet Service Provides but they are heavy and still costly. Also, operation in licensed bands incurs an additional cost. Although WiMAX can optionally operate in unlicensed band, the background interference reduces the transmission range and data rate considerably giving no benefit over Wi-Fi.

[15] www.wi-fi.org

Table 2.4: Comparison of WLAN, WPAN and WMAN [5].

Technology	IEEE WLAN (Wi-Fi)					IEEE WPAN				IEEE WMAN (WiMAX)
Standard	802.11	802.11a	802.11b	802.11g	802.11n	802.15.1 (Bluetooth)	802.15.3 (UWB)	802.15.3a (WiMedia)	802.15.4 (ZigBee)	802.16
Frequency band (GHz)	2.4	5.8	2.4	2.4	2.4–5.8	2.4	3.1–10.6	2.4	2.4	2–66
Range (m)	70	10	100	110	160	10	10	10	100	50,000
Data rate (Mbps)	2	54	11	54	248	3	55-1000	110–1000	0.25	134
Physical layer	DSSS, FHSS	OFDM	DSSS, CCK	OFDM	MIMO	FHSS	DS-UWB, OFDM	MB-OFDM	DSSS	MIMO-SOFDMA
Modulation	GFSK, BPSK, DBPSK, DQPSK	PBSK, QPSK, 16-QAM, 64-QAM	DPSK, DBPSK, DQPSK	BPSK, QPSK, 16-QAM, 64-QAM, DBPSK, DQPSK	BPSK, QPSK, 16-QAM, 64-QAM	GFSK, 2PSK, DQSP, 8PSK	OPSK, BPSK, OOK, PAM, PPM, Bi-Phase	QPSK, DCM	BPSK, OPSK	QPSK, QAM

To provide wireless VoIP service to mobile devices, we need to carefully choose a wireless technology to deliver the voice packets. Low cost of service is the ultimate bargaining chip of VoIP, therefore, any high cost solution will lessen the appeal of this technology as a platform to deliver voice packets over the Internet while providing the convenience of mobility. GSM and CDMA (CDMA2000) offer communication at 14 and 3 Mbps, respectively. The service is reliable and the handoff delay for roaming users is minimal. But the maintenance of the cellular network incurs a high cost. Bluetooth suffers from a short transmission range and low data rate and is only applicable to communication between personal devices. WiMAX offers a high data rate and long transmission range but the high cost of devices and operation in the licensed bands make the service cost high. Only Wi-Fi offers moderate transmission range and data rate at a very low cost. Moreover, the ubiquitous availability and widespread popularity make it an optimal solution for mobile VoIP. Most PDAs and hand held computing devices like the Android and iPad are equipped with Wi-Fi interfaces. Every international airport, many railway stations and public hotspots like cafes or stadiums provide free Wi-Fi services. Moreover, the decreasing cost of 802.11 devices make it a presumable part of nearly any home electric equipment in the near future forming low cost home mesh networks [63]. Therefore, Wi-Fi networks will provide the last mile coverage for VoIP traffic for the vast majority of customers in the future. A short comparison of WLAN, WPAN and WMAN is shown in Table 2.4. We review the IEEE 802.11 WLAN technology in the following section.

2.3 IEEE 802.11 Wireless Local Area Networks

IEEE 802.11 is a series of international standards which define the characteristics of Wireless Local Area Networks (WLAN). The IEEE LAN/MAN Standards Committee (IEEE 802) defined a number of standards in the 802.11 series for different WLAN features. The Wi-Fi alliance adapted different versions of the 802.11 standards as the Wi-Fi standard in order to maintain interoperability between devices manufactured by different vendors. The IEEE 802.11 standards define only the PHYsical (PHY) layer and Medium Access Control (MAC) layers. The upper layers of the TCP/IP or ISO OSI protocols remain unconstrained so that the MAC and PHY layers can be used with any routing algorithm, link control protocol or session management protocol to serve any type of applications. In this section, we briefly discuss the functionalities and key features of these standards.

IEEE 802.11 [64] is the original legacy standard defined in June 1997 and revised in 1999. It operates in the 2.4 GHz spectrum and uses an Infrared, FHSS (Frequency-Hopping Spread Spectrum) or DSSS (Direct-Sequence Spread Spectrum) based physical layer with maximum transmission range of only 70 meters. The Infrared based PHY layer can operate at 1 Mbps while the FHSS and DSSS based PHY layer offer 1/2 Mbps data rate. This standard was flexible and a number of design choices were left open which made the interoperability between WLAN devices of different vendors challenging.

IEEE 802.11a [65] was announced in September 1999 in the hope of increasing the data transmission rate. It uses Orthogonal Frequency Division Multiplexing (OFDM) based physical layer and allows a high data rate. Data can be transmitted at 6/9/12/18/24/36/48/54 Mbps. This standard supports the simultaneous use of 8 channels in the 5 GHz frequency band and includes an error correction mechanism. The 2.4 GHz spectrum being heavily used by many other house hold applications (e.g., microwave oven, Bluetooth devices, baby monitors and cordless DECT telephones), the 5 GHz spectrum offers much less noisy channels. However, the high carrier frequency also suffers from a short transmission range and poor obstacle (e.g., wall and other solid objects) penetration capability due to its shorter wavelength.

IEEE 802.11b [66] was also announced in September 1999 which extends the original standard by increasing the throughput and transmission range and lowering the manufacturing cost. It uses a DSSS based physical layer in a 2.4 GHz frequency band and defines 3 channels over which data can be transmitted at 1/2/5.5/11 Mbps. Eight-chip Complementary Code Keying (CCK) is used as the modulation mechanism to achieve this higher data rate. CCK uses a nearly orthogonal complex code set called the complementary sequence. The chip rate is kept constant while the data rate is varied according to the channel conditions by changing the spreading factor. The IEEE 802.11b standard also introduced the Auto Rate Fallback (ARF) mechanism to adjust the data transmission rate depending on the link quality.

IEEE 802.11e [67] was targeted to improve the quality of service. This standard identified the requirements of different types of packets in terms of the bandwidth and transmission delay so that the voice and video traffic are less impaired than other traffic by network congestions.

IEEE 802.11g [68] offered a high bandwidth with a low frequency band consumption. It allows transmission at 54 Mbps in the 2.4 GHz frequency range using either an OFDM or DSSS based physical layer. The allowed transmission rates are 1/2/6/9/12/18/24 Mbps. This standard is backward compatible with 802.11b.

IEEE 802.11i [69] was designed to improve the security of data transfer and used an Advanced Encryption Standard (AES) to manage and distribute keys. It implements encryption and authentication which can be used with the 802.11a, 802.11b and 802.11g standards.

The IEEE 802.11 legacy standard was more of a beta standard and due to interoperability reasons, it was never implemented. The MAC layer specification was revised in each version but remained more or less intact. However, the IEEE 802.11a, 802.11b and 802.11g standards redefined the physical layer specification in order to improve performance. On the other hand, 802.11e and 802.11i improved the quality of service and data security, respectively. IEEE 802.11b was the first and most commercially successful version of WLAN. 802.11a was not commercially successful mainly due to the high cost and complexity of manufacturing 5 GHz devices while the less expensive 802.11b products quickly occupied the market. The shorter wavelength used in 802.11a is easily absorbed by solid objects in their path. The poor obstacle penetration capability leads to shorter transmission range which requires a higher number of access points, compared to 802.11b/g, to cover an area. Later revisions of 802.11a came up with a longer transmission range but only at the cost of lower data rate similar to 802.11b. IEEE 802.11g also received a lot of attention and public acceptance primarily due to its backward compatibility with 802.11b. Only a few vendors manufacture 802.11a devices while 802.11b and 802.11g are commonplace. IEEE 802.11a, 802.1b and 802.11g primarily differ in the offered bandwidth and frequency band usage as shown in Table 2.4.

IEEE 802.11 divides the available spectrum band into a number of channels. To reduce the inter-channel interference, the channels need to be separated by a guard band. For instance, the 2.4 GHz band is divided into 13 channels (numbered from 1 to 13) where each channel is 22 MHz wide and spaced by 5 MHz guard bands. Japan adds another channel of 14 MHz. The availability of these channels is regulated by each country differently depending on how they allocate and license radio spectrums for different services. Japan authorizes the use of all 14 channels while European countries allow only the originally defined 13 channels. North America and Central and South American countries allow only the first 11 channels. But the 802.11 standard also defines a spectral mask which defines the permitted distribution of power across each channel. The mask requires that the signal be attenuated by at least 30 dB from its peak energy at ± 11 MHz from the center frequency. This requirement on the attenuation means that only every fourth or fifth channel can be used simultaneously

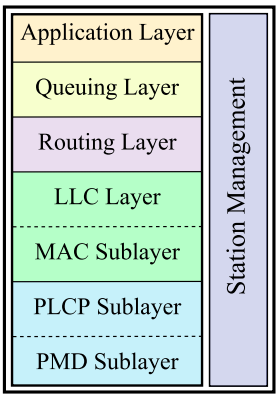


Figure 2.3: Architecture of an 802.11 compliant node.

without any overlapping. Generally channels 1, 6 and 11 are used in America and either channels 1, 5, 9 and 13 or 1, 6 and 11 are used in Europe.

The architecture of an 802.11 compliant node is shown in Fig. 2.3. The IEEE 802.11 standards define only the MAC layer and the physical layer. The physical layer is divided into the Physical Layer Convergence Protocol (PLCP) and the Physical Medium Dependent (PMD) sub layers. The PLCP prepares and parses data units that are transmitted or received using various 802.11 medium access techniques. The PMD performs the data transmission-reception and modulation-demodulation under the control of the PLCP. The MAC layer controls access to the channel. In the following, we briefly describe the MAC and PHY layers of the IEEE 802.11 standard.

2.3.1 Medium Access Control (MAC) Layer

The MAC layer provides the following services—

1. Authentication
2. MSDU delivery
3. Distribution
4. Deauthentication
5. Association
6. Integration
7. Privacy
8. Disassociation
9. Reassociation

A group of two or more 802.11 compliant nodes operating in the same frequency band is called a Basic Service Set (BSS). A BSS can comprise of client nodes only or the client nodes can operate under the supervision of an access point. Depending on the configuration, the WLAN can operate in one of three modes—

Independent Basic Service Set (IBSS) In an IBSS, the client stations (or nodes) communicate directly with each other. No access point is necessary to coordinate access to the channel and the client nodes discover routes and control access to the channel on their own on an ad hoc basis. This mode is sometimes called the Ad Hoc mode. An IBSS has limited access in the sense that the nodes in an

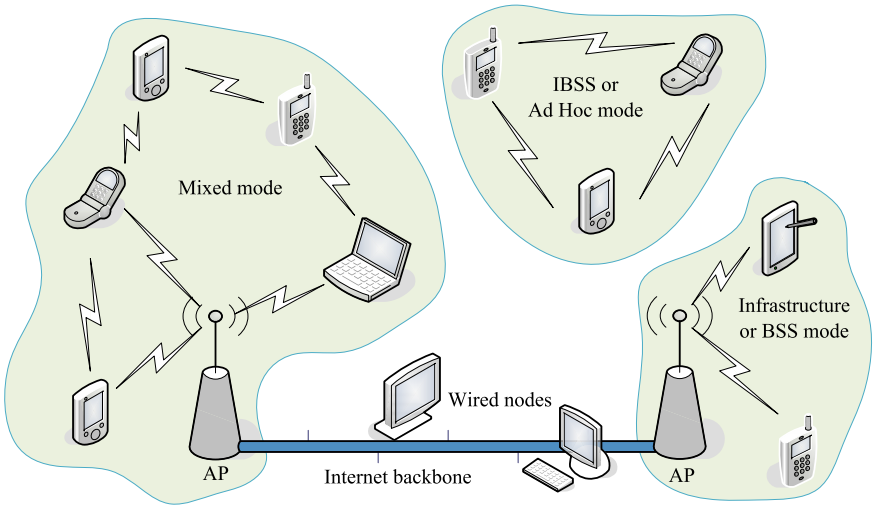


Figure 2.4: WLAN operation modes.

IBSS can not communicate with nodes in any other IBSS. Essentially, an IBSS is isolated from other networks in the world.

Infrastructure BSS When an access point is part of a BSS, the WLAN is termed as an Infrastructure BSS. The access point can coordinate medium access or simply provide Internet connection through a distributed medium access mechanism.

Mixed mode or Hybrid mode If an access point serves a BSS but only a portion of all the all client nodes can communicate directly with it, the WLAN is termed to operate in Hybrid or Mixed mode. The nodes which can not receive the transmissions from AP directly, can still communicate with the AP through the other client nodes using multihop routes.

The primary purpose of an access point is to provide Internet connectivity to the client nodes allowing them to communicate with stations in other wired or wireless networks. In this way, when two or more Infrastructure or mixed mode BSS are connected together by a Distribution System (DS) or wired network, the combined network is called an Extended Service Set.

IEEE 802.11 defines two medium access mechanisms which are described in the following section.

2.3.2 Medium Access Control Mechanisms

A wireless channel is a shared medium where simultaneous transmissions can collide with each other. But unlike Ethernet, a transmitting radio network interface can not detect the received power at the receiving node, and collisions can only be detected at

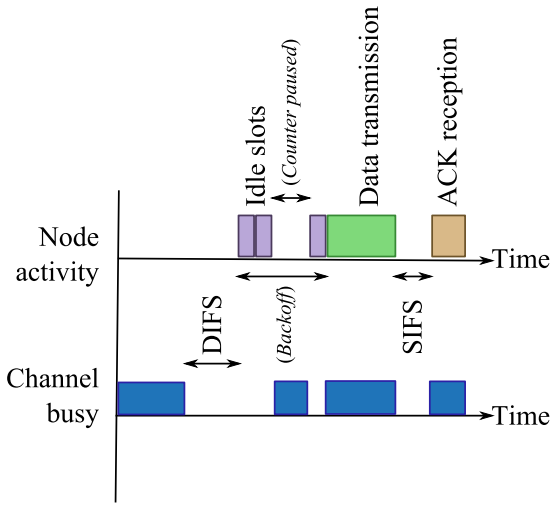


Figure 2.5: Basic handshake in the DCF based medium access mechanism.

the receiving end. Each Wi-Fi interface operates in a half-duplex mode only and can not detect collisions during their transmissions. This is why collision avoidance techniques are used instead of collision detection mechanisms. IEEE 802.11 uses a Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) based medium access control mechanism where a listen-before-talk strategy is used. Using this mechanism which is known as the Distributed Coordination Function (DCF) the individual nodes operate in a distributed manner to access the channel. Additionally, 802.11 offers a time synchronized medium access mechanism called the Point Coordination Function (PCF) which needs to be coordinated by an access point. We describe the DCF and PCF based medium access mechanisms in the following.

2.3.2.1 Distributed Coordination Function (DCF)

The basic DCF based medium access mechanism is illustrated in Fig. 2.5. When a node wants to transmit a packet, the node monitors the channel and waits until the channel is idle for more than a Distributed Inter Frame Space (DIFS) period. The channel can be sensed using either physical or virtual sensing (please see Section 2.3.3). If the channel is found to be idle for more than a DIFS, the node transmits the frame. Otherwise, it initiates a backoff mechanism. Between the transmission attempts of two consecutive packets, the backoff is used even if the channel is idle for more than a DIFS to avoid channel hogging. DCF uses discrete time backoff where the time is slotted and is calculated in terms of idle slots (or simply slots) and transmissions can occur only at the beginning of a slot. The length of an idle slot is defined for each physical layer

type in the standard and is long enough to detect an ongoing transmission. The length accounts for any propagation delay, the time needed to switch from the receiving state to the transmitting state and signal the MAC layer of the channel-busy state.

Once the backoff mechanism is initiated, the node sets its backoff counter to a uniform random number between zero and the initial contention window. The counter is decremented after each idle slot as long as the channel remains idle. If activity is detected in the channel, the counter is paused. Counter decrement is resumed again when the channel is detected to be idle for more than a DIFS.

The wireless node transmits the data frame when the backoff counter reaches zero. When the destination node receives a packet addressed to it, it waits for a Short Inter Frame Space (SIFS) and transmits an acknowledgment (ACK) frame that carries the sequence number of the successfully received data frame. Reception of an ACK frame at the sender station confirms the success of the associated transmission. But if the transmission collides with another concurrent transmission or the frame contains channel errors, the destination node can not retrieve its contents and does not transmit any ACK frame. If no ACK frame is received within an *ACK Timeout* following a transmission, the transmission is assumed to have failed. The length of the SIFS is shorter than the DIFS which ensures uninterrupted acknowledgment. Since every node waits for the channel to be idle for at least one DIFS period, access to the channel is secure for the destination node and no backoff is required during the handshake process comprising the transmissions of a data and an acknowledgment frames.

When multiple transmissions (of different frames by different nodes) collide, the frame that is received with the maximum power can still be retrieved if its received power exceeds the sum of all other received power by a *capture threshold*. This mechanism is called power capture. If the transmission fails due to a collision without capture or channel errors, the contention window of the sending station is increased using a backoff algorithm and the backoff procedure is performed again. On the other hand, if the transmission is successful, the contention window is reset to a small value according to the backoff algorithm and the next packet in the queue, if any, is picked and served. The backoff algorithm determines the level of the increase or decrease after each failure or success in transmission, respectively. A number of backoff algorithms are discussed in Section 2.3.4.

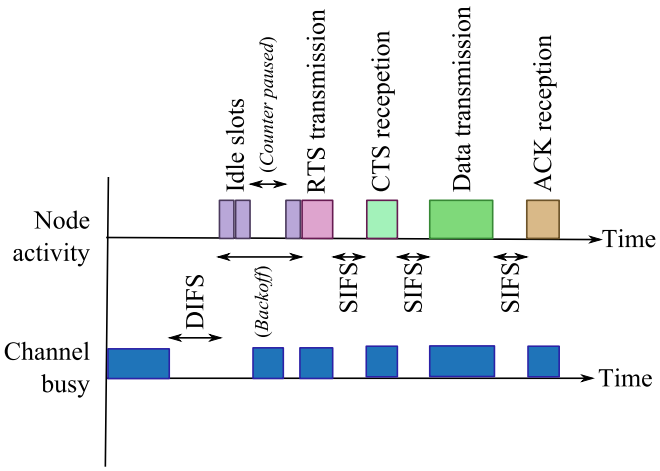


Figure 2.6: RTS/CTS handshake in DCF based medium access.

Collision occurs only at the receiving end and the transmitting station can not detect the success or failure of its transmissions. As a result, the hidden node problem and exposed node problem arise. The above mentioned handshake mechanism, which is generally called the basic DCF, therefore fails to safeguard a data packet transmission from collisions, the hidden node problem and the exposed node problem. To protect the data frame transmission from these problems, a channel reservation technique called the RTS/CTS based DCF handshake is used instead. The RTS/CTS based DCF handshake is adopted from the Multiple Access with Collision Avoidance for Wireless medium (MACAW) protocol proposed by Bharghavan *et al.* [70]. MACAW is an extension of the noteworthy MACA protocol [71] and includes RTS-CTS based medium reservation mechanism proposed by Biba [72] and employs a backoff algorithm. The RTS/CTS based handshake uses a channel reservation technique utilizing the virtual carrier sensing mechanism to avoid the above mentioned problems associated with the basic handshake.

In the RTS/CTS based handshake, a Request To Send (RTS) frame is initially transmitted instead of the data frame as shown in Fig. 2.6. On reception of the RTS, the receiver transmits a Clear To Send (CTS) packet after an SIFS. Both the RTS and CTS frames contain the probable duration of the handshake process that comprises the transmissions of the RTS, CTS, data and ACK frames. The handshake duration is used by every listening node to avoid accessing the channel during this period. This mechanism utilizes virtual sensing which is explained in Section 2.3.3. Therefore, the RTS/CTS frame interchange works like a virtual channel reservation technique. The sender waits for an SIFS and transmits the data frame after receiving the CTS

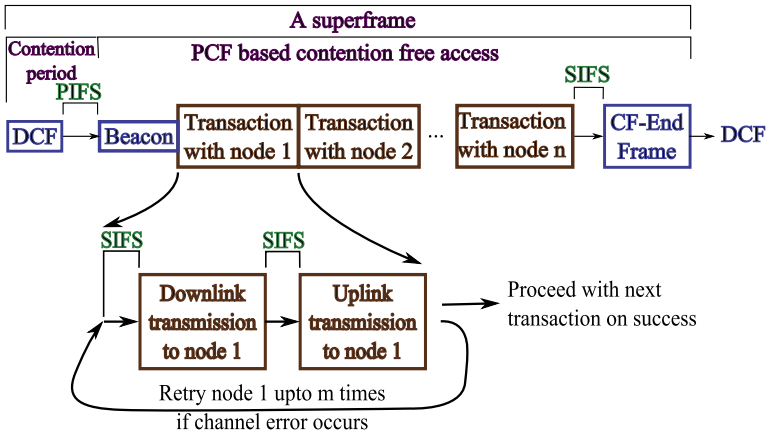


Figure 2.7: PCF based medium access mechanism.

frame. The receiver transmits an ACK after the SIFS confirming the success of the transmission, as done in the basic DCF.

Collisions can still occur with the RTS frames as it could with the data frames in the basic DCF handshake. But the RTS frames being small in size takes much less time to transmit and a collision wastes a much shorter channel time. However, transmissions of the additional RTS and CTS frames do incur a delay overhead. Therefore, the RTS/CTS based handshake performs better for larger data frames only while the basic handshake performs better for smaller frames. This is why the MAC layer uses RTS/CTS based handshake only for data frames that are longer than a predefined *RTS threshold*. In all other cases, the basic handshake is used. For example, VoIP packets are very short, usually carry about 20–160 bytes of payload and are transmitted using the basic handshake mechanism.

2.3.2.2 Point Coordination Function (PCF)

The Point Coordination Function is a time synchronized mechanism that is controlled by a Point Coordinator (PC). The PC is usually co-located with the access point (AP). PCF was made optional in the original standard and a number of PCF parameters were kept open which makes interoperability challenging. This is why the Wi-Fi alliance, who are primarily concerned with interoperability, did not include PCF based access mechanism in the Wi-Fi standard. Due to this reason, only a few vendors ever implemented PCF in their products. These products include AOpen's WarpLink AOI-706 [73] and the SDW-820 by Spectec [74].

In a PCF enabled network, the DCF based Contention Period (CP) and PCF based Contention Free Period (CFP) alternate in turns as illustrated in Fig. 2.7. One iteration of a CP and a CFP are conjunctively called a super-frame. At the beginning

of a super-frame, the PC waits for the channel to be idle for a PCF Inter Frame Spacing (PIFS) and then transmits a Beacon frame. The Beacon frame starts the CFP and contains the transmission interval of future Beacon frames and the expected length of the CFP. The client nodes in the BSS set their NAV (Network Allocation Vector) to this length and refrain from accessing the channel by means of virtual sensing.

During the CFP, the PC polls each node using a polling list. In the 802.11 standard, a round robin polling schedule is proposed. However, this strategy is unfair to different traffic categories as voice and video suffer from severe quality degradation due to a long delay but can tolerate minimal loss while, on the other hand, Telnet or FTP connections can tolerate a longer delay but can not tolerate packet loss. Therefore, a number of static and adaptive scheduling mechanisms are proposed in the literature (discussed in Section 2.5.5) in order to prioritize different traffic categories and improve the performance of voice and video traffic.

The polling list is populated whenever a node sends an Association request to the PC. The node is removed from the list after receiving a Disassociation request which a client node sends before moving away from the current WLAN. After transmitting the initial Beacon, the PC waits for a SIFS and then transmits a downlink polling frame to the first node in its polling list. The downlink packet can contain data and acknowledgment information, if there is any, along with the polling information. If there is no data or acknowledgment to be sent to this node, only the polling information is sent. After receiving a polling frame, the associated node waits for a SIFS and transmits an uplink frame. The uplink frame can contain acknowledgment information piggybacked in a data frame. The data and acknowledgment can be sent to the PC or any other node in the WLAN. However, if the downlink frame contains error, the node can not decode the polling information and will not respond. The absence of an uplink frame can also occur if the node suddenly dies (due to power failure or system crash) or moves away. If no uplink packet is received at the PC within a PIFS, the last downlink transmission is assumed to have failed and the PC continues with the next transmission in its schedule. To differentiate between a downlink transmission failure and the absence of uplink data or acknowledgment information at the client node, the client node still transmits a null frame when there is no data or acknowledgment information to be sent. The downlink and uplink transmissions in a super-frame associated with a given client node are conjunctively termed in this work as a *transaction*.

When transmission errors occur in the uplink or the downlink transmissions, which are identified by a received frame with errors or the absence of an uplink frame within a PIFS, respectively, the PC can retry transmitting to the same node until a retry limit expires. When the retry limit is exceeded, the packet is dropped and the PC continues to the next node in the polling schedule. As soon as the polling list is exhausted (i.e.,

every node is polled) or the predefined CFP repetition interval finishes, the PC waits for a SIFS and transmits a CF-End frame that announces the completion of the CFP. The CF-End frame carries a zero value as the remaining duration of the current CFP which is used by every node to reset their NAV and then every node starts to actively and physically sense the channel and compete for access. Therefore, as soon as the CF-End frame is transmitted, a contention period starts which continues until the PC transmits the next Beacon frame announcing another contention free period.

As mentioned before, IEEE 802.11e improves the quality of service for QoS aware applications like voice, video, etc. This is done by differentiating the traffic categories and prioritizing QoS aware traffic over background traffic. IEEE 802.11 defined a new medium access control mechanism called HCF which uses two other access mechanisms, namely, EDCA and HCCA. The HCF based medium access is explained in the following.

2.3.2.3 Hybrid Coordination Function (HCF)

IEEE 802.11e [67] introduced the Hybrid Coordination Function (HCF) in order to improve performance of voice and video traffic over WLANs by defining four Access Categories (AC) and prioritizing voice and video traffic over background traffic. HCF uses a DCF like CSMA/CA based access mechanism called Enhanced Distributed Channel Access (EDCA) and a PCF like access mechanism called HCF Controlled Channel Access (HCCA).

2.3.2.4 Enhanced Distributed Channel Access (EDCA)

The EDCA mechanism extends the DCF mechanism by introducing service differentiation for different traffic types. Access Categories (AC) are introduced for each of the four traffic types which are voice, video, best effort and background traffic. Each node still contends for channel access as done in the DCF based medium access mechanism, but uses different access parameters depending on the traffic type of the currently selected packet. Additionally, once a node has acquired access to the channel, instead of sending a single packet (like in DCF) it can transmit a number of packets as long as the total transmission time is within a time limit called the transmission opportunity (TXOP). The parameters that are selectively chosen for each AC are as follows—

- **CWmin** is the minimum contention window size. With a smaller CWmin, a node waits for less time in the backoff state and attains a higher transmission probability.

Table 2.5: Medium Access Parameters for Different Access Categories

Access Category	Traffic Type	CWmin	CWmax	AIFSN	TXOP (ms)
AC_BK	Background	31	1023	7	0
AC_BE	Best Effort	31	1023	3	0
AC_VI	Video	15	31	2	3.008/6.016
AC_VO	Voice	7	15	2	1.504/3.264

- **CWmax** is the maximum contention window size and, similar to CWmin, a higher priority AC has a lower CWmax.
- **AIFSN** is the Arbitration Inter Frame Space. Each node waits for an AIFSN number of idle slots instead of a DIFS at the beginning of the backoff. A node with a lower AIFSN thus has a greater probability of securing access to the channel earlier than another node with a higher AIFSN. Therefore, prioritized ACs are assigned a shorter AIFSN.
- **TXOP** is the time limit for consecutive transmissions once access to the channel is secured. After getting access to the channel, a node can transmit multiple packets as long as its TXOP does not expire.

The different values of these parameters for the four ACs defined in EDCA are shown in Table 2.5. One of the two different values of TXOP for AC_VI and AC_VO are chosen depending on the physical layer type in use.

2.3.2.5 HCF Controlled Channel Access (HCCA)

HCCA works in a PCF like fashion but while the CFPs start after a regular interval and alternate with CPs, an HCCA can be started at any time during a CP. This kind of CFP is called a Controlled Access Phase (CAP). The PC can transmit the Beacon frame initiating a CAP at any time and start a contention free communication with the nodes. Additionally, HCCA recognizes the access categories and, unlike PCF, is not limited to a round robin polling of the stations. Rather, it can identify the traffic streams and poll the nodes on a per session basis using a scheduling algorithm. Furthermore, the nodes can send information on their queue length to the AP which can be used to optimize the scheduling mechanism. Like EDCA, HCCA also uses TXOP so that once a node is polled, it can transmit a number of frames as long as the total transmission time does not exceed its TXOP limit.

In the DCF and EDCA modes, every node is required to sense the channel before a transmission and during the backoff periods. These access mechanisms primarily differ from the MACAW protocol in the carrier sensing mechanisms which are explained next.

2.3.3 Carrier Sensing Mechanisms

Carrier sensing means to detect activity in the channel or medium. IEEE 802.11 uses a "listen before talk" mechanism and each transmission is preceded by a phase when a node willing to transmit needs to sense the channel in order to detect any ongoing transmission. This task can be carried out by the physical layer or the MAC layer in one of the following two ways—

Physical Carrier Sensing Physical carrier sensing means to physically sense the channel to detect any ongoing transmission. This mechanism is carried out in the physical layer as requested by the MAC layer.

Virtual Carrier Sensing Virtual carrier sensing is the logical carrier sensing performed at the MAC layer so that certain handshake procedures can reserve the channel. The RTS, CTS and Beacon frames carry the duration for which the current handshake transmissions will keep the channel busy. Every node that receives such a frame, sets their NAV to this duration. The NAV is a counter that decrements its value after each idle time slot. Unlike the backoff counter (which is also a counter), NAV does not check the medium condition and decrements itself even if the channel is busy. Before any transmission, a node checks its NAV and transmits only if the NAV has a zero value. A non-zero NAV indicates that the channel is reserved and the node has to wait.

In the DCF and EDCA mechanisms, the backoff algorithms perform an important function which lessens the probability of successive collisions and allows the client nodes to adapt to changing traffic conditions. These backoff algorithms are explained in the next section.

2.3.4 Backoff Algorithms

Backoff algorithms are used to increase or decrease the contention window size in order to adapt the nodes in a BSS to changing traffic conditions. As the traffic load increases, the probabilities of transmission and collision also increase. The backoff algorithm increases the contention window size after each collision so that the nodes backoff more. As a result, the transmission probability decreases resulting in a lower collision frequency. With a lower traffic load, the contending nodes experience fewer collisions which lowers their contention window size, results in higher transmission probability and the network utilization increases.

The IEEE 802.11 standard defines only the Binary Exponential Backoff algorithm with a finite or infinite retry limit but many researchers have attempted to increase the

network performance and prioritize voice traffic by developing new backoff algorithms. The most notable backoff algorithms are discussed in this section.

2.3.4.1 Binary Exponential Backoff (BEB)

As the name suggests, the binary exponential backoff algorithm increases the collision window exponentially. After each failed transmission, the transmitting node doubles its contention window size. To avoid a too long backoff delay, the contention window is not increased after reaching a predefined maximum value the CWmax. After each successful transmission, the contention window is reset to its minimum value CWmin.

BEB comes in two flavors, infinite retry (or ∞-retry) and finite retry limit (or m-retry) BEB. In the infinite retry BEB mechanism, a packet is transmitted until success. On the other hand, in the finite retry BEB, the packet is dropped after transmitting it for a maximum of the *retry limit*. Whenever the packet is dropped, the contention window is reset to its minimum value.

As the transmission time of a packet is roughly proportional to its length, a longer packet uses the channel longer and is more prone to collisions and channel-errors compared to a shorter packet. Sometimes the next hop destination also determines the failure probability as some nodes are more likely to suffer from the hidden node problem than others. Additionally, in a multi-rate enabled WLAN, the modulation scheme and data transmission rate are selected for a packet considering the capability of the receiving node. Depending on the Signal to Noise and Interference Ratio (SNIR), the modulation scheme can result in a higher bit error rate. For the above mentioned reasons, some packets may face a higher probability of failed transmissions than others. Repeated failures increase the medium access delay of a packet. At the same time, as the packet is hogging the MAC layer, other packets behind it in the queue are subsequently delayed.

From a VoIP perspective, a late packet is no more useful than a lost packet. Therefore, it is better to drop a recurrently failing packet which is most likely to be already late and give the following packets a greater chance to reach their destinations in time. However, infinite retry BEB does not drop any failing packets and allows for channel hogging. As a result, by the time when one recurrently failing packet is successfully transmitted, all the other packets following it have become useless anyway. For this reason, voice applications perform better with the m-retry BEB.

2.3.4.2 Fair Backoff

If the backoff window size of one of the nodes in a collision domain is less than the others, the node waits for less time before transmitting and has a greater chance in

winning access to the channel. Once the node wins by transmitting successfully, it resets its contention window in the BEB algorithm to the minimum value which increases its probability to succeed in the following round. If the node's traffic load is sufficient to use the total available bandwidth, the other nodes may starve which leads to unfairness. An extension of the BEB algorithm is proposed by Bharghavan *et al.* [70] where the MAC frames contain the current contention window so that every node can copy it to their own contention window after a successful transmission. In this mechanism, every node maintains the same level of backoff ensuring fairness and avoiding channel hogging. But the success of a transmission is only confirmed by an ACK frame and therefore, this method requires each node to receive and process every transmission and keep track of the contention window that every packet was transmitted with until they are acknowledged. Thus, high processing power and memory capacity are required.

2.3.4.3 Slow-Decrease Backoff

In a fairly dense network, collisions can occur at a high rate. As the nodes face collisions at a higher rate, they increase the contention window to reduce the number of collisions at the cost of a longer delay in medium access. However, according to the standard BEB algorithm, the contention window is reset to its minimum value after a successful transmission. The contention window is also reset when the number of retries exceeds the retry limit in case of the finite retry BEB algorithm. When the network traffic load remains the same, this sudden decrease in contention window increases the risk of further collisions by the same nodes. Wu *et al.* [75] proposed an extension of the BEB algorithm where the contention window is slowly decreased. After a successful transmission, the contention window is set to half of the current contention window or the minimum contention window, whichever is greater. If the retry limit exceeds, however, it indicates that the network is overloaded and the contention window is kept as it was, i.e., the maximum contention window size.

2.3.4.4 Quadratic Backoff

The standard BEB algorithm doubles the contention window after every collision or failed transmission which works well with a small network and low traffic load. If the collisions occur due to a sudden increase in the network size, the algorithm fails to cope with the fast changing traffic load. Abu-Tair *et al.* [76] proposed an adaptive backoff algorithm where the collision rate is used to select one of two schemes. When the collision rate is smaller than a predefined threshold, the standard binary exponential backoff algorithm is used. When the collision rate is higher than this threshold, a

quadratic algorithm is used where the contention window is squared instead of doubling it. After a successful transmission, the contention window is reset to its minimum value as done in the standard BEB algorithm. Simulation results show that this proposed scheme offers a higher throughput and channel utilization as well as a lower delay and loss in medium access. However, an adaptive backoff algorithm is only necessary if the network size changes frequently and sudden increases in the traffic load are expected.

2.3.4.5 Tender Backoff

Due to the exponential nature of the BEB algorithm, the contention window size increases sharply (i.e., the rate of increase in contention window size rises fast) with recurring collisions, which can lead to an unnecessarily long delay if such collisions are occasional. Lin and Pan [77] introduced a tender backoff algorithm that adds two more backoff states to reduce such delays. Here, a contention window size of 384 is inserted between 256 and 512 and another contention window size of 768 is inserted between 512 and 1024. Similar to [75], the contention window is halved after a successful transmission. However, any transition to and from the newly added states does not involve doubling or halving the contention window. This mechanism makes changes in the contention window size smoother leading to less delay in the medium access in some cases. A higher throughput than both BEB and [75] is shown to have achieved by reducing the unnecessary delay.

2.3.4.6 Double Increment Double Decrement

Chatzimisions *et al.* [78] extended [75] where the contention window is doubled or halved after a failed or successful transmission, respectively. However, unlike [75], the packet is not dropped even if the retry limit exceeds making the backoff algorithm an infinite retry algorithm which is termed as Double Increment Double Decrement (DIDD) algorithm. This mechanism reduces the packet loss at the cost of a longer delay in the medium access. The performance gain is analyzed using a Markov chain.

A detailed analysis of each of these algorithms is outside the scope of this thesis. However, Vukovic *et al.* [79], Vukovic and Smavatkul [80], Bender *et al.* [81] presented such analyses including the delay and loss. IEEE 802.11 defined only the binary exponential backoff algorithm with a finite or infinite retry limit. This is the reason why all commercial devices comply to only this algorithm as they support only 802.11a, 802.11b or 802.11g. Therefore, any practical network employs the DCF based medium access mechanism with the BEB based backoff.

With the understanding of the 802.11 MAC layer specifications, we describe the supported physical layers briefly in the following section.

2.3.5 Physical (PHY) Layers

The IEEE 802.11 standard defines a number of physical layer specifications, e.g., FHSS, DSSS OFDM and Infrared which are described briefly in this section.

Frequency Hopping Spread Spectrum (FHSS) A FHSS based physical layer transmits on a chosen frequency for a very short time and then switches to another frequency using a predefined frequency hopping pattern that is known to both the transmitter and receiver. This switching lowers the impact of high energy interference in a narrow band as well as the mutual interference of multiple FHSS transmitters positioned close to each other. In FHSS, a 2.4 GHz ISM (Industrial, Scientific and Medical) band is split into multiple 1 MHz bands. The transmitter switches among these bands after operating for less than 400 ms in each band. The hopping patterns are defined as 3 sets, each consisting of 26 hopping sequences. The sets are defined in a way that minimizes the mutual interference among co-located access points. The chosen frequency is modulated using a two level Gaussian Frequency Shift Keying (GFSK) at 1 Mbps or four level GFSK modulations at 2 Mbps data rate. FHSS is robust to interference, cost effective and simplistic but the commercial implementations of FHSS are less successful and common than DSSS.

Direct Sequence Spread Spectrum (DSSS) DSSS is the most successfully and commonly used physical layer technique. Apart from IEEE 802.11 WLANs, it is also used in cellular networks and Global Positioning Systems (GPS). In DSSS, the data stream is multiplied by a Pseudo Random binary Number (PRN) sequence (also called the "chipping sequence") of a higher bit rate (called the "chipping rate"). In the PRN, the difference between the numbers of 1s and 0s in the sequence is less than or equal to 1 and the autocorrelation function of the sequence is close to 0 everywhere except at the shift 0. These properties of the PRN are similar to white noise that has a flat line spectrum. Therefore, the multiplication of the data stream by the PRN flattens the spectrum of the resulting signal. The multiplied data is then transmitted using any standard modulation technique. At the receiving end, the received data is demodulated but the original data after the multiplication can not be recovered unless the chipping sequence is known. The received data is multiplied by a PRN generated by the receiver to retrieve the originally transmitted data.

Orthogonal Frequency Division Multiplexing (OFDM) uses closely spaced subcarriers which are orthogonal to each other. Carrier spacing is kept equal to reciprocal of symbol period so that the interference contributions of the subcarriers sums to zero in the signal period. Data are divided into a number of data streams, one per subcarrier, and modulated using a conventional modulation scheme at a low symbol rate. OFDM exhibits high spectral efficiency, resilience to radio frequency interference and lower multi-path distortion. OFDM can cope with severe channel condition, e.g., attenuation of high frequencies in copper wire and frequency selective fading in multipath scenario, without using complex filters. This is why it is heavily used in digital television (DVB-C2 standard), broadband internet (using ADSL and VDSL), wireless networking using Wi-Fi (IEEE 802.11a,b,n), WPAN (IEEE 802.15.3a), mobile broadband wireless (IEEE 802.20) and mobile-WiMAX (IEEE 802.16e).

Infrared (IR) Infrared PHY provides a 1 Mbps or, optionally, 2 Mbps data rate. The IR based physical layer uses nearly visible light of 850–950 mm wavelength for signaling which is similar to common consumer devices like remote controls. But this signal is not required to be directed, i.e., the sender and receiver do not need to be in a clear line of sight. In such cases, the reflected light can be utilized. But in environments where neither reflecting surfaces, nor line of sight is available, IR based devices may suffer from a reduced transmission range. The standard transmission range is 10 m but using highly sensitive receivers, the range can be increased to 20 m in a controlled environment. Infrared is not popular in the industry and there is no successful commercial implementation.

2.4 Performance of Wireless Local Area Networks (WLAN)

The performance of VoIP applications in WLANs is largely determined by the performance of the medium access control mechanism. In this section, we analyze the WLAN performance briefly.

2.4.1 Wireless Network Performance Measures

Researchers used different measures to determine the WLAN performance. The most commonly used performance measures are introduced in this section.

Throughput Throughput is the number of bits transmitted per second by the medium access layer. The saturation throughput is the throughput achieved by a saturated network, i.e., a network where every node has some packets to transmit

at any given time. Saturation throughput is one of the most commonly used performance measures in the literature as it gives an idea of the upper limit of the achievable throughput. However, the collision rate is very high under the saturated condition which incurs an overhead in terms of the channel idle time and the time wasted in collided transmissions. A little higher throughput than the saturation throughput can be achieved at near saturated condition which is shown in [2, 82]. Throughput in an unsaturated network was estimated in [19, 83, 84] where the nodes have no packet to send for a considerable amount of time. It is a useful measure for applications where a large amount of data needs to be transferred, for example, in FTP download or peer-to-peer file sharing. However, throughput is not an adequate measure of the performance of QoS aware applications like VoIP or IPTv.

Goodput Goodput is defined as the number of useful bits transmitted per second at the application layer and, therefore, the bits transmitted by the routing or link control algorithms are not considered. The goodput is a better indication of the usefulness or performance of a network, but similar to the throughput, it is still not an adequate estimate of the performance of QoS aware applications.

Medium Access Delay The delay in the medium access is one of the key performance measures for delay sensitive applications. Due to the absence of a proper QoS mechanism in the IEEE 802.11 standards, the delay and delay jitter can not be controlled which can result in severe quality degradation for QoS aware applications, e.g., voice and video traffic. The inverse of the medium access delay is the service rate of the queue, therefore, the medium access delay also dictates the delay and loss in the queue. This is why the exact nature of the delay variation with changing network condition and traffic load must be analyzed with great care before deploying wireless networks.

Medium Access Loss When a transmission suffers from a collision or channel error, the received packet is dropped by the receiving station. The sender then waits for acknowledgment until a predefined time and retries to transmit the packet. Depending on the backoff algorithm, the packet may be dropped after retrying for a predefined number of times (the "retry limit" in m-retry BEB). The ratio of the number of packets dropped by the MAC layer to the number of packets the MAC layer attempted to transmit is called the medium access loss.

Queuing Delay The medium access delay defines the service time for the queue. Depending on the queuing model, it is possible to determine the delay and loss in the queue. The queuing delay is the time from when a packet is enqueued until

the time it is removed from the queue by the MAC layer. If the service rate of a queue becomes lower than its arrival rate, specially with deterministic traffic like constant bit rate VoIP codecs, the queue starts growing and the queuing delay increases.

Queuing Loss The queue can only store a fixed number of packets. In addition to voice packets, it also needs to store background traffic packets, routing packets and link control packets. Any packet that arrives at an already full queue is dropped and the ratio of dropped packets to the total number of packet arrivals is called the queuing loss. The delay and loss in the queue play a crucial role in defining the voice quality of a network.

Delay Jitter Variation in the end-to-end delay experienced by packets is called delay jitter. As mentioned earlier, the voice packets are needed at the receiving station at a regular interval which the network can not guarantee and a dejitter buffer is used to reduce the delay variation. However, too much jitter in the end-to-end delay can be beyond the capacity of the dejitter buffer which leads to a high delay in the buffer. Therefore, the delay jitter plays a vital role in determining the voice quality.

Dejitter Buffer Loss At any given time, if a packet which is needed by the VoIP application is not available in the dejitter buffer, the packet is considered lost and it will cause a silent pause in the played out sound. If the packet arrives later, it can no longer be used, and the dejitter buffer will drop the packet. The amount of packets dropped by the dejitter buffer is called the dejitter buffer loss. With a longer dejitter buffer, the probability of a packet to be available on time increases. On the other hand, a shorter dejitter buffer means only a few packets can be buffered and the probability of packet availability may decrease depending on the network performance. If a dejitter buffer becomes full, which is more likely with a shorter buffer length, further arrivals may be dropped which also increases the dejitter buffer loss. Therefore, a longer dejitter buffer helps in reducing the dejitter buffer loss.

Dejitter Buffer Delay The delay a packet faces in the dejitter buffer is called dejitter buffer delay. It is the interval starting from the time when a packet reaches the dejitter buffer and ends at the time when it is played out. To reduce the dejitter buffer loss due to overflow, the buffer has to be sufficiently long to accommodate a good number of voice packets. Moreover, to reduce the buffer loss by dropping a packet that is too late to play out, initially a pause is inserted so that the buffer can fill up to a required level. The longer the pause is, the higher is the probability that a packet reaches the dejitter buffer before its play

out deadline. However, the pause is inserted in the end-to-end path of the voice stream and incurs delay impairment to the voice quality. Therefore, to reduce the end-to-end delay, it is essential to keep the buffer length short. A long dejitter buffer reduces the dejitter buffer loss but also increases the dejitter buffer delay while a short dejitter buffer increases the dejitter buffer loss but decreases the dejitter buffer delay. The choice of an appropriate length for the dejitter buffer is specific to the implementation of the dejitter buffer.

A number of works in the literature investigated the performance of the IEEE 802.11 networks which is of great essence in order to analyze the voice performance of VoIP in such networks. We review a number of such works in the next section.

2.4.2 Capacity Estimation of WLANs

Bianchi *et al.* [85] studied WLAN performance using simulations and also presented an adaptive mechanism which dynamically controls the contention window to improve the network performance. Assuming an ideal channel and no hidden terminals in [85], the wireless stations estimate the total number of stations n_s and the length of the DCF handshake t_s. The initial contention window is set dynamically to $n_s\sqrt{2t_s}$. The estimated network size n_s is estimated by observing the number of busy slot times which can deviate significantly from its actual value. To ensure a smooth but continuous change of this estimate, a further modification was proposed where the initial contention window was set to $(1 + h/\sqrt{n_s})\sqrt{2t_s}$. Here, h is a constant which assumed a value of 2 in the experiments. Additionally, the RTS/CTS based handshake instead of basic DCF was suggested for use with packets that are larger than 6000 bits.

Bianchi [2, 86] employed a Markov model to analyze the saturation throughput of a single hop WLAN. This Markov model is shown in Fig. 2.8 where the circles represent medium access states. The arcs represent the state transitions with the probability shown as their label. The states in each row correspond to a retry stage. The retry stages are labeled as $0, 1, \ldots, m$ and there are maximum m such retry stages. The initial contention window size is W which is doubled after each failed transmission until a maximum of W_m is reached. The states in each column correspond to the value of the backoff counter that is set to a uniform random number from $[0, 1, \ldots, W_i - 1]$ when the node reaches the i-th backoff state. The label (i, j) of each state, therefore, corresponds to a (*retry stage*, *counter value*) pair. A transmission is attempted when the counter reaches zero, therefore, transmissions can occur only from the $(i, 0)$ states for $i \in 0, 1, \ldots, m$ which are all in the first column. A transmission can suffer from a collision with probability p. A collided frame can not be recognized by the receiving station and the receiving station does not acknowledge. Whenever a transmission is acknowledged, with probability $1 - p$, the transmission is taken to be successful and

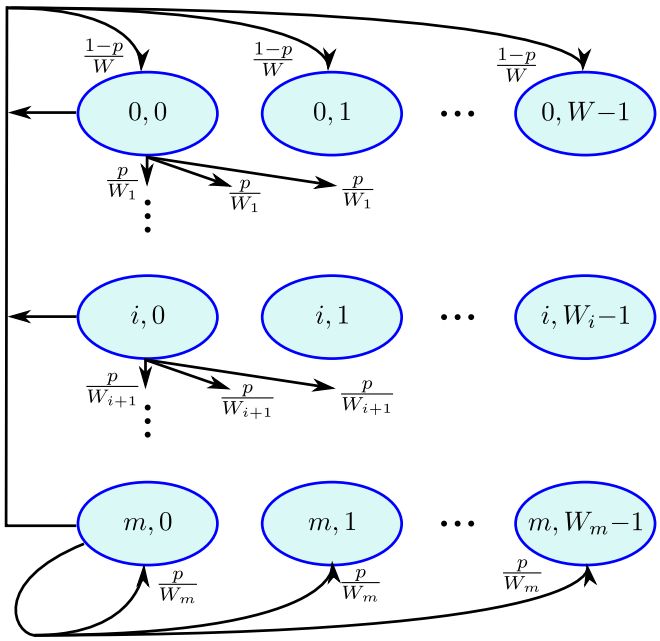


Figure 2.8: Markov model to represent DCF based medium access utilizing ∞-retry BEB as used by Bianchi [2].

system resets to the 0-th retry stage. If a transmission is not acknowledged within an *ACK Timeout*, the system goes to the next retry stage and doubles its contention window. Once the last (m-th) retry stage is reached, the contention window is not increased any more and the system recursively stays there (due to the infinite retry limit scheme) until a transmission succeeds.

The above model considers ∞-retry BEB only and ignores the impact of the retry limit, imperfect channel and capture effect. Moreover, the unsaturated traffic condition was not considered although real world networks operate under unsaturated condition for a considerable amount of time [19]. Despite its simplicity, the model can be used to estimate the performance of saturated WLANs with a reasonable accuracy for both basic and RTS/CTS based DCF. It was pointed out in [2] that as the offered load increases, the throughput grows towards a maximum value. With any further increase in the offered load, the throughput decreases and settles at an asymptotic value and the network remains in the saturated condition. Therefore, the maximum throughput is achieved at a near saturated condition but not at the saturated state.

For the ease of analysis, the collision and transmission probabilities are assumed constant regardless of the number of retransmissions. This assumption is deemed more accurate for larger networks and contention window sizes. A unique solution can be found by numerically solving a system of non-linear equations in terms of the transmission and collision probabilities. The results show that the maximum

throughput achieved by the basic and RTS/CTS mechanisms are very close. The model in [2] was further generalized by Bianchi and Tinnirello [87] to assist in modeling a wider range of backoff algorithms.

Ziouva and Antonakopoulos [88] presented a similar Markov model for the performance analysis of a DCF based saturated WLAN using some modified assumptions. The primary difference from [2] is the inclusion of a busy medium condition. With the busy medium condition, the probability of the channel being idle is included whenever the backoff counter is decremented. The impacts of the retry stages and the collision window sizes were also investigated. It was identified that the number of retry stages and the contention window size strongly affect the throughput of the basic DCF while RTS/CTS DCF is more robust to these changes.

Wu *et al.* [89] extended the Markov model in [2] for m-retry BEB where a packet is dropped after $m + 1$ transmissions. Additionally, an extension to the existing DCF based medium access was proposed which was named as DCF+. The proposed mechanism was used when a sender transmitted data using the basic DCF and the receiver also had a packet to send to the sender. Upon reception of the data frame, the receiver transmitted a special ACK frame that was used to set the NAV similar to that in a RTS packet in a RTS/CTS handshake. In response to such an ACK frame, the sender transmitted a CTS packet and the rest of the transmissions of an RTS/CTS handshake follows. Since both the first ACK and the CTS frames are used to reserve the channel, collisions do not occur in this two-way data interchange. Performance gain is achieved by removing contention for the second data frame. Although the availability of a data frame at the receiver is unlikely for background traffic, it is common in two way voice connections. Therefore, this mechanism can save a good amount of idle backoff time in each voice stream and at each hop. Simulations in NS-2 showed a considerable throughput increase especially in dense networks.

Chatzimisios *et al.* [90, 91, 92, 93] extended the model in [89] to incorporate the impact of imperfect channel condition as Bit Error Rate (BER). In addition to throughput, delay and loss in the medium access were also investigated. It was noted that the packet drop probability increases rapidly until BER became 10^{-4} and reaches 1.0 at a BER of 10^{-3}. But the impacts of power capture and unsaturated traffic were not considered in [88–93].

Barowski *et al.* [83] extended Ziouva's model [88] and Liaw *et al.* [84] extended Wu's model [89] to consider unsaturated traffic for ∞-retry and m-retry BEB, respectively. Both approaches used a new state in the Markov model to represent the empty queue state. The transitions in and out of this state were derived by using standard queuing analysis. But the impacts of the imperfect channel and capture effect were ignored in [83, 84]. Additionally, [83] considered the energy consumption and proposed a three

dimensional model for relay-only nodes, that is, the nodes which relay packets but do not generate them. This model was proposed for a multihop network where some of the nodes relay packets that were generated by others. However, despite some guidelines on the three dimensional model, no performance measures were derived for this model. It was assumed that at any given time, the number of data source nodes was known to a relaying node, however, no guideline was provided to calculate the number of traffic generating nodes. Queue length was used as the third dimension of the state space in addition to the retry stage and backoff counter value. It is well known that in multihop WLANs the traffic arrival rate varies giving rise to a critical zone [94]. But the progressively increasing traffic load or critical zone were not considered in [83]. We address these issues in Chapter 4.

Daneshgaran *et al.* [19] merged the models in [2, 83] and additionally considered imperfect channel [90] and capture effect over a Rayleigh fading channel [95, 96]. It was demonstrated that throughput assumes a linear function of the packet arrival rate until a critical packet arrival rate is reached. For small packet arrival rate λ, the queue fullness and transmission probability approaches zero. Throughput for such a small λ was shown to be close to $n_s l_p \lambda$ where n_s and l_p are the network and payload sizes, respectively. A critical arrival rate was also determined which denotes the upper limit of this linear approximation.

Table 2.6: Comparison of Markov model based analyses of 802.11 WLANs employing DCF based medium access.

Specific Study	Unsaturated Traffic	Retry Limit	Imperfect Channel	Power Capture
Bianchi *et al.* [85]	No	Infinite	No	No
Bianchi [86]	No	Infinite	No	No
Bianchi [2]	No	Infinite	No	No
Bianchi and Tinnirello [87]	No	Finite	No	No
Ziouva and Antonakopoulos [88]	No	Infinite	No	No
Wu *et al.* [89]	No	Finite	No	No
Chatzimisios *et al.* [91]	No	Finite	Yes	No
Chatzimisios *et al.* [92]	No	Finite	Yes	No
Chatzimisios *et al.* [93]	No	Finite	Yes	No
Chatzimisios *et al.* [90]	No	Finite	Yes	No
Barowski *et al.* [83]	Yes	Infinite	No	No
Liaw *et al.* [84]	Yes	Finite	No	No
Daneshgaran *et al.* [19]	Yes	Infinite	Yes	Yes

The Markov model base analyses of WLANs offer a very accurate model of the medium access mechanism with reasonable mathematical complexity. The existing studies [2, 19, 83, 84, 86, 87, 89–93] explored the impacts of imperfect channel, capture effect, m-retry BEB and unsaturated traffic in different works. An ideal channel is only possible in theory, and a real channel will always have noise and interference. The impact of an imperfect channel was considered in [19, 91–93]. Power capture can bring a positive effect to the network performance and this feature is now available in almost every 802.11 device. Power capture was considered in [19]. Most networks operate in the unsaturated condition for a considerable amount of time and analytical models should consider both saturated and unsaturated traffic. Unsaturated traffic was modeled in [19, 83, 84]. Finite retry backoff algorithm drops a packet that has suffered from a predefined number of transmission failures. For delay sensitive applications like VoIP, such a packet is most likely already be too late to meet its dejitter buffer deadline. Thus, removing a repeatedly failing packet gives the following packets a greater opportunity to make their destinations in time. This feature is very crucial for delay sensitive applications like VoIP as explained before. Finite retry was considered in [84, 89, 91–93]. A comparison of these works is listed in Table 2.6. Although all the above mentioned factors are present in a real environment for VoIP calls, no existing model considers all of these factors concomitantly. We develop a Markov model considering all the above factors in Chapter 3. Moreover, no existing analytical model considered multichannel, multihop WLAN for VoIP traffic. Although [83] provided some guidelines on multihop WLANs, the model is incomplete in the sense that no detailed analysis, performance measures or capacity estimates were provided. Additionally, the increasing traffic load and critical zone formulation near the access point were not considered. We model a multichannel, multi-interface, multihop WLAN in Chapter 4. With the understanding of WLAN performance, we review the existing works which estimated the call capacity of WLANs in the next section.

2.4.3 Obstacles to Provide VoIP over WLAN

Although VoIP can be used in any IP based network, certain challenges are faced in implementing voice networks over IEEE 802.11 WLANs. In this section, we discuss the most important and challenging issues which network administrators should consider before designing and deploying such networks. Traditional PSTN offers a good voice quality which is also referred to as *toll quality*. PSTN is able to maintain its voice quality throughout a call since it reserves a port of each switching center for every

call. But such reservation techniques are not applicable to IP networks and, in turn, to VoIP. As a result, the voice quality in VoIP calls is vulnerable and subject to high end-to-end delay, delay jitter and packet loss [44]. In the following, we discuss the parameters which are likely to impact the performance of voice services.

End-to-end Delay A high end-to-end delay incurs a high voice quality impairment. The end-to-end delay is calculated from the mouth (talker) to the ear (listener) and is a sum of various delay components. The encoder picks a number of voice frames, encodes and puts them in a UDP packet. Therefore, the total length of voice that is carried in a single packet is incurred as the packetization delay in the end-to-end path of the voice stream as well. Whenever the queue is non-empty, the packet faces a queuing delay. Transmitting the packets through the network incurs the majority of the delay. Finally, the dejitter buffer incurs a delay before the packets are picked by the decoder. The packetization, encoding and decoding processes incur some minute delays for their operation but are minimal with today's high speed computing systems. The dejitter buffer delay is specific to the dejitter strategy, i.e., static or dynamic, and its length but the choice of codec can play a vital role in determining the end-to-end delay. For instance, G.723.1 uses 30 ms voice frames compared to the 10 ms voice frames of G.711 and G.729. Therefore, with the same level of aggregation, G.723.1 incurs a higher packetization delay. G.729 incurs an additional 5 ms of look ahead delay compared to 0 ms and 7.5 ms look ahead delays for G.711 and G.723.1, respectively. On the other hand, G.711 generates an 80 bytes payload per voice frame, compared to the 10 bytes of payload per frame with G.729, and results in longer transmission time and longer MAC delay.

A high delay increases the echo impairment to the voice quality [44]. Moreover, the interactivity between the talker and the listener is hampered if the end-to-end delay is very long. The delay limits of one way transmission for different applications defined by the ITU-T G.114 [97] are shown in Table 2.7. For delays above 150 ms, the listener perceives hesitation in the talker's voice due to the delay in response. When the delay is more than 400ms, the voice call is deemed to be interrupted by the listener. However, for highly interactive calls and in the presence of packet loss, users may face inconvenience even with an end-to-end delay of less than 150 ms. The ITU-T E-model considered the impact of the delay as delay impairment factor and as shown in Fig. 3.1 on page 98, the delay impairment increases slowly until the end-to-end delay becomes higher than 160 ms and rises fast for higher values.

Delay Jitter Although the encoder at the talker end sends packets at a fixed rate, the

Table 2.7: One way delay limits according to ITU-T G.114

End-to-end Delay (ms)	Interactivity of Tasks
0–150	Acceptable for most user applications.
150–400	Acceptable but some data and voice applications may experience service quality degradation.
Above 400	Unacceptable for general network planning purposes (except international connections over satellite).

packets arrive at the destination at an uneven rate due to network congestions and bottlenecks. The dejitter buffer tries to reduce this jitter and smooth out the traffic arrival rate at the cost of some additional delay. But if the delay jitter is very high, some packets may not be available at the destination when they are expected to be played out, thus causing an audible gap in the voice stream. If the packet arrives at the destination at a later time, it is dropped. Subsequent packet drops due to a high delay jitter incurs severe voice quality degradation. Depending on the delay budget, a 30–75 ms of delay jitter can be admissible [44]. The delay jitter is usually very specific to instantaneous traffic load and network congestion scenarios, and is independent of design parameters like the codec choice and network configuration.

Packet Loss Packet loss can occur in an overloaded network or due to instantaneous congestion. In an overloaded network, many nodes try to transmit packets simultaneously causing frequent collisions. A collided packet is retransmitted, but the medium access layer drops a packet when it has retransmitted the packet beyond an acceptable limit. When the medium access delay becomes high, the service rate of the queue becomes low and the queue becomes prone to getting full and any further packets that arrive at an already full queue are dropped. In addition to the queue and MAC layer, packet loss can also occur in the dejitter buffer as mentioned before.

When VoIP is used over WLAN, these vulnerabilities are magnified due to the inherent incapabilities of the IEEE 802.11 standard. We discuss the major obstacles in ensuring voice quality over WLANs in the following.

2.4.3.1 Lack of a QoS Assurance Mechanism

The IEEE 802.11 standards do not have a proper quality of service assurance mechanism. Although 802.11e focused on improving the performance of voice and video traffic, it only introduced service differentiation to give voice and video traffic a better

share of the channel resources. But this mechanism does not guarantee an upper limit of end-to-end delay or loss. Moreover, the 802.11e standard is still not implemented in the commercial products. The most successful commercial standards are 802.11b and 802.11g which do not even differentiate between traffic types. Since the voice quality is impaired by any delay, loss and delay jitter, voice calls can suffer from severe quality degradation under a high load.

2.4.3.2 Low Bandwidth

Wireless standards offer only a fraction of bandwidth that is available in the wired networks. Now-a-days, Ethernet connections offer up to 1 Gbps bandwidth while the maximum theoretical bandwidth in WLANs is only 54 Mbps. Additionally, due to the collision avoidance nature of DCF based medium access, a considerable amount of time is wasted in the backoff state. Furthermore, control packets and frame headers are always transmitted at the basic rate of 1 Mbps. This is why only a fraction of the theoretical bandwidth can be utilized in reality.

2.4.3.3 Short Transmission Range

Wi-Fi offers only a short transmission range with poor obstacle penetration. This limits the coverage of access points and wireless routers. For typical networks in corporate offices, university campuses and hospitals, etc., the offered short transmission range is deemed to be adequate as many access points can be carefully placed. But providing coverage in semi-utilized public hotspots like cafes, train stations or shopping malls requires a trade off between the cost and the coverage benefits.

2.4.3.4 Sharing Communication Channel with Other Traffic

In wired phone lines, e.g., PSTN, the ports/channels are reserved for voice communication only. However, the ubiquitous acceptance of VoIP is primarily due to its capability of using an existing network and sharing the network with other applications. Therefore, some background traffic is expected with voice calls which causes VoIP to perform worse [6]. To avoid such problems and with the availability of multiple channels in the 802.11 standards, some wireless channels can be kept for the exclusive use of voice communications while the other channels can be used in a shared mode.

2.4.3.5 Poor Handoff Performance

As mobile devices tend to roam about, frequent handoff of mobile nodes from one BSS to another is one of the major problems that degrade performance of VoIP over Wi-Fi.

Each handoff from one access point to another incurs a delay in the end-to-end path of the voice stream [98]. Additionally, packet loss during a handoff incurs an additional impairment to the voice quality. GSM networks reserve some control channels to manage resource allocation but 802.11 standards do not use such a mechanism. Smooth hand over is a great challenge for VoIP over WLAN.

2.4.3.6 Power Saving for Mobile Devices

The convenience of mobility comes at a price as the mobile hand-held devices have a limited power supply. Saving battery power while maintaining the voice quality is another challenging issue.

Due to the inherent features of the IEEE 802.11 standard, it is a challenging task to maintain voice quality over WLANs. Especially, during peak hours, a high call volume is expected which suddenly increases the network traffic load. If these networks are not carefully designed, the sudden increase in network traffic will result in call jitter or call quality variation for all ongoing calls. Call drop and rejection can also occur when the network becomes overloaded. To protect against such drawbacks, the WLANs must be carefully designed considering expected call volume, network configuration and, most importantly, voice quality requirements. Network administrators should carefully estimate the voice capacity of these networks considering standard voice quality measures before deployment. A number of researchers investigated the voice capacity of WLANs which are discussed in the following section.

2.5 Voice Capacity Analysis of WLANs

The voice capacity of WLANs intrigued many researchers and a number of works in the literature investigated the performance of IEEE 802.11 networks with voice traffic. We review the most notable of these works in this section.

A good number of works investigated the voice capacity of 802.11 WLANs using analytical, simulation or testbed based studies. Simulation based studies are very helpful in validating existing models and testing particular configurations of a network. They can also highlight the general trends of goodness or badness of a network feature. But such studies are quite rigid in providing an exact analysis of a network since only a fixed set of parameters and their combinations can be used in a simulation or testbed while in network design and planning, network designers need to consider a wide variety of parameters, devices and configurations to choose from. Simulation and testbed based studies offer only limited insight since the expected performance after changing some network aspects can not be determined easily, without performing all the simulations again, which is possible only by using an analytical model. Analytical models parameterize the inherent characteristics of the network aspects which can easily be altered to determine the network performance under changed conditions. Additionally, all the insights gained from a simulation based approach are also provided by an analytical model. A number of the most popular and widely accepted studies on voice performance over WLANs are reviewed here.

2.5.1 Single Channel, Single hop Networks

Garg and Kappes [13] investigated the voice capacity with the G.711, G.729 and G.723.1 codecs using testbeds of 802.11a and 802.11b WLANs. The connection of the RTP based voice sessions were monitored using a commercial tool through measurements of the delay, delay jitter and packet loss. It was identified that the idle time spent by a node per frame is a non-linear function of the network size. However, this delay was expected to be small in a saturated network and was approximated by a delay in a network of only two saturated nodes, which was empirically measured to be 8.5 idle slots. Although the channel remains idle for a small amount of time in a saturated network, the backoff delay per node increases with the network size. As the number of stations increases, the probability of collisions also increases, which results in a high backoff and medium access delay. Despite the complexity of the DCF based medium access mechanism, a very simplistic throughput usage estimation was used in [13] to determine the voice capacity. Despite the low computational complexity of the model, the capacity estimation by throughput usage can not ensure voice quality which is affected greatly by both end-to-end delay and loss. The effects of the aggregation level, error prone channels or capture effect were also not considered in [13]. It was identified that the data rate was affected by a poor signal strength resulting from spatial distribution but no guideline for its measure was mentioned. The capacity of

802.11a (54 Mbps) was found to be five times higher than 802.11b (11 Mbps). The proportional increase in the call capacity with data rate resulted from the oversimplifying assumption of constant backoff delay.

Hole and Tobagi [14] carried out a similar study in an IEEE 802.11b testbed of Cisco 7960 phones using the G.711 and G.729 codecs. MOS was used to determine the voice quality of the VoIP calls. However, like [13] the method presented in [14] also used some oversimplified assumptions that include—

- No collisions occur
- No channel errors occur
- No packet loss in the dejitter buffer

While these assumptions simplify the analysis, they also weaken the validity of the model. It was argued that the idle time was minimal when the network traffic load approached its capacity since the devices count down the backoff slots concurrently. An idle backoff period of $\frac{CWmin}{2}$ number of slots was assumed where $CWmin$ was the initial contention window. Moreover, it was assumed that half of the traffic was transmitted by the AP which is valid only in single hop networks. In multihop networks, the client nodes forward the data of other nodes to and from the AP and the majority of the transmissions are carried out by the client nodes. It was claimed that the effect of collisions on the call capacity was minimal and the idle backoff time approached zero in a saturated network.

The use of short PLCP preamble [66] was found to increase voice capacity by 25–50% in [14]. But the utilization of this short preamble can not be relied upon since it was optional in the standard and may not be inter-operable between devices of different vendors. G.729 supported more calls than G.711 in the simulations due to its higher level of voice compression. The impact of channel errors was incorporated as the bit error rate. It was identified from simulations that no call can be supported with $BER \geqslant 10^{-3}$. Both [13] and [14] ignored the impacts of capture effect and unsaturated traffic. Moreover, the analytical methods in [13, 14] did not use any standard voice quality measure.

Medepalli *et al.* [99] presented an analytic and simulation based investigation to determine voice capacity of the IEEE 802.11a, 802.11b and 802.11g single hop networks. Voice capacity was suggested to be a function of the channel bandwidth, codec packetization interval and packet size. In the 802.11g WLANs, capacity was adversely affected by RTS/CTS based handshake. The G.711 codec was used to generate voice packets following the artificial speech synthesis standard ITU-T P.59 [100]. Conversational speech was recognized as sequences of one-way talk-spurt, pause, double-talk

Table 2.8: Temporal parameters in conversational speech.

State	Duration (sec)	Rate (%)
Talk spurt	1.004	38.53
Pause	1.587	61.47
Double talk	0.228	6.59
Mutual silence	0.508	22.48

Table 2.9: Distribution of conversation states.

Conversational State	Mean of Exponential Distribution (ms)
A talking, B silent	854
Both silent	456
Both talking	226
B talking, A silent	854

and mutual silence periods in [100]. The average duration and rate of these events are shown in Table 2.8. Cumulative distribution functions of the talk-spurt and pause durations were approximated using exponential functions. A four state Markov chain was suggested in [100] to model the talking and silent states of the talker and listener as shown in Fig. 2.9. Medepalli *et al.* [99] used independent and identically distributed exponential random variables to model the duration spent in each of the states. The expected mean of these variables are shown in Table 2.9. But no voice quality assurance technique was used which limits the usefulness of the model.

Voice capacity was approximated using the resource consumption per call and the channel time was used as a measure of consumed resource. However, the collisions were assumed to be independent only to determine the number of retries from a Truncated Geometric random variable and the effects of imperfect channel, capture effect and aggregation level were ignored. Unsaturated traffic and delay and loss in the queue were also not considered. The call capacity was compared between 802.11a, 802.11b and 802.11g WLANs and in the presence of video traffic. Video packets being large in size consumed more channel resource and adversely affected the voice capacity.

Anjum *et al.* [15] investigated the voice performance in WLANs in the presence of background data traffic. A testbed of 802.11b WLAN was used with commercially available off-the-shelf devices like the Orinoco Gold and Agere Orinoco AP-1000. MGEN[16] and VGEN (in house tool developed in [15]) were used to generate voice-stream-like UDP data traffic and Ethereal[17] was used to record the network performance measures.

[16]http://cs.itd.nrl.navy.mil/work/mgen/index.php
[17]http://www.ethereal.com

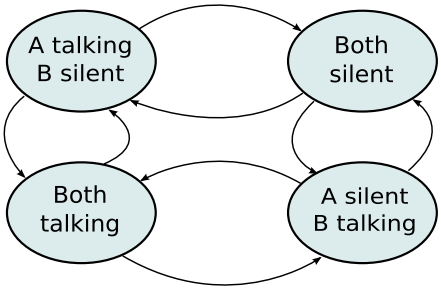


Figure 2.9: Markov model for conversational speech according to ITU-T P.59.

A queuing mechanism based on Backoff Control Priority Queuing (BC-PQ) [101] was used in [15] to improve the voice capacity. This mechanism was implemented in the AP buffer and was used to prioritize voice packets by traffic differentiation and backoff mitigation. Here, voice packets were given priority over data packets and any voice packet that arrived in the queue was put at the head of the queue. Additionally, the AP did not backoff at all for the voice packets. The prioritization of the voice streams improved the voice capacity. A packet loss of 0.2% and an end-to-end delay of 200 ms were used as the thresholds of acceptable voice quality. However, due to the rigid prioritization, voice traffic can completely hog the channel causing the data traffic to starve.

Zahedi and Pahlavan [16] also investigated the performance of voice communication in the presence of data traffic. Voice traffic was given a higher priority over the data traffic in order to maintain the delay requirement. To reduce the traffic load of the network, Improved Multi-Band Excitation (IMBE) codec was used which operates at 4.8 Kbps and consumes minimal bandwidth per call. The Frame Error Correction bits were removed to make voice packets shorter resulting in 88 bits of payload per packet while the payload size of data packets was 2576 bits. The overhead of employing TCP was identified as a key factor limiting the voice capacity. UDP and TCP were used for voice and data traffic, respectively. Each node was modeled as an $M/G/1$ queue in which the voice packets were prioritized over the data packets. An end-to-end delay of 100 and 50 ms was used as the threshold of acceptable voice quality. The packet loss was calculated from the packets that did not reach destination in 50 and 100 ms. But packet loss in the queue and MAC layer were not considered. Maximum packet loss of 0.1% was used as a limiting threshold in determining the voice capacity.

Malone *et al.* [102] investigated the trade off between delay and loss due to buffering in 802.11 WLANs. The investigation found that there is a sharp transition between a low delay-loss to a high delay-loss condition. The transition point is a function of the network traffic load and defines the voice capacity of the network. In particular, the impact of the buffer size on voice performance was investigated. The simulations were

performed using two way voice calls of 64 Kbps on-off stream with silence suppression. The on-off periods were assumed exponentially distributed with a mean of 1.5 second. Voice capacity was found to be severely affected by a small buffer, especially the performance of the AP degraded very quickly since it maintains a connection with each node resulting in a very high traffic arrival rate. Therefore, a small buffer of the AP results in severe voice quality degradation even with a small number of calls.

Dini *et al.* [103] investigated voice performance in the WLANs through experimentations in a testbed of 802.11g and identified Packet Loss Ratio (PLR) as a fundamental parameter that degrades the user perceived voice quality. Similar to [102], the access point queue was determined to be the most vulnerable point in terms of packet loss. Experiments using the G.729, G.711 and G.726 codecs dictated that PLR has to be kept below 10^{-3} (as mentioned in [14]) for acceptable voice quality. A PLR estimation technique involving parsing of the sequence numbers from the RTP headers of voice packets was also proposed in [103] to be used in the call admission control mechanisms.

2.5.2 Single Channel, Multihop Networks

Shin and Schulzrinne [104, 105] presented an analytical model and simulation based capacity estimation using 802.11b and 802.11e wireless testbeds. It was identified that although 802.11e can maintain QoS for VoIP traffic, it does not improve the voice capacity. The theoretical models in [104, 105] used oversimplified and non-realistic assumptions like ideal channel and no capture. Moreover, the impacts of multiple channels or interfaces were not considered. The delay in medium access was roughly estimated as $\frac{W}{2}$ idle slots which is true only when there are no collisions and the system remains in the first retry stage. Additionally, when a network approaches its capacity, the frequency of collisions increases resulting in a high backoff and medium access delay. The uplink delay was deemed minimal and ignored although the experimentations showed it to be about 30 ms. The impacts of the aggregation level, codec, queue and dejitter buffer were also not considered.

Armenia *et al.* [106] investigated voice performance using the Ad hoc On Demand Distance Vector (AODV) and Optimized Link State Routing (OLSR) routing protocols in a multihop 802.11b testbed. Throughput, end-to-end delay and delay jitter were used to measure the voice quality. Voice calls were generated using GnomeMeeting 1.0.2 software that uses the H.323 signaling protocol and a number of codecs including iLBC, GSM-06.10, MS-GSM, G.711-aLaw, G.711-μLaw and G.723.1. The G.711 codec (both aLaw and μLaw) was found to attain the highest throughput and lowest delay and delay jitter compared to the other codecs. This is due to the comparatively larger payload size of G.711 that minimizes the overhead of protocol headers.

2.5.3 Multi-channel, Multihop WLANs

Liu *et al.* [107] investigated the QoS of real time traffic in multi-channel, multi-radio, multi-hop networks. Performance of voice traffic in the presence of data traffic was experimented using a testbed of 32 nodes. Voice calls were generated using the GSM 06.10 codec and the application layer goodput and packet loss were estimated using Watchdog[18] client-server tools to measure the quality of service for voice and video traffic. Interference between different flows was found to be significant which led to unacceptable performance. Even the use of orthogonal channels for different NICs on the same node performed poorly. However, the use of non-overlapping channels increased the network capacity to some extent. It was concluded that coordination among nodes is of vital importance and to achieve an optimal network performance, all client nodes must use the channels collaboratively.

2.5.4 Multi-channel, Multihop Mesh Networks

Niculescu *et al.* [108] investigated voice capacity improvement techniques in multihop wireless networks using experiments in a 802.11b mesh testbed. The testbed consisted of 15 Stargate nodes and was deployed in the NEC[19] research lab, Princeton, NJ. The presence of TCP data traffic, packet loss in medium access due to operations in the unlicensed band and the high overhead of protocol headers were found to become severe in a multihop mesh scenario. When all the nodes were within each other's interference range in a chain topology, the UDP throughput was found to be decreasing with and be inversely proportional to the number of hops. G.729 codec was used with 10 and 20 ms of voice frames and the call quality was measured using the E-model. To improve the voice capacity, use of multiple interfaces, improved routing algorithm and packet aggregation were advised. When the interfaces were connected to separate channels, the interference due to transmissions in adjacent hops was minimized. Additionally, at the beginning of the call or for a change of route, a node sent a fake probe packet with voice characteristics, e.g., the frame size, UDP header, etc., through five different routes. A route that resulted in an acceptable voice quality was chosen. The protocol headers incurred a high overhead when the number of calls was increased. Packet aggregation and header compression were used conjunctively in [108] at each hop for the packet which shared the same next hop destination. A similar investigation was performed by Ganguly *et al.* [109] who also used the G.729 codec to make calls and the E-model to determine voice quality. In particular, packet loss and delay were investigated in a multihop mesh network using a testbed and NS2 simulations. Similar

[18] http://www.rhombustechnologies.com/main.asp?page=watchdog
[19] http://www.nec.com

to [108], the protocol header overhead and interference were identified as key factors that degrade voice quality and packet aggregation along with header compression were suggested in order to increase the voice capacity.

Falconio *et al.* [110] investigated the performance of VoIP services in a multi-interface multi-hop 802.11 mesh backhaul network using simulations in NS-2. The network interfaces operated in orthogonal channels so that a node could transmit and receive using different NICs simultaneously. Due to the overlapping of carrier frequency ranges, the contention became severe and the resulting high delay restrained the voice calls in the backhaul network to only a few hops of route length. The E-model was used to determine voice quality of the G.711, G.729 and GSM based VoIP calls. A chain topology was used to determine the maximum number of hops that a voice call can survive while maintaining voice quality. When the routers were equipped with a single interface each, the interference between adjacent nodes resulted in high delay and loss. It was suggested that the utilization of multiple interfaces tuned in different channels can reduce the interference in adjacent cells and improve the voice performance. Falconio *et al.* [110] considered uni-directional voice only which was extended by Falconio *et al.* [111] where two-way voice calls were used. But neither of [110, 111] considered real life factors like capture effect which can bring a positive effect to the network performance.

2.5.5 Voice Capacity using the PCF based Medium Access

While all of the above works considered the DCF based medium access only, a few researchers investigated the voice capacity with the PCF based MAC using both simulations and analyses.

Veeraraghavan *et al.* [20] investigated voice performance using the PCF based medium access mechanism. VoIP call capacity was estimated with and without silence suppression and a considerable capacity increase was achieved with the former approach. A larger inter-poll period was suggested to increase voice capacity. But a larger inter-poll period also increases the end-to-end delay and after a threshold is reached, this technique can not increase the voice capacity any more. The impacts of imperfect channels, packet loss and codec parameters were ignored and the delay and loss in the queue were also not considered.

Chen *et al.* [112] used a simulation approach using the OPNET simulator and investigated performance of VBR (Variable Bit Rate, with silence suppression) VoIP calls in a PCF based WLAN using the channel access delay as a threshold for voice quality. Silence suppression incurs inconvenience to the listener and most real VoIP softwares, including Skype, do not use it [113]. Therefore, in this dissertation use

Table 2.10: Comparison of Voice Capacity Analyses over IEEE 802.11 WLANs

MAC	Study	Network Type[a]	Codec	Voice Frame (ms)	Method	Quality Measure	Channel Type	Power Capture[b]	Queuing Effect	Unsaturated Traffic
DCF	Garg and Kappes [13]	SCSH	G.711, G.729, G.723.1	10∼20	Analysis	E-model	Ideal	Ignored	Ignored	Ignored
					Testbed	Commercial tool	Ideal	Ignored	Ignored	Ignored
	Hole and Tobagi [14]	SCSH	G.711, G.729	10∼50	Analysis	Delay limit	Ideal	Ignored	Ignored	Ignored
					Simulation	MOS	Imperfect	Ignored	Ignored	Ignored
	Medepalli *et al.* [99]	SCSH	G.711	10∼50	Analysis	None	Ideal	Ignored	Ignored	Ignored
	Anjum *et al.* [15]	SCSH	G.711	10	Testbed	Delay, Loss	Ideal	Ignored	Considered	Ignored
	Zahedi and Pahlavan [16]	SCSH	IMBE	18	Analysis	Delay	Ideal	Ignored	Considered	Considered
	Shin and Schulzrinne [104, 105]	SCSH	G.711	20	Analysis	None	Ideal	Ignored	Ignored	Ignored
		SCMH			Testbed	E-model	Ideal	Ignored	Ignored	Ignored
	Armenia *et al.* [106]	SCMH	G.711, iLBC, GSM-06.10, MS-GSM	30	Testbed	Delay, Jitter	Ideal	Ignored	Ignored	Ignored
	Liu *et al.* [107]	MCMH	GSM 06.10	20	Testbed	Packet loss	Ideal	Ignored	Ignored	Ignored
	Kawata *et al.* [17]	SCSH	G.711	20	Analysis	Delay	Ideal	Ignored	Ignored	Ignored
	Niculescu *et al.* [108]	MCMHM	G.729	10–20	Testbed	E-model	Ideal	Ignored	Ignored	Ignored
	Falconio *et al.* [110, 111]	MCMHM	G.711, G.729, GSM	10	Simulation	E-model	Imperfect	Ignored	Ignored	Ignored
PCF	Veeraraghavan *et al.* [20]	SCSH	Truespeech	30	Analysis	Delay	Ideal	—	Ignored	Ignored
	Chen *et al.* [112]	SCSH	G.711	10	Simulation	Delay	Ideal	—	Ignored	Ignored
	Kawata *et al.* [17]	SCSH	G.711	20	Analysis	Delay	Ideal	—	Ignored	Ignored
	Ma *et al.* [3]	SCSH	—	—	Analysis	Delay	Imperfect	—	Ignored	Ignored

[a]SCSH: Single Channel Single Hop, SCMH: Single channel Multihop, MCMH: Multi-channel Multihop, MCMHM: Multi-channel Multihop Mesh
[b]Applicable to DCF only

consider CBR (Constant Bit Rate) VoIP calls only. Using the G.711 codec in a WLAN with 11 Mbps data rate, the simulations in [112] revealed that PCF can support higher number of voice calls than DCF based medium access. But the impact of imperfect channel and delay and loss in the queue were ignored.

Although most VoIP applications use CBR traffic where a node is expected to have a packet to send after a certain interval, this is not the case for voice with silence suppression. With silence suppression, a node does not generate packets when the user is silent. IEEE 802.11 PCF uses a round robin polling schedule which leads to channel time wastage if a node does not have any packet to send. Kim and Suh [114] used an adaptive polling list to improve the voice capacity by minimizing such polling packet transmissions. The scheme in [114] considered only the uplink traffic and, assuming silence suppression to be used, modeled a one-way voice stream with a two-state Markov model as shown in Fig. 2.10. Here, the talkspurt and silence states were assumed to be exponential random variables with mean values of 1.0 and 1.35 sec, respectively. A statistical activity detection module was developed to determine the nodes in a talkspurt phase and was used to poll only those active nodes. The nodes that were expected to be in their silence phase were not polled. The performance benefit was gained from saving the time that would be wasted in potential silent nodes. However, the phase transitions of the Markov model can be expected to be correct only over a long run and can give wrong decisions occasionally. But such a situation would lead to packet loss which was not considered. Impacts of downlink traffic or imperfect channel were also ignored.

Kawata *et al.* [17] proposed a simplistic analytical model to estimate the voice capacity using the DCF and PCF based medium access. But unrealistic assumptions like ideal channel and no collisions make the models inapplicable to a real scenario. Two modifications of the standard PCF mechanism were also proposed to improve the voice capacity. Non-voice packets were forced to be transmitted in the contention period which gave voice packets a higher priority. Additionally, if a node transmitted a null frame (meaning it has no packet to send) for three consecutive times, the node was removed from the polling schedule. Conversely, the scheduling allowed a node to be included in the polling schedule when it transmitted a voice packet during the contention period. Whenever the queue of a node started to increase, i.e., if the queue had more than one packet, it transmitted those packets during the contention period. An overall 20% increase in the voice capacity was achieved with these techniques.

Quan and Hui [115] proposed an improved Deficit Round Robin (DRR) polling schedule to enhance voice performance. DRR is a weighted round robin scheduling algorithm applicable to variable packet sizes when the mean packet size is unknown. A maximum packet size threshold was determined in [115] and any packet that exceeded

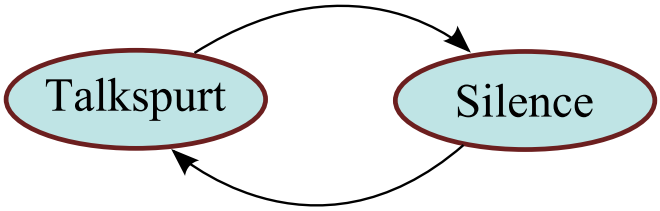


Figure 2.10: Two-state Markov model for one-way voice.

this threshold was held back for the next run of the scheduler. Although DRR works efficiently with variable length packets, it performs poorly under bursty traffic resulting in increased delay. To overcome this problem, two separate queues were used in [115] for VoIP packets and data packets, respectively, where the VoIP queue was allotted a larger quota for transmission. Any packet that exceeded the quota was delayed for transmission in a later run of the scheduler, thereby giving the voice packets a higher priority than the data packets. Most of the works that investigated VoIP performance with PCF based medium access concentrated on the polling schedule management.

Ma *et al.* [3] considered the effect of imperfect channel in a PCF based WLAN and investigated the voice capacity. However, the difference in channel time wastage due to channel errors in uplink and downlink packets was ignored which resulted in an incorrect estimate of the channel access delay. The delay in channel access was used to determine the upper limit of voice capacity which is an adequate measure, as discussed before, to ensure voice quality in the IEEE 802.11 WLANs. Instead of any standard codec, a CBR voice traffic generated at 64 Kbps was assumed and the effect of codec characteristics like aggregation level and packetization delay were ignored in [3]. Exactly one packet was assumed to be available at the beginning of each CFP which is unrealistic for a CBR with constant bit rate used with different CFP lengths. Moreover, the delay and loss in the queue incur severe quality degradation which were also ignored in [3].

The works reviewed in this section (and summarized in Table 2.10) show that the analytical approaches in determining the voice capacity ignored real world factors like imperfect channel and capture effect. Moreover, unsaturated traffic was also not considered although most networks operate under the unsaturated condition for a considerable amount of time. Most importantly, the analytical models did not use any standard voice quality measure and considered only delay or loss as a threshold for call capacity, although voice quality is impaired by the combined effect of both. We emphasize on voice quality in this thesis because if the voice quality is unsatisfactory, users will never use VoIP over WLANs.

2.5.6 Key Aspects to Consider in Estimating Call Capacity of WLANs

From the review of the existing works in the last section, we identify the following real network aspects that should be considered in estimating the call capacity of WLANs.

- ITU-T G.114 [97, 116] defined the upper limit of end-to-end delay as a mere guideline and it does not reflect the true voice quality impairment since voice quality is impaired by the combined effect of both delay and loss in the network. A standard voice quality estimation method, e.g., MOS, PESQ or E-model, should be used while estimating voice capacity of the WLANs.

- In the IEEE 802.11 WLANs, most of the delay and loss are associated with the medium access mechanism which, in turn, dictates the delay and loss in the queue. Therefore, more accurate models of the DCF and PCF based medium access mechanisms should be used in determining the delay and loss in medium access and queue. Since the high delay of DCF based MAC is associated with the backoff mechanism, the backoff model needs to be finely detailed and consider m-retry BEB since voice traffic performs better with a finite retry backoff algorithm.

- Considering the impact of real world factors, including imperfect channel and capture effect, is of utmost importance because it allows to determine a realistic estimate of voice capacity.

- Additionally, the delay and loss in the queue were ignored in most of the works which degrade the voice quality severely. Standard queuing analysis should be used in conjunction with a more detailed model of the MAC layer.

- Configurations of the voice codec and characteristic of the dejitter buffer influences voice capacity greatly and should be considered carefully.

2.6 Voice Performance Enhancement in WLANs

The low voice capacity of WLANs inspired a good number of works in the current literature which aimed at improving the standard protocols and devised mechanisms to support a higher number of calls. A number of authors investigated VoIP performance improvement by employing techniques associated with multi-rate adaptation, packet aggregation, dejitter buffer management, header compression, service differentiation with or without employing transmission opportunity of the IEEE 802.11e standard, medium access mechanism or call admission control. Some authors also focused on improving the wireless coverage of voice services and interference between adjacent

APs, handoff management or energy efficiency. We review the most cited works of different approaches in this section.

2.6.1 Multi-rate Link Adaptation

Whenever the link quality deteriorates, i.e., the wireless channel gets affected by noise and interference, and packet loss increases. Multi-rate adaptation is one way to reduce the packet loss by changing the data transmission rate as proposed by Simoens and Bartolome [117]. However, when the data rate is reduced, the available bandwidth might not still be sufficient to support voice quality. Specially, when the data rate is high due to low compression and aggregation, the rate adaptation can cause call jitter. Kawata and Yamada [118] presented an adaptive cross-layer mechanism where the encoding rate and packetization interval were adjusted depending on the time varying transmission rate of the physical layer. But communicating a modification of such codec parameters requires additional data transmission which incurs a delay overhead. The use of a media gateway (which can reside within the AP) was suggested to adjust the codec parameters by re-packetizing the voice packets and communicating with distant peers. In this mechanism also, some additional delay is incurred in the media gateway to re-packetize each and every packet. Moreover, voice privacy is broken as the AP is allowed to decrypt the packets.

2.6.2 Extended Coverage

The self configuring nature of DCF allows the client nodes to set up routes and forward traffic for one another even in the absence of an AP. This mechanism allows multihop communication and a single AP can serve Internet connection to nodes deployed or roaming about in a wide area. We address extending the coverage using multihop topology in Chapters 4 and 6. In this section, we review a few recent works on coverage for voice services.

Hsieh *et al.* [119] investigated spotty wireless coverage in the National Taiwan University main campus testbed. Multihop relay and dual mode handsets were utilized in [119] to serve isolated nodes in the "dead spots" where no wireless connection was available. The first approach used traffic relay by client nodes over only one hop to avoid the overhead of multihop routing protocols. The second method employed a GSM module of a dual mode mobile phone or PDA. A cross layer signal processing algorithm was used to decide on the choice of the network so that the performance of the voice service would not be hampered. Although the coverage was increased considerably, the delay jitter increased due to relaying and delay mismatch between GSM and WLAN calls.

Mondal *et al.* [38] investigated multihop voice communication using an enterprise testbed. The primary reason for voice quality degradation in such networks was identified to be significant packet loss. Transmitting each packet five times was suggested as a way to lessen the voice packet loss. VoIP codecs generate packets at a high rate and additional transmissions of each packet can cause higher collisions in each collision domain resulting in an even higher packet loss and increased delay. Therefore, this heavy replication was executed only between the end-points and nearby relays, and the duplications had been removed in [38] before the voice packets went into the public Internet. This mechanism reduced poor quality calls from 35% to 10% and increased good quality calls from 45% to 70%. Although the packet replication was reduced to avoid flooding the wired lines and the Internet, it still flooded the wireless network which is more prone to congestion.

2.6.3 Dejitter Buffer Management

Variation in the transmission delay and loss is identified as a key obstacle in ensuring good voice quality over the Internet. As mentioned before, a dejitter buffer can reduce packet loss at the cost of a higher packet delay and vice versa. Both the delay and loss incur impairment and this trade off between the delay and loss is of key importance in maintaining voice quality. A number of researchers investigated dejitter buffer performance with different techniques to keep both the end-to-end delay and loss within their acceptable limits. Ramjee *et al.* [120] used the mean and variance of packet delay in conjunction with a recursive linear filter to determine voice quality impairment. The end-to-end delay was predicted for each packet from the most recent observations. Four algorithms were proposed to predict the end-to-end delay which was then used to adjust the dejitter buffer so that the dejitter buffer loss could be minimized. Sreenan *et al.* [121] proposed a histogram based method with similar objectives. Here, a histogram of packet delay was generated after observing a predefined number of packets which was then used to approximate the average packet delay. The buffer length was dynamically set to the latest minimum packet delay in order to minimize packet loss due to delay jitter.

The exact voice quality impairment level can not be determined from only delay or loss, but the concerted impact incurred by both the delay and loss dictates the correct voice quality which is reflected in the E-model. Atzori and Lobina [122] used a simplified E-Model in designing a playout buffer. An adaptive dejitter buffer was used in [122] where the buffer length was dynamically adjusted to keep both the delay and loss at an acceptable level. To determine the appropriate trade off between the delay and loss, the simplified E-Model proposed by Cole and Rosenbluth [57] was

used. Packet delay, loss and inter-arrival time were recorded during a talk spurt. This information was then used to calculate the average delay, loss, packet inter-arrival time and standard deviation of inter-arrivals during a silent period. At the beginning of a new talk spurt, these values were used to determine the optimal length of the dejitter buffer that minimized the total voice quality impairment incurred due to the end-to-end delay and loss.

2.6.4 Packet Aggregation

The voice capacity of WLANs is largely throttled by the CSMA/CA mechanism which spends a considerable amount of time in the backoff state. As the medium access layer has to backoff and contend for channel access for each and every packet, the traffic load plays a crucial role in determining the network congestion and voice capacity. Traffic load is defined by the number of competitors and their packet arrival rates, and can be reduced in a number of ways. Firstly, the voice frames are usually aggregated in UDP packets by the packetizer. Secondly, these UDP packets can be aggregated statically or dynamically at any layer in the end-to-end path. As any kind of data aggregation incurs a delay overhead, special care is required to ensure that the voice quality is not degraded beyond an acceptable range. This can be easily achieved in the application layer, i.e., using voice frame aggregation. However, with packet aggregation schemes, the application layer voice quality has to be considered in the MAC layer aggregation, where the network congestion can be measured, and usually a cross layer approach is needed which can be either static or dynamic.

Packet aggregation can be applied on either inter-call or intra-call packets. Intra-call aggregation performs poorly since older packets are kept waiting to form a larger aggregated packet. In the intra-call aggregation method with a good number of voice streams, an adequate number of packets is more likely to be available in the queue to create the aggregated packet. But intra-call packet aggregation requires less computation at each hop since the total payload has one final destination and re-packetization is never required. In the inter-call aggregation scheme, however, it is likely that the voice frames in one aggregated packet need to go to different nodes and, in the worst case scenario, the packets have to be re-packetized at each hop aggregating only the packets for each of the next hop destinations. As a result, higher computational and buffering capabilities are required in the relaying nodes.

Wang *et al.* [123] proposed an inter-call packet aggregation method that utilized multicast transmissions to deliver aggregated packets to their destination nodes. The improvement was achieved by reducing the protocol overhead and channel time wastage during contention. But multicast frames are never acknowledged and retransmissions

are not performed. Therefore, successful transmissions can not be confirmed and each collision or channel error will incur a packet loss of not one but a number of voice packets. Given the high probability of collisions in 802.11b networks [124], this mechanism will suffer from high packet loss.

Yun *et al.* [124] proposed a dynamic, adaptive intra-call packet aggregation strategy to increase the voice capacity by decreasing the packet arrival rate according to the link congestion level. Additionally, bandwidth equalization between the uplink and downlink transmissions was proposed. Packet aggregation was done only when the network approached an overloaded condition. Whenever the MAC layer picked the Head of Queue (HoQ) packet, all the other packets in the queue which belonged to the same voice call were aggregated. Unlike intra-call aggregation by application layer codecs, this activity was performed at each relaying node. If some calls use a lower aggregation level, i.e., generate packets at a higher rate, they can enjoy a greater bandwidth as it is more likely that the queue will always have some packets to aggregate for these calls. However, the transmission time of an aggregated packet is much more than that of a regular packet and so the other calls will receive a lesser bandwidth leading to unfairness.

The AP often becomes the bottleneck since it has a greater packet arrival rate for symmetric traffic like VoIP. For an i number of two-way voice calls, the AP transmits $(i-1)$ times more packets compared to client nodes. To equalize the bandwidth utilization of the uplink and downlink traffic, [124] also proposed setting the contention window size of the AP to $\frac{1}{i-1}$-th of the standard collision window. But the non-proportional nature of backoff delay in the 802.11 standard was not considered.

Although the traffic load can be reduced by inter-call aggregation, it requires the relaying nodes to parse the packet contents and re-packetize them. Allowing any node other than the destination to look into a packet invades the privacy of the communication and can not be allowed. Moreover, it requires the relaying nodes to be more resourceful in terms of buffer size and computational power. Therefore, in this thesis we do not consider inter-call frame aggregation.

2.6.5 Header Compression

The protocol headers incur a high overhead for short voice packets. For example, with an aggregation level of two, each sender using the G.729 codec generates a voice packet with 20 bytes. However, to transmit it through an IP network, it has to carry additional headers of 74 bytes including the RTP, UDP, IP and MAC headers. If the latest IPv6 protocol is used instead of the traditional IPv4 protocol, then this overhead becomes 94 bytes in total. This large overhead incurred by the protocol headers instigated many

research projects to consider compression of such headers in order to save bandwidth which is very scarce in wireless networks. Fortunately, there is a good amount of redundancy in the header fields, specially among packets that belong to the same voice connection. Therefore, it is possible to send the common information only once at the beginning of a voice conversation and send updated information only when necessary. This approach shortens the header for most of the packets and saves a good amount of bandwidth. For instance, RFC 1144 by Jacobson [125] compresses the 40 bytes of the TCP and IP headers into 4 bytes before transmitting. The header compression module detects retransmission attempts and sends updates on any changed header fields whenever needed. RObust Header Compression (ROHC) by Bormann *et al.* [126] is a noteworthy standard for header compression that considers erroneous wireless channels and offers an efficient header compression technique for RTP-UDP-IP, UDP-IP and ESP-IP (Encapsulating Security Payload with IP) headers. However, the DCF mechanism wastes a considerable amount of time in the backoff state. Therefore, the bandwidth that is saved by utilizing header compression is not necessarily available in full for new voice connections [127]. Fortuna and Ricardo [127] investigated the performance of ROHC in 802.11 networks and suggested that a 23% improvement in the call capacity can be achieved from its use. Another header compression was proposed in RFC 2508 by Casner and Jacobson [128] which improved the ROHC technique [126] by compressing the 40 bytes header of IP, UDP and RTP into only 2 bytes of header provided that checksum is not used. Since VoIP is usually transmitted using a RTP-UDP-IP protocol stack, it can improve the call capacity considerably. But the checksum is very useful in determining the transmission errors which are expected in all real (imperfect) channels. However, [128] can still compress the same headers into a minimum of 4 bytes with the checksum enabled. If some redundant fields are changed, the method in [128] uses an explicit signaling message to update both parties. But the performance of the update procedure depends on the link round trip time and with lossy links, this technique does not perform well.

2.6.6 Packet Segmentation for Non-voice Packets

VoIP packets are smaller in size but are generated at a high rate. In traditional DCF, the medium access layer has to compete for each packet irrespective of their size or arrival rate. As a result, the delay overhead per VoIP packet is very high. In a network where non-realtime packets, e.g., Telnet or FTP, are larger, the voice packets attain a lower goodput than the non-realtime packets. One way to prioritize the voice packets is segmenting the non-realtime packets. Zhitao *et al.* [129] proposed an adaptive segmentation strategy for non-realtime packets in order to improve VoIP

performance. This technique reduces the residual transmission time of any background traffic on the voice traffic. Channel utilization was used to determine the maximum segment size. When channel utilization becomes very high, a congestion is in order and the segment size was suggested to be reduced quickly to avoid a bottleneck. Although this mechanism prioritizes voice packets, it increases the overall traffic arrival rate of the nodes which results in a higher packet loss.

2.6.7 Selective Dropping of Stale Packets

A packet that arrived late can not be played by the listener, therefore, a late packet is no more useful than a lost packet. If a packet suffers too many retransmissions but is not dropped by the medium access layer or the queue, there is a good chance that the packet is already too late to be played out but it will still be transmitted incurring a wastage of channel time during contention at each hop in the end-to-end path as it is relayed. Baldwin *et al.* [130] used an adaptive contention window control mechanism and a dead-line based queue management approach. Any packet in the queue that missed its deadline was dropped. This mechanism gave the other packets a greater chance to reach their destination in time. But the deadline can be decided only by the receiver and communicating the deadline incurs an additional overhead and requires changes in the standard packet header format.

2.6.8 Service Differentiation

The AP is usually the only node that is connected to the Internet and all the voice traffic must go through it. Although the AP transmits much more traffic than any other wireless station, it has to compete with the wireless nodes for access to the channel since the IEEE 802.11 standards do not differentiate between uplink and downlink data. Gao *et al.* [131] used service differentiation of the IEEE 802.11e standard to improve the VoIP capacity in WLANs and proposed a higher priority access category to be allocated to the AP and a priority lower access category to the mobile stations. A Markov model based analysis was used to determine the performance increase. An improvement of 20–30% in the voice capacity was found when using this method. Although imperfect channel and unsaturated traffic was considered, the positive impact of power capture was ignored and standard voice quality measures were also not used.

2.6.9 Transmission Opportunity

As mentioned in Section 2.3.2.3, the IEEE 802.11e HCF based medium access allows a node that has acquired access to the channel, either by DCF backoff or by receiving a polling packet from the AP, to transmit a number of packets sequentially given the

total transmission time remains within a predefined time limit called transmission opportunity (TX_OP). Dangerfield *et al.* [132] and Stoeckigt and Vu [133] utilized the transmission opportunity to enhance the voice performance which was verified through simulations. Additionally, a queuing model was developed in [133] to analytically evaluate the performance gain. As mentioned before, the queue at the AP has to buffer more packets compared to the client nodes and is more prone to becoming full. The impact of the buffer size at the AP is analyzed in [133] along with a guideline on the minimum buffer size required to support the maximum number of voice calls.

2.6.10 Energy Efficiency for Voice Traffic

A periodic wake-up and sleeping schedule is at the core of any standard energy saving mechanism for mobile nodes. Voice packets, being intolerant to end-to-end delay, pose a challenge in energy saving since the sleeping periods would increase the end-to-end delay of the voice traffic. To overcome these challenges, it is crucial to have the application and medium access layers collaborate in order to determine a sleep schedule with a minimum delay overhead.

Tsao and Huang [134] proposed a cross-layer energy saving mechanism exploiting the fact that, unlike the traffic types that are highly sensitive to packet loss (e.g., SSH, Telnet, etc.), voice streams can tolerate a small amount of packet loss. Depending on current packet loss rate and the target voice quality, this mechanism dynamically forces the packets to be transmitted in multicast mode which disables the medium access control layer acknowledgments. In this way, the time and energy required for transmitting acknowledgments were saved when voice quality was over the required level and some packet loss could be tolerated. However, this mechanism effectively disables the retry mechanism in the case of a transmission error. The retry mechanism was designed solely to reduce the packet loss which increases voice quality. The balance of voice quality and energy efficiency was dynamically decided in [134].

Zhu *et al.* [135] proposed a dynamic sleep strategy called the Collision Detective Dynamic Sleep Strategy or CDDSS to improve both the capacity and energy efficiency in 802.11 WLANs. In this strategy, the wireless nodes adapted sleep interval and packetization interval according to the average collision probability. Unlike a static sleep strategy, sleeping was used only when the additional delay due to sleeping would not degrade the voice quality. The method used the collision probability to detect the network congestion level and adopted a sleep interval so that the nodes could transmit in turns, thereby, reducing the collisions and giving an increased call capacity. But the capacity estimation ignored the impact of imperfect channel, capture effect and most importantly, standard voice quality impairments.

2.6.11 Adaptive Medium Access

Medium access control parameters can be exploited to avoid network congestion by adapting to changed traffic load. Moreover, voice traffic can be assigned a greater priority by modifying these parameters. Abu-Tair *et al.* [76] suggested modifying the backoff parameters to reduce the packet collision probability. In particular, the contention window size was adaptively chosen with respect to the network traffic load to reduce the number of collisions. In this mechanism, the number of collisions was recorded and depending on a threshold for the collision rate, the contention window size was increased either exponentially or quadratically. In the standard DCF, the contention window size is reset to its minimum value after a successful transmission. With a high traffic load in a dense network, this rapid decrease in the contention window size results in further collisions. Moreover, if the network size increases rapidly, the exponential increase in the BEB algorithm is deemed too slow. This is why, [76] used exponential and quadratic increase for light and high traffic load, respectively. The proposed mechanism reduced the packet loss which has an unfavorable impact on the voice quality.

A rapid increase of the contention window, as proposed in [76], can also result in a longer delay in the backoff state if the same number of collisions occurs. However, using a threshold to switch the contention window increasing function, it is possible to reduce the total number of collisions which will result in shorter delay. A reduction in access delay of 10% and a decrease in packet loss by 18% was reported in Abu-Tair *et al.* [76] using simulations in NS-2.

2.6.12 Interference of Adjacent Wireless LANs

In a mesh of wireless networks, the coverage of access points usually overlap to ensure complete coverage. But the overlapped areas can face interference from adjacent APs' transmissions. Chan and Liew [136] investigated the impact of interference in a cellular WLAN scenario where the APs are in the vicinity of each other. The access points were located in the hexagonal fashion used in GSM or CDMA based cellular mobile networks. To reduce inter cell interference, a conflict graph was used in call admission control. Additionally, a new access mechanism was proposed to isolate inter-cell interference by allocating the VoIP sessions of adjacent cells different time slots to access the channel. The IEEE 802.11 sleep mode was utilized to localize the contention for channel access to chosen time frames in each channel. The improvement was achieved through fewer collisions in each time frame at the cost of an increased delay in medium access. Capacity increase by 36.8∼52.1% and 29.15∼35.3% were found in the experiments with NS2 using the admission control and and the proposed time divisioned CSMA technique, respectively.

However, interference between co-located non-802.11 networks also hamper the voice performance which was investigated by Angrisani *et al.* [22] using a 802.11g testbed with Bluetooth and Additive White Gaussian Noise (AWGN). The tests were performed within a shielded semi-anechoic chamber compliant with the electromagnetic compatibility requirements for radiated emission tests. Cross-layer measurements at the network, physical and application layers were used to determine the impact of physical layer interference on the application layer voice quality. The experiments revealed that Bluetooth interference, having a low transmission power of 0∼20 dBm, causes only minimal and tolerable interference. AWGN, although causing abrupt delay jitter, also showed tolerable interference to voice quality. However, in an overlapping WLAN-WLAN scenario, the interference was found to be very high which resulted in increased packet loss and delay jitter, especially at high data rates.

2.6.13 Call Admission Control

Due to high voice packet generation rate, the voice networks are prone to getting saturated giving rise to an increased collision rate and resulting in high delay and loss. When a WLAN is already serving the maximum number of supported calls, adding another call can lead the whole network to saturation, resulting in voice quality degradation of every ongoing call. A Call Admission Control (CAC) system decides whether a new VoIP call can be admitted without breaching the QoS expectation of the ongoing calls. The CAC module generally resides in the access point and its sole purpose is to decide whether the current network resources, e.g., Internet bandwidth, channel data rate, SNIR, etc., can support another voice call while still meeting the QoS requirements of every active call. If the network resources are not sufficient, any new call requests will be denied by the CAC module.

In the current literature, the traffic load is identified as a key factor that dictates voice performance. Lowering the traffic load requires an increase in the inter packet generation time which, in turn, increases the end-to-end delay of voice packets. Therefore, the delay versus traffic load trade off is a crucial decision in the voice network designing. A call admission control strategy was proposed by Chen *et al.* [137] which adaptively controlled transmission interval in order to increase the call capacity. The CAC strategy (called ATICAC) adaptively chose a transmission interval according to the network status so that the probabilities of transmission and collision were lowered preventing the network from going into saturation. The analytical model did not consider imperfect channel condition or power capture, used the saturation traffic model by Bianchi [2] and tried to reduce the collision probability by 10%. It was assumed that a collision probability reduction by 0.1 was sufficient to keep the network

unsaturated. When a new connection request came, the ATICAC module checked the current collision probability to decide if the network would go into saturation and whether increasing the current transmission interval can protect it. If a new call would cause the network to become saturated, the call was rejected. A 50-100% increase in the number of calls was demonstrated through simulations performed in NS2.

Ortiz *et al.* [138] used effective bandwidth (the minimum bandwidth required by a flow) in the call admission control similar to the wired networks, e.g. ATM, and proposed a simplistic CAC model. A new voice connection was accepted as long as the required additional effective bandwidth was still available in the associated channel. The effective bandwidth is highly dependent on the traffic type and can be determined as explained in [139] and [140, Chapter 8]. Voice quality impairments and the effects of aggregation, queuing and medium access mechanism were ignored in [138].

A number of works suggested a user-centric approach in call admission control where the client nodes themselves decide individually whether to accept a new call. The admission decision can be taken considering channel utilization [141] or channel activity time [142]. Dini *et al.* [141] used the ratio of successful and failed transmission attempts to determine the effective traffic load. Each node estimated the ratio to determine whether a new call could be admitted by the AP without degrading the voice quality of any ongoing calls. The assumption in [141] that transmission failures are mostly due to collisions limits the application of this analysis to nearly ideal channels only. Channel activity time or the time between idle periods can lead to a wrong estimation of the traffic load since a good number of transmissions contain errors with an imperfect channel which increases channel activity estimation with much less traffic load. Therefore, the method in [142] is subject to error which was addressed in [141]. Dini *et al.* [141] observed the channel condition by checking the MAC CRC retry flag of every frame. Heterogeneous codecs and multi-rate PHY were ignored in [142] which were discussed in [141].

2.6.14 Handoff Management

As mobile nodes move around, a node may sometimes move outside the area covered by an AP and into the periphery of some other AP. In such a case, the node's routing information needs to be updated. Additionally for voice calls, a call admission decision may need to be made. This process is called the handoff where all routing and traffic control information and authority are handed over to a new AP. The handoff procedure requires some additional packet transmissions which incur delay and packet loss leading to call jitter and sometimes, call drop. A number of researchers investigated techniques to make the handoff procedure more seamless.

As a mobile node moves away from the AP the power of the downlink signal, i.e., signal strength, gets weakened. After a certain distance, the signal to noise and interference ratio becomes lower than a predefined threshold called the *call search threshold* that indicates that the AP is too far away and the mobile node needs to look for another AP with a better signal strength. However, unlike cellular networks such as GSM, there is no control channel in WLAN to inform a roaming node about new APs. Therefore, the handoff process usually involves a channel scan process followed by a re-authentication process. The channel scan is performed to discover potential APs in the new area while the re-authentication procedure updates the routing information and makes admission control decisions. The channel scan incurs most of the delay overhead in the handoff procedure. For example, the average time for re-authentication and scanning are 20 ms and 350–500 ms, respectively [143]. If an AP is not found in time and the signal power of the downlink frame from the old AP becomes lower than the *out-of-range threshold*, the communication with the old AP becomes interrupted and the ongoing call is dropped. This scenario represents a typical handoff failure.

Channel scan can be performed by passively listening to a channel for Beacon frames and actively sending a probe request frame. Huang *et al.* [144] proposed a channel scan scheduling mechanism with consideration for voice activity. A three state Markov model was used to model a voice conversation with silence suppression so that any silent periods could be foreseen. Whenever the node did not need to transmit or was not expecting a downlink packet from the AP, it searched in other channels for a possible AP by listening to Beacon frames. The signal to noise ratio of the Beacon frames indicates the goodness of signal strength of a potential new AP. If another AP was found with a lower SNIR, handoff was performed during a mutual silence period of the voice conversation so that voice quality was not hampered. However, although the scanning can be abandoned to transmit an early uplink frame, there is no way to notify when an early downlink frame is arriving. Failure to receive such downlink frames can lead to packet loss and it was suggested that to keep the packet loss below 1%, no more than two channels should be scanned in each mutual silence period.

The work in [144] was extended by Huang and Chang [143] to improve handoff management using a hybrid channel scan mechanism where prior knowledge about the neighboring APs, e.g., their channel number and Beacon arrival time, were assumed to be available. The mobile node utilized this information to schedule their scans in selected channels at particular time frames. The estimated SNIRs for each neighboring AP were estimated and an active scanning was performed by sending probe frames to the APs. Additionally, the channel switching time was estimated statistically. A better handoff performance in terms of a lower packet loss was achieved compared to [144]. Although the prior knowledge availability was justified by means of broadcast

packets by other roaming mobile nodes, the additional delay overhead to transmit such broadcast packets in a WLAN with frequently moving nodes was not discussed.

2.6.15 Voice Performance in Wireless Mesh Networks

Wireless Mesh Networks (WMN) are aimed at extending network coverage by placing static wireless mesh routers at regular distances so that a multihop communication path can be established to an AP even from a distant mobile node. The only difference between a multihop WLAN and a WMN is the utilization of mesh routers in place of regular client nodes. Mesh routers only implement physical, medium access and routing layers, and therefore, perform packet forwarding faster. Generally, the mesh routers are placed statically, have multiple Network Interfaces (NI) and enjoy a better power source. As a result, the coverage is ceaseless unlike that in a WLAN where the coverage can be interrupted as mobile nodes move away.

Since the medium and physical layer standards remain unchanged, most of the analyses of multihop WLANs are also applicable to WMNs. However, a number of works focused specifically on mesh networks to enhance their performance in terms of the voice capacity. A cross layer scheme to mitigate the PHY and MAC layer overhead by packet aggregation in mesh networks was introduced by Okech *et al.* [145]. The scheme used virtual queues for each outgoing link (of each NI) where the next hop distances were already known. To minimize the required number of channel access, an optimal packet size was determined dynamically by considering the signal to noise and interference ratio. However, as the voice frames wait in the queue to be part of an aggregated packet, the end-to-end delay increases which, in turn, increases the dejitter buffers loss and results in a higher delay impairment and lower voice capacity.

2.7 Research Challenges for VoIP over WLANs

VoIP offers a low cost voice communication mechanism that is widely accepted by both home and corporate users. But mobile phones are still popular due to the convenience of mobility. In order to deliver voice packets at mobile devices, we need to use a low cost wireless last mile career. IEEE 802.11 WLANs offer wireless coverage without incurring a high cost, and the ubiquitous availability of Wi-Fi radio interfaces in mobile devices makes it an ideal solution in providing the last mile coverage for voice calls. But maintaining the voice quality is challenging in such networks due to the absence of a QoS mechanism in the IEEE 802.11 standard. Therefore, WLANs should be carefully designed using a detailed analytical model that considers all the important real world factors. The existing works in the current literature (discussed

in Section 2.5) investigated different aspects of VoIP over Wi-Fi but ignored a number of key factors and used over simplified assumptions. As a result, the following research questions still remain open.

2.7.1 Analytical Model considering Voice Quality Impairments and DCF MAC performance

The existing voice capacity studies are primarily testbed and simulation based whose applications are limited compared to analytical models. The few analytical models used over simplified assumptions which make them unrealistic. For instance, assumptions of no collisions, no packet errors or fixed backoff delay were used which can not be attained in practical networks. Moreover, the analytical models did not use standard voice quality measurements in estimating the call capacity. The existing works used bandwidth consumption, channel time usage or end-to-delay to estimate the voice capacity which are inadequate in determining the exact quality impairment for a combination of delay, loss, delay jitter, etc. Thus, the voice quality remains uninsured and unpredictable in the designed networks. Most of the delay and loss in the WLANs occur in the MAC layer and are governed by the offered traffic load. The relationship between the delay and loss with the network size and traffic arrival rate is non-linear which has been identified by many works discussed in Section 2.4.2. Due to the oversimplified assumptions in determining the delay and loss, the existing voice capacity analyses fail to model the delay and loss associated with the DCF based medium access. A complete analytical call capacity model that considers the standard voice quality measures and the specific nature of performance of the DCF based medium access mechanism still remains an open question. To address this research challenge, an analytical call capacity model for VoIP over a DCF based WLAN can be designed. To ensure the voice quality, the model should use a standard quality estimation method such as the ITU-T E-model, and carefully determine voice quality requirements. Network aspects that incur impairments to voice quality should be carefully considered. In 802.11 WLANs, most of the delay and loss are incurred in the medium access layer, therefore, an accurate model of the DCF based medium access mechanism employing the finite retry BEB should be used in the call capacity estimation. Voice traffic characteristics in terms of codec specific features, the configurations of codec and dejitter buffer and their effect on voice quality should be incorporated in addition to network aspects so that the designed WLANs can maintain an acceptable level of voice quality. Such a model will be of great assistance to network designers and administrators

2.7.2 Realistic Network Factors including Imperfect Channel and Capture Effect

The delay and loss in WLANs are primarily associated with the medium access mechanism. Since voice traffic performs better with the finite retry BEB, compared to infinite retry BEB, a complete model of the DCF based medium access should be used in the call capacity models. Moreover, all real channels suffer from noise and interference which cause packet errors and degrade the WLAN performance. On the other hand, most Wi-Fi devices support capture effect which can reduce the packet loss due to collisions. But no existing study of the IEEE 802.11 DCF mechanism considers the effects of finite retry, imperfect channel, capture effect and unsaturated traffic concomitantly which is pointed out in Section 2.4.2. A voice capacity model can be useful only if all of these factors are considered carefully. Additionally, the model should be verified using simulations to validate all assumptions.

2.7.3 Delay and Loss in the Queue

It is identified from analyses and simulations in this research, that delay and loss in the queue are key factors that degrade voice quality. The choice of codec configuration and WLAN parameters define the network traffic load and dictate the performance of the MAC layer which, in turn, defines the service rate of the queue. Therefore, in addition to queue buffer length, configuration of the voice codec and DCF MAC parameters influence the delay and loss in the queue which contribute to the end-to-end delay, loss and voice quality impairments. But existing studies ignored the performance of the queue for voice traffic. A call capacity model needs to incorporate the impact of the queue performance considering voice traffic and WLAN performance in fine detail.

2.7.4 Multi-channel WLANs and Multi-interface Nodes

IEEE 802.11 supports the simultaneous use of multiple channels which can be best utilized using multiple interfaces. Future devices are likely to include more than one network interface in order to avoid network outages [25]. However, no existing analytical call capacity model considers the impact of multiple channels or network interfaces. The incorporation of these two features redefines the collision domain which ultimately governs the backoff delay and performance of the DCF based medium access. Especially, given that the IEEE 802.11 WLANs offer a low voice capacity, the incorporation of multiple channels is of great importance since it can increase the voice capacity considerably. However, apart from collision domain, such a scenario also requires the use of multi-server queuing system, which should also be considered in the capacity model.

2.7.5 Multihop Wireless Networks

The assumptions used in the existing works limit their application strictly to single hop networks only although multihop networks are commonplace. Multihop WLANs play a key role in offering large coverage using only a few access points. As intermediate nodes relay packets to and from the AP on behalf of the outer nodes, the effective traffic arrival rate increases for nodes that are nearer to the AP. As a result, the traffic arrival rate, and with it, the delay and loss distribution of collision domains in the same WLAN become non-uniform. This phenomenon also gives rise to a critical zone in the closest vicinity of the AP and strongly restrains the network capacity [94]. The formation of critical zones limits both WLAN performance and voice capacity. But the existing analytical models considered neither the changing traffic load nor the formation of critical zones.

2.7.6 Voice Capacity using the PCF based Medium Access

IEEE 802.11 PCF offers time synchronized medium access that can reduce delay and loss and increase the voice capacity. Therefore, analyzing voice performance using the PCF based medium access is of vital importance. However, the existing studies either considered ideal channel only or used simplistic methods to determine the super-frame length for imperfect channels, either of which led to erroneous performance measures (please see Section 5.1) and inadequate analysis of delay and loss. While estimating voice capacity of PCF based WLANs, an accurate model of the PCF mechanism needs to be used that also considers the effect of imperfect channels.

2.7.7 Time Synchronized Access in Multihop WLANs

While PCF offers low delay and loss by the use of time synchronized medium access, its use is restricted to single hop WLANs only. On the other hand, DCF offers wide coverage through multihop communication at the cost of high delay and loss. The best features of the two mechanisms could be employed together in a hybrid medium access mechanism that can offer low delay and loss in the multihop WLANs. But time synchronized medium access requires a point coordinator, or its equivalent, to control access to the channel. With the advances in mobile technologies, battery power, memory and other resources in mobile devices are no longer scarce or costly and the client nodes can be utilized in this regard. However, such a mechanism requires careful analyses to avoid channel hogging. Client nodes are expected to be mobile which also needs to be considered.

2.7.8 Other Challenges

As mentioned in Section 2.4.3, handoff management and power saving in mobile devices are crucial factors that can incur voice quality impairment. Usually some delay and loss are incurred during a handoff as additional packets need to be transmitted to update routing information and association with the old and new APs. Similarly, power saving by periodic sleep and wake-up schedule also incur some additional delay impairment to the voice quality. Some researchers also investigated the use of multi-rate MAC mechanism which can reduce the channel error rate by adapting the data rate according to channel noise and interference. However, these issues related to seamless handoff management, power saving or multi-rate MAC are outside the scope of this thesis.

2.8 Summary

In this chapter, we discussed the important role of VoIP in corporate and private communications and the need to provide voice services in mobile devices. We identified that the IEEE 802.11 standards offer a low-cost solution to provide the last mile coverage for voice calls. However, as mentioned in Section 2.4.3, due to the inherent limitations of WLANs, voice quality often degrades in such networks, especially under a high traffic load. Such networks need to be carefully designed considering voice quality requirements. Real network aspects like imperfect channel and capture effect, realistic traffic model and the inherent characteristics of the medium access mechanism should be considered. An analytical model that incorporates the above considerations will be of great assistance to network designers. In this regard, we reviewed the most notable works on both voice capacity and performance analyses of 802.11 WLANs. We identified their advantages and limitations, and summarized the open research questions in Section 2.7. In this respect, we target to design an analytical call capacity model in the next chapter with voice quality in the focus. The model considers the delay and loss in medium access and queue. The effects of imperfect channel and capture effect are considered and both saturated and unsaturated traffic conditions are incorporated. Most importantly, the combined effect of the end-to-end delay and loss on voice quality is determined using the ITU-T E-model so that the voice quality does not degrade in the 802.11 WLANs.

Chapter 3

Voice Capacity of Single Hop WLANs employing DCF

3.1 Introduction

IEEE 802.11 WLANs are prone to getting overloaded due to low bandwidth availability. In the absence of a QoS mechanism, the voice quality is severely adversely affected when the network gets overloaded, especially during busy hours. The voice quality jitter (or quality variation) in such a network causes users' inconvenience leading to customer dissatisfaction. Additionally, ongoing calls may be dropped (or disconnected), and new connections may be rejected if the network is not properly planned considering the expected call volume. Maintaining an acceptable level of voice quality is very important since customers will not use VoIP over WLANs if the voice quality is unacceptable and the customers are dissatisfied [12]. To maintain a good quality of voice calls, the WLANs have to be carefully designed considering the voice quality requirements. These networks have to be sufficiently provisioned considering the expected call volume so that voice quality is never hampered. In this chapter, we estimate the voice quality requirements in terms of quality impairment factors. We utilize this estimate in conjunction with a Markov model for DCF based communication to formulate a call capacity model. The model considers all the real world factors present in a real network so that it can be used in network designing to ensure the voice quality. We introduce the network model first and then determine the voice quality requirements in the following sections.

3.2 Network Description

In this chapter, we model the voice capacity of a single channel, single hop wireless network. Although a number of works in the current literature (reviewed in Section 2.5.1) offered analytical call capacity models, no existing work considered the voice quality impairments using an adequate method. Most works considered only a maximum limit of either end-to-end delay or loss; but as discussed in the previous chapter, it is the combined effect of both delay and loss that impairs the voice quality which none of the analytical capacity models in the literature has considered. Using only one of these limits results in an unrealistic over-estimation of the voice capacity. On the other hand, considering the both results in an under-estimation since the voice quality is unnecessarily over-restrained. When a standard voice quality estimation method, e.g., E-model, PESQ or MOS, is used to consider a given combination of delay and loss, an accurate estimation of the voice quality can be obtained which can then be used in a call capacity model.

In the IEEE 802.11 networks most of the delay and loss are incurred in the medium access and queue. But the over-simplified assumptions, e.g., no collisions or ideal error-free channel, used in the existing works [13, 14] render them inapplicable to real environments. Moreover, power capture and unsaturated traffic condition were ignored in all such works. The delay and loss in the queue were also not considered which are often the key determining factors of the voice quality.

To address the above mentioned issues, we consider the voice quality impairments using the ITU-T E-model. The justifications behind this choice of voice quality estimation method are explained in Section 2.1.5. We determine the network performance measures that are required to maintain an acceptable level of voice quality using a Markov model for the DCF mechanism. We consider the impacts of an imperfect channel, the capture effect and the delay and loss in the queue on the voice capacity. To model a real scenario closely, we also consider both unsaturated and saturated traffic conditions. We begin modeling the capacity of a single channel, single hop WLAN by first determining the voice quality requirements in the following section.

3.3 Voice Quality Requirement using ITU-T E-model

As discussed in Section 2.1.5, MOS, PESQ and E-model are the most widely used voice quality estimation methods. MOS requires human experimenters to listen to a voice sample that has passed through a network under investigation while PESQ requires a signal to be transmitted through the network. Both MOS and PESQ require the network to be used which is unavailable during the network design and planning

phase. On the other hand, the ITU-T E-model uses only network and device configurations in estimating the voice quality. Therefore, the E-model is the appropriate method to estimate voice quality during network planning. The E-model determines the voice quality impairments as a result of different network conditions. We use a top down approach that reverse calculates the voice quality impairments using network performance measures and determines the network configuration required to maintain an acceptable voice quality.

The E-model provides a prediction of the voice quality as perceived by a telephone user by considering a variety of conversational conditions in the end-to-end path of the connection. The E-model assumes that physical parameters can be mapped to a user's psychological satisfaction levels and represents the voice quality with a transmission rating factor R. The rating in R can be easily mapped to other quality measures, such as, the Mean Opinion Score (MOS), percentage Good or Better (%GoB) and percentage Poor or Worse (%PoW). Using a concept from the OPINE model which states that the independent psychological factors on the psychological scale are additive, the R score is represented as the weighted sum of different impairments as shown in (3.1).

$$R = R_0 - I_s - I_d - I_{e\_eff} + A. \tag{3.1}$$

The above components were introduced in Section 2.1.5.3 and are briefly described below (further details can be found in [18]).

3.3.1 Basic Signal to Noise Ratio R_0

The basic signal to noise ratio R_0 considers noise sources such as the circuit noise and room noise, and represents the received speech signal level relative to circuit and acoustic noise. The basic signal to noise ratio is defined as

$$R_0 = 15 - 1.5(SLR + N_0), \tag{3.2}$$

where SLR is the Send Loudness Rating, and N_0 is the power addition of all noise sources and is represented as

$$N_0 = 10 \log \left(10^{\frac{N_c}{10}} + 10^{\frac{N_{os}}{10}} + 10^{\frac{N_{or}}{10}} + 10^{\frac{N_{fo}}{10}} \right). \tag{3.3}$$

Here, N_c is the sum of all circuit noise powers and N_{fo} is the noise floor at the receiver side. The room noise at sender and receiver sides are denoted by N_{os} and N_{or}, respectively, which are represented as equivalent circuit noise. N_c, N_{os} and N_{or} are measured with reference to the 0 dBr point which, according to ITU-T G.223, is the mean absolute power level of speech plus the signaling current transmitted over a telephone line in one direction during busy hours. The value of this power level is -15

dBm0 and has a mean power of 31.6 microwatts. The noise floor at the receiver side is defined as

$$N_{fo} = N_{for} + RLR, \tag{3.4}$$

where RLR is the Receive Loudness Rating and N_{for} is the absolute noise floor at the receiver side which is usually set to -64 dBmp.

3.3.2 Simultaneous Impairment Factor I_s

The simultaneous impairment factor I_s in (3.1) combines all impairments, e.g., too loud speech levels and quantization noise, etc., occurring more or less simultaneously with the voice signal. The simultaneous impairment factor considers the effects of the Overall Loudness Rating (OLR), non-optimum sidetone and quantizing distortion, and is represented as

$$I_s = I_{olr} + I_{st} + I_q. \tag{3.5}$$

The impairment I_{olr} caused by a too low overall loudness rating is, in turn, a function of OLR, RLR and N_0, and is expressed as

$$\begin{aligned} I_{olr} &= 20\left(\left(1+\left(\tfrac{X_{olr}}{8}\right)^8\right)^{\frac{1}{8}} - \tfrac{X_{olr}}{8}\right), \text{ where} \\ X_{olr} &= OLR + 0.2\,(64 + N_0 - RLR)\,. \end{aligned} \tag{3.6}$$

On the other hand, I_{st} and I_q are functions of the SideTone Mask Rating $STMR$, Talker Echo Loudness Rating $TELR$ and the number of Quantization Distortion Units qdu. The quantization distortion is the inaccuracy introduced during the analog to digital conversion process. It is the rounding error between the analog input voltage to the ADC and the digitized output value. This is a non-linear and signal dependent parameter. Previously, qdu was the basis of an end-to-end transmission planning mechanism called the "14 qdu rule" that considered impairments due to digital signal processing. A mapping of qdu to the R score can be found in the G.113 standard [55]. However, qdu is inadequate to represent voice quality impairment and the use of the E-model is suggested by ITU-T instead.

3.3.3 Delay Impairment Factor I_d

The delay impairment factor I_d in (3.1) considers any delay and echo effects in the end-to-end path of the voice stream. The delay impairment factor is a function of the one-way delay that is calculated from the mouth (talker) to the ear (listener). The total one-way delay is the sum of the codec delay, look ahead delay, one-way network

delay, digital signal processing delay and receive jitter buffer playout delay. The delay impairment is formulated as

$$I_d = I_{dte} + I_{dle} + I_{dd}. \tag{3.7}$$

The delay impairment factor due to talker echo I_{dte} is a function of N_0, RLR, $TELR$ and the mean one-way delay T of the echo path, and is expressed as

$$I_{dte} = \left(\frac{R_{oe} - R_e}{2} + \sqrt{\frac{(R_{oe} - R_e)^2}{4} + 100} - 1 \right) \left(1 - e^{-T}\right), \tag{3.8}$$

where $R_{oe} = -15(N_0 - RLR)$ and R_e is defined as

$$R_e = 80 + 2.5(TELR - 40 \lg \frac{1 + \frac{T}{10}}{1 + \frac{T}{150}} + 6e^{-0.3T^2} - 14). \tag{3.9}$$

The impairment factor I_{dle} represents the impairments due to listener echo which is a function of R_0, the Weighted Echo Path Loss $WEPL$ and the round-trip delay in a 4-wire Loop t_a, and is expressed as

$$\begin{array}{l} I_{dle} = \frac{R_0 - R_{le}}{2} + \sqrt{\frac{(R_0 - R_{le})^2}{4} + 169}, \text{ where} \\ R_{le} = 10.5 \frac{WEPL + 7}{\sqrt[4]{t_a - 1}}. \end{array} \tag{3.10}$$

The component I_{dd} represents the impairment caused by an overly long absolute one way delay T_a which occurs even with perfect echo canceling, and is expressed as

$$I_{dd} = \begin{cases} 0 & \text{when } T_a < 100 \text{ ms}, \\ 25 \left(\left(1 + X^6\right)^{\frac{1}{6}} - 3 \left(1 + \left(\frac{X}{3}\right)^6\right)^{\frac{1}{6}} + 2 \right) & \text{when } T_a > 100 \text{ ms}, \end{cases} \tag{3.11}$$

where $X = \frac{\log\left(\frac{T_a}{100}\right)}{\log 2}$.

The following three delay parameters vary with the end-to-end transmission time—

- The absolute one-way-delay between the sender and receiver T_a.
- The mean one-way delay T between the receiver side and the point in a connection where a signal coupling occurs as a source of echo.
- The round-trip delay t_a in a 4-wire loop where the "double reflected" signal causes impairments due to listener echo.

Most of the above discussed parameters do not vary in a VoIP scenario and ITU-T strongly recommends the use of their default values in such a case. Representing the end-to-end delay with d_e we use $d_e = T = T_a = \frac{t_a}{2}$ (as suggested in [57, 146]).

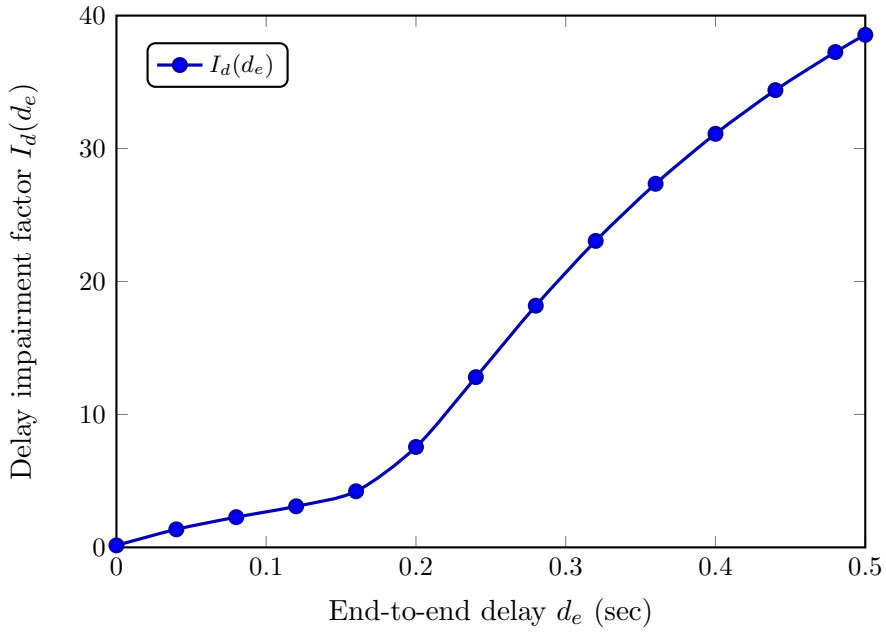


Figure 3.1: The relationship between the delay impairment factor $I_d(d_e)$ and end-to-end delay d_e.

Additionally, we use the recommended values for RLR, $TERV$ and $WEPL$ which reduce the delay impairment factor as a function of only the end-to-end delay d_e. From here onward, we use $I_d(d_e)$ to represent the delay impairment caused by an end-to-end delay of d_e. The relationship of the end-to-end delay and delay impairment is shown in Fig. 3.1. Inspired by the almost (piecewise) linear pattern of this function, Cole and Rosenbluth [57] approximated I_d using a piecewise linear function. The approximation in [57] is not accurate in all ranges of d_e and a piecewise quadratic curve was used to approximate I_d in [23]. However, the calculations required in network design and planning are carried out in an offline, non-realtime manner; and with today's high speed computational systems, the computational complexity is no longer an obstacle. Therefore, to achieve a greater accuracy, we use the original ITU-T equations to determine the delay impairment factor.

3.3.4 Effective Equipment Impairment Factor I_{e_eff}

The effective equipment impairment factor I_{e_eff} in (3.1) represents the psychological perception of the inconveniences introduced by low bit rate codecs and packet loss. The effect of a low bit rate codec is not related to most other parameters and depends on a subjective mean opinion score. However, voice quality impairment for a codec due to random packet loss is highly dependent on the network performance. In early versions

Table 3.1: Codec specific parameters for determining effective equipment impairment factor.

Codec	**Algorithm**	**Operating Rate (kbps)**	**Equipment Impairment at Zero Loss,** I_e	**Packet Loss Robustness,** B_{pl}
G.729	CS-ACELP	8	11	19
G.711	PCM	64	0	4.3
G.723.1	MP-MLQ	6.3	15	16.1
GSM-EFR	ACELP	12.2	5	10

of the E-model, the equipment impairment was presented as tabulated values. These values were approximated for easy calculation in [57] using a logarithmic function that was determined using curve fitting techniques. However, in the latest revision of the standard, I_{e_eff} is modeled as a function of the equipment impairment at zero loss I_e, the end-to-end packet loss e_e and packet loss randomness.

$$I_{e\_eff} = I_e + (95 - I_e)\frac{e_e}{\frac{e_e}{BurstR} + B_{pl}}. \tag{3.12}$$

Here, B_{pl} is the packet loss robustness factor. Both I_e and B_{pl} are codec specific parameters and their values for the most popular codecs are shown in Table 3.1 (data taken from [55, 56, 147] where similar values for most other codecs can be found).

The burst ratio $BurstR$ is defined as

$$BurstR = \frac{\text{Average length of bursts}}{\text{Average length of bursts with random loss}}. \tag{3.13}$$

As a result, $BurstR$ becomes 1 with random packet loss. The prevalence of bursty traffic is specific to the application type and network performance. VoIP applications generate regular size packets after regular intervals and therefore the packet arrival rate is even. Network congestion can lead to bursty traffic which can be detected only by network administrators. Without such information, we use random packet loss with $BurstR = 1$. Therefore, the effective equipment impairment factor becomes a function of only the end-to-end loss e_e which is denoted as $I_{e\_eff}(e_e)$ from here onward. The relationship between the effective equipment factor and end-to-end packet loss is illustrated in Fig. 3.2 for the G.729, G.711 and G.723.1 codecs.

3.3.5 Advantage Factor A

The advantage factor A in (3.1) incorporates the advantages of access and represents the user's expectation of voice quality based on how the call was initiated. For instance,

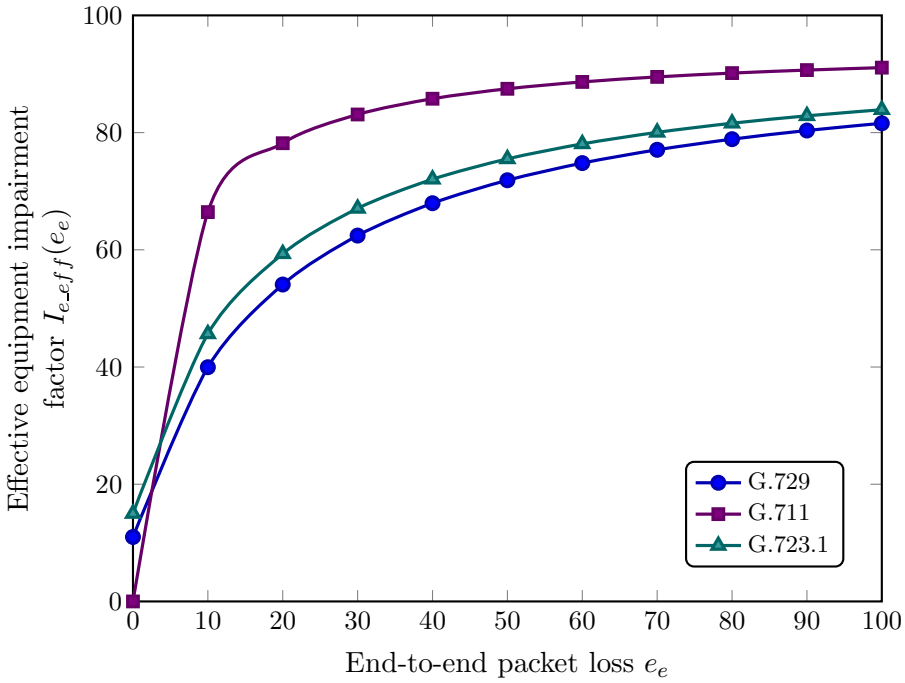


Figure 3.2: The relationship between the end-to-end packet loss and the effective equipment impairment factor.

Table 3.2: Prescribed ranges of the advantage factor.

Communication Scenario	**Maximum** A
Conventional (wirebound or wired network)	0
Mobility by cellular network within a building	5
Mobility in a geographical area or moving in a vehicle	10
Access to hard-to-reach locations, e.g., via multi-hop satellite connections	20

a little degradation in the voice quality is more likely to be accepted by a mobile phone user than a fixed line user. These advantages also include the advantage of mobility in areas that are hard to cover. For instance, a user in a remote area will be less dissatisfied with slightly degraded voice quality in his mobile phone than a user of a wired line. The prescribed values of the advantage factor from [55] are shown in Table 3.2. A value of 10 was suggested for cellular networks in [57] from a private communication between Cole and Rosenbluth [57] and Mark Perkins of AT&T. In a well planned WLAN, users enjoy a level of mobility that is comparable to cellular networks. Despite the slightly lower voice quality with VoIP (compared to wired PSTN lines), its low cost of service gives an additional benefit to the user. But to avoid any over-estimation of the call capacity, we use an advantage factor of 5 in this thesis.

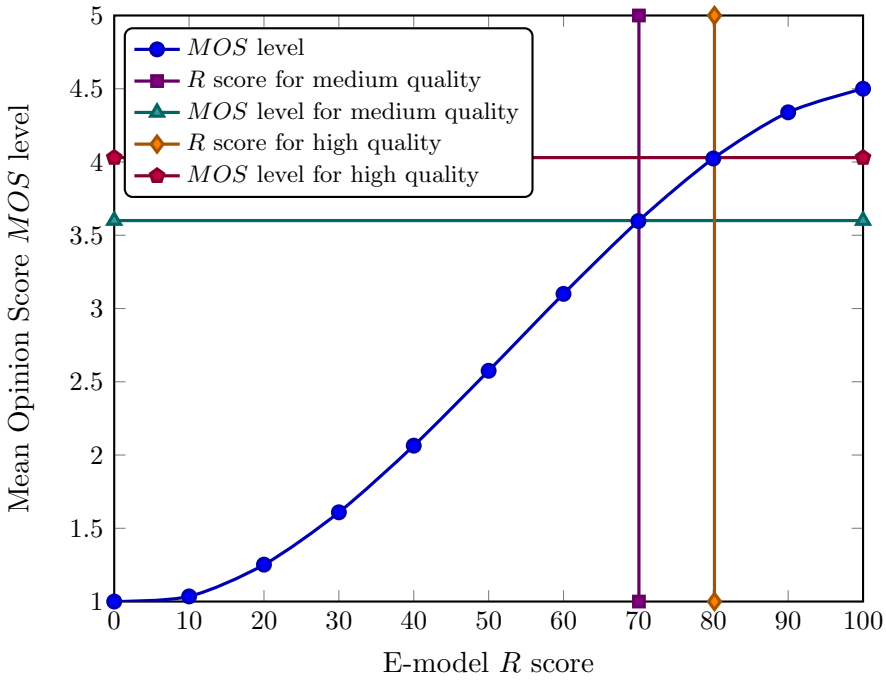


Figure 3.3: The R score versus MOS level relationship.

3.4 Impairment Budget

Since MOS is the most widely used voice quality estimation method, we map the voice quality requirement in MOS to equivalent E-model R score and then represent this requirement in terms of network performance measures. The MOS values for different voice quality levels are shown in Table 3.3. The minimum MOS levels for high and medium quality calls are 4.03 and 3.60 [18, 57], respectively. An R score using the E-model can be mapped to a corresponding MOS level using the following equation.

$$MOS = \begin{cases} 1 & \text{for } R < 0, \\ 1 + 0.0035 \times R + 7 \times 10^{-6} \times R(R-60)(100-R) & \text{for } 0 \leqslant R \leqslant 100, \\ 4.5 & \text{for } R > 100. \end{cases} \tag{3.14}$$

This relationship is depicted in Figure 3.3. Using the relationship in (3.14) we determine the minimum R score value for medium and high quality calls as shown below.

$$R \geqslant \begin{cases} 70.07 & \text{for medium quality,} \\ 80.16 & \text{for high quality.} \end{cases} \tag{3.15}$$

The minimum values of the MOS level and R score required for medium and high quality calls are also shown in Fig. 3.3.

Table 3.3: Voice quality ratings using the Mean Opinion Score.

Quality of Speech	MOS **Score**
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Table 3.4: Default values and ranges for E-model parameters.

Parameter	**Unit**	**Default**	**Range**
Send Loudness Rating, SLR	dB	+8	0∼+18
Receive Loudness Rating, RLR	dB	+2	-5∼+14
Sidetone Masking Rating, $STMR$	dB	15	10∼20
Listener Sidetone Rating, $LSTR$	dB	18	13∼23
Talker Echo Loudness Rating, $TELR$	–	65	5∼65
Sender's D-value of telephone, D_s	–	3	-3∼+3
Talker's D-value of telephone, D_s	–	3	-3∼+3
Weighted Echo Path Loss, $WEPL$	dB	110	5∼110
Sender's room noise, P_s	dB	35	35∼85
Receiver's room noise, P_r	d_B	35	35∼85

Although the determination of R score requires a number of parameters, ITU-T published additional network planning standards which provide default values for most of these parameters. These values are strongly recommended for use when the corresponding parameters are not under investigation. Some of these values are presented in Table 3.4. In our current work, we investigate the impacts of IEEE 802.11 MAC and PHY layer characteristics on the voice quality. We focus on the end-to-end path that carries the voice packets rather than the sender and receiver's environmental conditions. Therefore, we disregard variations of the loudness ratings, sidetone masking and room noise and use the ITU-T recommended values for these parameters. Using the ITU-T recommended values in equations (3.1)∼ (3.13) we can reduce the requirement of R in (3.15) into an impairment budget I_b of only the delay and loss impairments as shown below.

$$I_b = I_d(d_e) + I_{e\_eff}(e_e) \leqslant \begin{cases} 28.3553 & \text{for medium quality,} \\ 13.1952 & \text{for high quality.} \end{cases} \tag{3.16}$$

Here, I_b is the total variable impairment incurred during the transport of voice packets through the underlying network, and is the sum of the impairments due to end-to-end delay and loss. We explain the end-to-end delay and loss in the following sections.

3.5 End-to-end Delay

The end-to-end delay is measured from the mouth (talker) to the ear (listener) which comprises the following delay components.

Packetization Interval With aggregation level n_a, the packetizer puts an n_a number of voice frames in a single UDP packet. Since the first bit of a packet is withheld from transmission until the last bit is encoded and packetized, the packetizer introduces a packetization delay of $n_a d_f$ to the end-to-end path. Here, d_f is the length of a voice frame in seconds.

Look Ahead Delay Some coders use a look-ahead mechanism where the encoder reads (looks into) part of the following voice frame to better compress the current frame. The mechanism introduces an additional look-ahead delay d_l.

Queuing Delay A voice packet is wrapped with RTP, UDP and IP headers and put in the queue. The delay in the encoding process and adding the protocol headers, which sometimes require calculating checksums, can be expected to be minimal in today's high speed computing systems. For the same reason, the enqueuing and dequeuing delays can also be assumed to be minimal. But the time a packet waits in the queue before being fetched for transmission is considerably large. In fact, we identified from simulations that the queuing delay is one of the key factors in determining the voice quality. We denote the delay in the queue by d_q.

Channel Access Delay Whenever the MAC layer is free, it picks the Head of Queue (HoQ) packet and attempts to transmit it. Due to the collision avoidance nature of the DCF based medium access mechanism, the channel access delay, denoted by d_c, is also considerably high.

Internet Delay For long distance voice calls, the WLANs typically provide the last mile coverage while the Internet connects the end-point WLANs. The delay over the Internet is denoted with d_i.

Dejitter Buffer Delay When the receiving station receives a voice packet, it stores the packet in a dejitter buffer. The packets are picked from the dejitter buffer at regular intervals. The length of the buffer and the buffering mechanism dictates the delay d_j in the dejitter buffer. We use a static dejitter buffering mechanism as explained in [57]. The buffer can store $2b$ voice packets and incurs in a dejitter buffer delay of $d_j = bg$, where g is the mean packet inter-arrival time.

The above mentioned delay elements are additive and taking all the delay elements into account the end-to-end delay can be presented as

$$d_e = d_l + n_a d_f + d_i + d_q + d_c + d_j. \tag{3.17}$$

The delay in the Internet depends on the distance of the end-points of a connection and its route. Network designers can estimate its value from the expected call pattern. In the absence of such information, we use $d_i = 0$ as used in previous works [13, 14], however, our model accommodates such a delay when an estimated value is available.

3.6 End-to-end Loss

The end-to-end packet loss can occur in couple of ways which are discussed below.

Queuing Loss Every queue has a finite capacity to store packets. We denote the capacity of a queue by s_q which means that at any given time, the queue can hold a maximum of s_q packets. If the service rate of the queue is less than the arrival rate, the queue will become full. Any packet that arrives at an already-full queue will be dropped. These dropped packets give the queuing loss e_q which is calculated as the ratio of the number of packets dropped to the number of arrivals to the queue.

Channel Access Loss The DCF based medium access mechanism tries to minimize collisions using a backoff technique. Nevertheless, collisions do occur and a collided packet can not be recognized by the receiving station and is not acknowledged. Similarly, transmitted frames that are exposed to channel error can not be salvaged. But a collided packet is not necessarily lost (i.e., will not necessarily contribute to packet loss) since the sender station retransmits the packet unless the predefined retry limit is exceeded, i.e. a packet is dropped only when the number of retransmissions exceeds the retry limit. The packet loss in channel access e_c is calculated as the ratio of the number of packets dropped by the MAC layer as a result of the retry limit expiration to the number of packets fetched from the queue by the MAC layer.

Dejitter Buffer Loss Like the interface queue, the dejitter buffer also has a finite capacity. A packet that arrives to an already-full dejitter buffer will be dropped. Additionally, if a packet misses its deadline, i.e., if a packet reaches the dejitter buffer after its play out time has elapsed, the packet becomes useless and will be dropped anyway. A long dejitter buffer can minimize the packet loss but increases the dejitter buffer delay. Conversely, a short dejitter buffer decreases

the dejitter buffer delay but increases the packet loss. Dynamic dejitter buffering mechanisms adaptively choose the buffer length to balance the delay and loss. However, with the dejitter buffer mechanism specified in the last section, the dejitter buffer loss e_j can be presented as [57]

$$e_j \sim P\{l > bg\}, \tag{3.18}$$

where l is the expected delay jitter and $P\{X\}$ denotes the probability of an event X. Applying Chebyshev's inequality, it is possible to show that

$$d_j < \frac{v(l)}{(bg-g)^2}, \tag{3.19}$$

where $v(X)$ is the variance of a variable X. However, when b is sufficiently large (in the dejitter buffer of $2b$ packet capacity as explained in the last section) the dejitter buffer loss can be kept to an insignificant value.

The packet loss components are multiplicative and the end-to-end loss e_e can be represented as

$$e_e = e_q + (1 - e_q)\, e_c + (1 - e_q - (1 - e_q)\, e_c)\, e_j. \tag{3.20}$$

The performance of the WLAN greatly impacts the delay and loss in medium access. The delay in medium access is also the inter service delay of the queue. We first model the queue and then the MAC layer in the following sections but it should be noted that the delay and loss in both the medium access and queue are inter-related.

3.7 Modeling the Queue

Whenever the MAC layer becomes free, it fetches the HoQ packet from the queue and tries to transmit it. Therefore, the MAC layer acts as a server for the queue. The packet arrival rate λ and service rate $\mu = \frac{1}{d_c}$ play crucial roles in determining the performance of the queue. We model the queue as a $M/M/1/s_q$ queue [148] with Markovian arrival and departure where s_q denotes the queue size in number of packets. With a queue utilization ratio $\rho = \lambda d_c$, the probability P_i of exactly i packets being in the queue is given by

$$P_i = \rho^i P_0. \tag{3.21}$$

Here, P_0 is the probability that the queue is empty. P_0 can be determined from the following equality

$$\sum_{i=0}^{s_q} P_i = 1. \tag{3.22}$$

The empty queue probability is given by

$$P_0 = \begin{cases} \frac{1-\rho}{1-\rho^{s_q+1}} & \text{for } \rho \neq 1, \\ \frac{1}{s_q+1} & \text{otherwise.} \end{cases} \tag{3.23}$$

We define the probability p_q of a non empty queue as

$$p_q = P\{\text{Queue is not empty}\} = 1 - P_0. \tag{3.24}$$

3.7.1 Queuing Loss

All packets which arrive at an already-full queue are dropped and therefore, queuing loss e_q is given by

$$e_q = \begin{cases} \frac{\rho^{s_q}(1-\rho)}{1-\rho^{s_q+1}} & \text{for } \rho \neq 1, \\ \frac{1}{s_q+1} & \text{otherwise.} \end{cases} \tag{3.25}$$

3.7.2 Queuing Delay

Since packets are dropped at a rate of λe_q, the effective packet arrival rate is $\lambda(1-e_q)$. Applying Little's theorem we find $d_q\lambda(1-e_q) = \sum_{i=1}^{s_q}(i-1)P_i$ which, when simplified, gives the queuing delay as

$$d_q = \begin{cases} \frac{\rho\left(\rho^{s_q+1}s_q - \rho^{s_q}s_q - \rho^{s_q+1}+\rho\right)}{\lambda(\rho-1)(\rho^{s_q}-1)} & \text{for } \rho \neq 1, \\ \frac{s_q-1}{2\lambda} & \text{otherwise.} \end{cases} \tag{3.26}$$

3.8 Modeling Channel Access using Finite Retry BEB based DCF

As mentioned in Section 2.4, a number of works modeled the DCF based channel access mechanism. The Markov model based analyses show a reasonable accuracy with moderate mathematical complexity. However, no existing capacity model considered finite retry BEB, imperfect channel, power capture and unsaturated traffic conditions concomitantly. Therefore, here we develop a Markov model for the DCF mechanism employing m-retry BEB and the above mentioned factors to analyze the performance of a homogeneous WLAN. The steady state probabilities of the Markov model are used to determine the delay and loss in medium access which are then fed into the E-model to determine voice quality. The model considers the following real environment factors—

- Unsaturated and saturated traffic conditions

- Imperfect channel
- Power capture effect
- Finite retry BEB

3.8.1 The Markov Model

We denote the number of nodes (wireless stations) in the WLAN with n_s. The number of retry stages and the retry limit of the BEB algorithm are denoted with m' and m, respectively. The initial contention window size is denoted with W which is doubled after each transmission failure (by the BEB algorithm) for a maximum of m' times. A packet that has suffered transmission failures for $(m+1)$ times is dropped. The contention window size W_i at retry stage i is defined as

$$W_i = \begin{cases} 2^i W & \text{for } 0 \leqslant i \leqslant m', \\ 2^{m'} W & \text{for } i > m'. \end{cases} \tag{3.27}$$

The IEEE 802.11 standards specify the retry limit, CW_{min} and CW_{max}. Here, the latter two are the minimum and maximum contention window sizes, respectively, which define W and m' as

$$W = CW_{min} + 1, \tag{3.28}$$

$$m' = \log_2 \frac{CW_{max}+1}{CW_{min}+1}. \tag{3.29}$$

An ideal channel in which all transmissions are error free is only possible in theory. In every real channel, ambient noise subsists and impacts all ongoing transmissions. Even a transmission that occurred at one place acts as noise at another after high attenuation as it travels through free space, earth or walls and by being superimposed by other signals. When a noise signal intermingles with a transmission signal, bit errors occur. The bit error rate (BER) is a function of the channel noise, SNIR, coding scheme, data rate, device inaccuracies and the modulation scheme.

The bit error rate can be determined for a given signal to noise ratio, modulation scheme and particular model of network interface. An example relationship of the bit error rate with SNIR and different modulation schemes for an Intersil HFA3861B network card can be found in [149, Appendix A]. The packet error rate p_e is defined as the ratio of the number of packet transmissions that suffer from channel errors to the total number of packet transmissions. For a given network interface operating in a given channel using a particular modulation scheme of choice, bit error rate becomes a function of the physical layer transmission rate. The data bits are usually transmitted at a higher rate than the preamble and PLCP headers, and suffer from a different bit

error rate. If the bit error rates for the preamble (and PLCP) transmission rate and the data transmission rate are BER_p and BER_d, respectively, the packet error rate for a preamble and PLCP header size of $(pre + plcp)$ bits and a MAC layer payload size of *payload* bits can be shown to be

$$p_e = 1 - (1 - BER_p)^{pre+plcp} \, (1 - BER_d)^{payload}. \tag{3.30}$$

However, incorporating the SNIR and modulation scheme in a capacity model restricts the use of the model to one specific environment, a specific modulation mechanism and a vendor specific network card. Therefore, in this thesis we consider the impact of an imperfect channel as the packet error rate. Using the packet error rate to represent the impact of an imperfect channel enables the use of our model with any modulation scheme, network interface card or channel condition.

We denote the probability of an Event with $P\{\text{Event}\}$ and define the following necessary notations and their relations for the probability parameters used in this model. We denote the probability that a given node transmits in the channel at a given time slot by τ.

$$\tau = P\{\text{A random node transmits in a random slot}\}. \tag{3.31}$$

The channel will be busy if one or more nodes transmit in a chosen time slot. With τ being the transmission probability, $(1 - \tau)$ gives the probability that a node does not transmit. Therefore, the $(1 - \tau)^{n_s}$ is the probability that no node transmits, and the probability p_b of the channel being busy can be defined as

$$p_b = P\{\text{Channel is busy}\} = 1 - (1 - \tau)^{n_s}. \tag{3.32}$$

A collision occurs when two or more nodes transmit at the same time. When the channel is already busy, a collision will occur if one or more nodes transmit with probability $1 - (1 - \tau)^{n_s - 1}$. Therefore, the collision probability p_c can be defined as

$$p_c = P\{\text{Collision} \mid \text{Transmission}\} = \frac{1 - (1 - \tau)^{n_s - 1}}{p_b}. \tag{3.33}$$

When a collision occurs, all the colliding packets are not necessarily corrupted. A packet that is received with the maximum power can still be salvaged because the receiver uses the power ratio of the data signal to noise and interference signals to determine the data signal contents. If the ratio of the power of the packet received with the maximum power to the sum of power of all other received packets is greater than a predefined capture threshold z, then the packet with the maximum power can be received (or captured). The value of z depends on the vendor specific implementation of the network interface card. We denote the probability of a capture by p_p. The

capture probability is a function of the capture threshold, correlation receiver's gain and power of interfering packets. The probability of observing exactly i interfering packets in a given time slot is denoted by π_i, and is defined as

$$\pi_i = \binom{n_s}{i+1} \tau^{i+1} (1-\tau)^{n-i-1}. \tag{3.34}$$

When i packets interfere with an ongoing transmission in a Rayleigh fading channel, the probability β_i that the ratio of the power of the packet with the highest power to the sum of power of all interfering packets exceeds the capture threshold can be defined as [19, 95, 96]

$$\beta_i = \frac{1}{(1+zG)^i}, \tag{3.35}$$

where G is the processing gain of the correlation receiver. If the spreading factor of the sender's code spreader is denoted with s (e.g., $s = 11$ for an 11 chip Barker sequence), the processing gain is given by

$$G = \frac{2}{3s}. \tag{3.36}$$

Considering that any number of transmissions may collide, the capture probability is given by

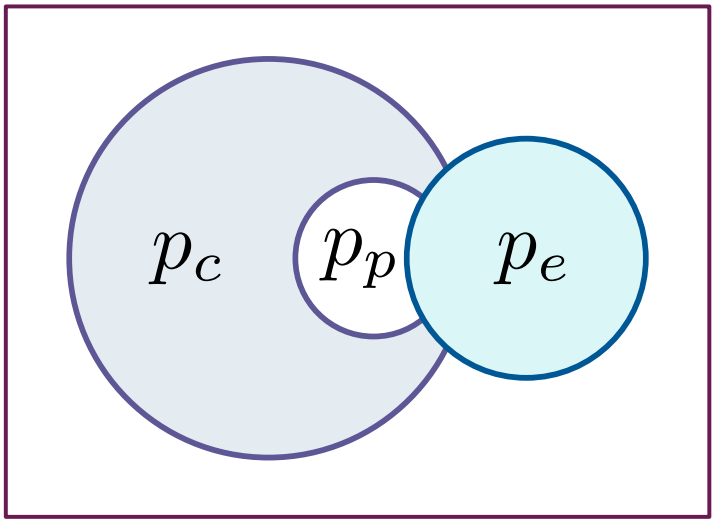
$$p_p = \sum_{i=1}^{n_s-1} \pi_i \beta_i = \sum_{i=1}^{n_s-1} \binom{n_s}{i+1} \tau^{i+1} (1-\tau)^{n-i-1} (1+zG)^{-i}. \tag{3.37}$$

A transmission may contain channel errors or suffer from a collision without a capture or both. In such cases, the receiver does not transmit the ACK. The sender waits for the ACK frame for an ACK_Timeout period and continues with the next transmission. p_f is the probability that a transmission failed, i.e., an ACK was not received and the packet requires at least another retransmission. Similarly, p_s is the probability that a transmitted frame is received correctly, i.e., it is acknowledged and $p_s = 1 - p_f$. Using the above explanation of the success and failure of a transmission in relation to p_c, p_p and p_e, we can define p_f as

$$p_f = (p_c - p_p) + p_e p_p + (p_e - p_e p_c). \tag{3.38}$$

This relationship can be easily understood and presented using the Venn diagram in Fig. 3.4 where the shaded and white regions refer to the probabilities of failure p_f and success p_s, respectively.

Similar to [2, 19, 75, 83, 84, 90], we assume that the probability of transmission in a random slot and the probability of a collision are independent of the retry history. This condition implies that τ and p_c are independent of the current retry stage.



Shaded regions correspond to transmission failure probability p_f and the rest corresponds to transmission success probability p_s.

Figure 3.4: Venn diagram for success and failure of a transmission.

The Markov model we used to model the finite retry BEB based DCF mechanism is shown in Fig. 3.5. The states (elliptical shapes) in a row are in the same retry stage while retry stages increase from top to bottom. The columns denote different counter values increasing from left to right. The notation (i,k) in the figure denotes the i-th retry stage with a counter value of k. When a transmission from a state $(i, 0)$ fails, the retry stage increments to $(i+1)$ and the system shifts downward by one row. The new counter value is set to an integer k which is chosen randomly (uniform) from the set $[0, W_{i+1} - 1]$. Thus, the new state becomes $(i + 1, k)$. The state E at the top left corner of the figure represents the state when the queue is empty. This state is the key to modeling unsaturated traffic. The arrows represent transitions from one state to other possible states and each arrow is labeled with the probability of the associated transition.

Although our model resembles those presented in [90] and [19], it differs in a number of ways. [19] modeled the ∞-retry BEB where a packet is retransmitted until success. Therefore, the outgoing arcs from state (m',0) with probability p_f are recursive to the m'-th retry stage and retry stages $m' + 1, m' + 2, ..., m$ are not present. Our Markov model is designed for the m-retry BEB and, therefore, the system is reset from the m-th retry stage irrespective of any success in transmission. Therefore, the outgoing arcs from state $(m, 0)$ are neither recursive, nor depend on the success or failure of the last transmission. On the other hand, [90] modeled the saturated traffic condition only where every node has some packets to transmit at any given time. Therefore,

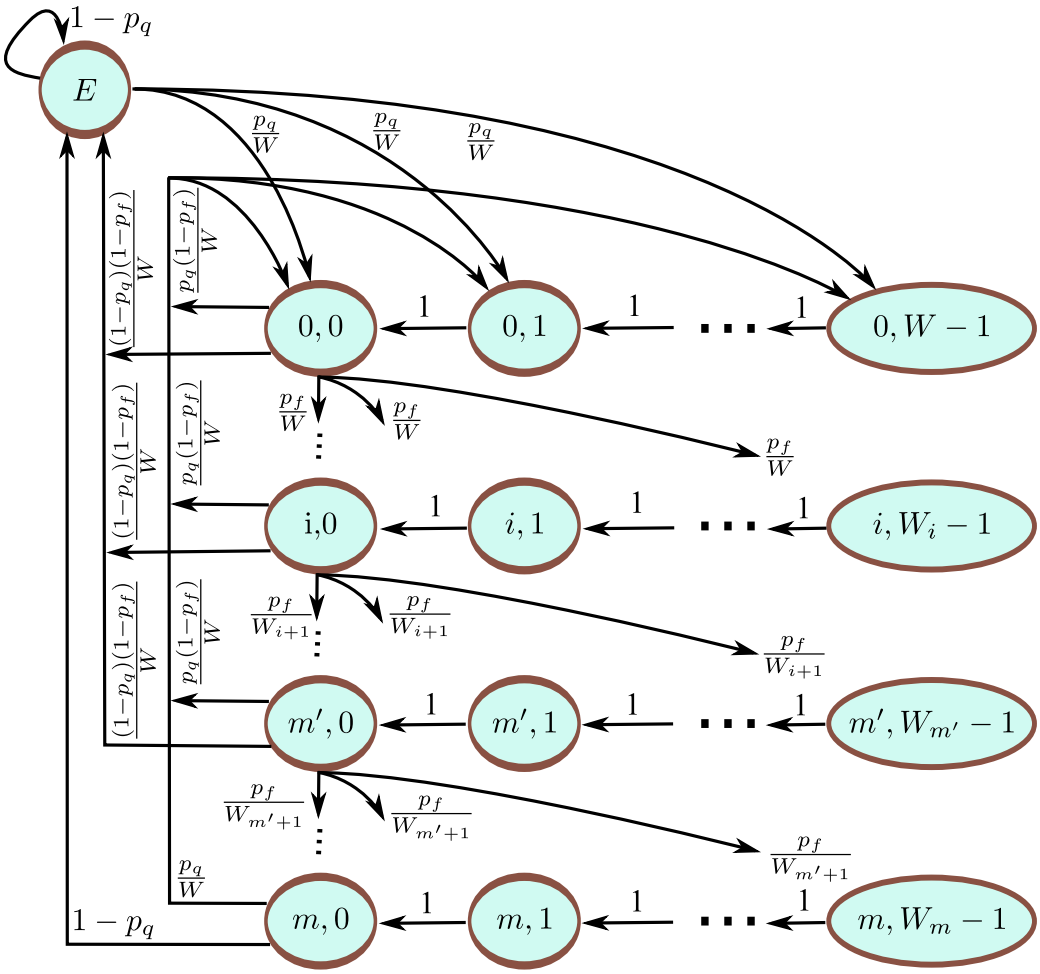


Figure 3.5: Markov model for finite retry BEB based DCF.

the empty queue state E is not present in [90] and whenever the retry stage is reset, the Markov model always goes to the 0-th retry stage. The system is reset when a transmission is successful or the retry limit exceeds. In our model, the system goes to the state E if the queue is found empty during a reset. Otherwise, the retry stage is set to 0 and the counter value is chosen from $k \in [0, W-1]$. Although [84] investigated unsaturated traffic load for m-retry BEB, the effects of imperfect channel and power capture were ignored.

3.8.2 One Step Transition Probabilities

The states in our Markov model are bi-dimensional processes $\{s(t), b(t)\}$ where $s(t)$ is the retry stage and $b(t)$ is the counter value at time t. The one step transition probability from state (i,k) to (i',k') is denoted as $P\{i',k'|i,k\}$. In the following, we define the Markov model transition probabilities (Fig. 3.5) depending on the action of

the counter and retry stage as follows.

1. The counter is decremented when the channel is idle.

$$P\{i,k|i,k+1\} = 1 \text{ for } 0 \leqslant i \leqslant m \text{ and } 1 \leqslant k \leqslant W_i - 2. \tag{3.39}$$

2. The retry stage is incremented after a transmission fails.

$$P\{i,k|i-1,0\} = \frac{p_f}{W_i} \text{ for } 1 \leqslant i \leqslant m \text{ and } 0 \leqslant k \leqslant W_i - 1. \tag{3.40}$$

3. The retry stage is reset when a transmission is successful from state $i < m$ and the queue is non-empty.

$$P\{0,k|i,0\} = \frac{p_q(1-p_f)}{W} \text{ for } 0 \leqslant i \leqslant m-1 \text{ and } 0 \leqslant k \leqslant W - 1. \tag{3.41}$$

4. The retry stage is reset when the retry limit is exceeded and the queue is found non-empty.

$$P\{0,k|m,0\} = \frac{p_q}{W} \text{ for } 0 \leqslant k \leqslant W - 1. \tag{3.42}$$

5. The system reaches state E after a transmission is successful and the queue is found empty.

$$P\{E|i,0\} = (1-p_q)(1-p_f) \text{ for } 0 \leqslant i \leqslant m-1. \tag{3.43}$$

6. The system reaches state E after the retry limit is exceeded and the queue is found empty.

$$P\{E|m,0\} = (1-p_q) \text{ for } 0 \leqslant i \leqslant m-1. \tag{3.44}$$

7. The system reaches the initial retry stage with the arrival of a packet to an empty queue.

$$P\{0,k|E\} = \frac{p_q}{W} \text{ for } 0 \leqslant k \leqslant W - 1. \tag{3.45}$$

3.8.3 Stationary Probabilities

We denote the stationary probability distribution of any state (i,k) by $b_{i,k}$ which can be defined as

$$b_{i,k} = \lim_{t\to\infty} P\{s(t) = i, b(t) = k\} \text{ for } i \in [0,m] \text{ and } k \in [0, 2^i W - 1]. \tag{3.46}$$

Let the steady state probability of the system being in state E be denoted by b_E. As long as the retry limit is not exceeded, the retry stage increments after each failed transmission and

$$b_{i,0} = p_f\, b_{i-1,0} = p_f^i\, b_{0,0} \text{ for } 1 \leqslant i \leqslant m. \tag{3.47}$$

Using the similarity of the state changing probabilities across all retry stages, we can show that

$$b_{i,k} = \frac{W_i - k}{W_i} \begin{cases} p_q(1-p_f) \sum_{j=0}^{m-1} b_{j,0} + p_q b_{m,0} & \text{for } i = 0, \\ p_f \, b_{i-1,0} & \text{for } 1 \leqslant i \leqslant m. \end{cases} \tag{3.48}$$

The steady state probability of entering and leaving any state is equal. Imposing this condition on state E, we obtain

$$b_E = \frac{1}{p_q} \left\{ (1-p_q)(1-p_f) \sum_{i=0}^{m-1} b_{i,0} + (1-p_q) b_{m,0} \right\}. \tag{3.49}$$

Imposing the same condition on state $\{0,0\}$ gives

$$p_f b_{0,0} = (1-p_f) \sum_{i=0}^{m} b_{i,0} + p_f \, b_{m,0}. \tag{3.50}$$

Summation of steady state probabilities of all states (states $\{0,0\}, \{1,0\}, \ldots, \{m-1,0\}$) where a transmission failure increments the retry stage gives

$$\sum_{i=0}^{m-1} b_{i,0} = b_{0,0} \sum_{i=0}^{m-1} p_f^i = \frac{1-p_f^m}{1-p_f} b_{0,0}. \tag{3.51}$$

Using (3.51) and (3.48) in (3.49), we derive

$$b_E = \frac{1-p_q}{p_q} b_{0,0}. \tag{3.52}$$

The steady state probabilities of all states should sum to 1, i.e.,

$$\sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} + b_E = 1, \tag{3.53}$$

which defines $b_{0,0}$ as shown in (3.54).

$$b_{0,0} = \frac{2p_q(1-p_f)(1-2p_f)}{\left(\begin{array}{r} p_q W (1-p_f)\left(1-(2p_f)^{m'+1}\right) + p_q(1-2p_f)(1-p_f^{m'+1}) \\ +p_q p_f^{m'+1}(1-2p_f)(2^{m'}W+1)(1-p_f^{m-m'}) \\ +2(1-p_f)(1-2p_f)(1-p_q)(1-p_f^m) \\ +2p_f^m(1-p_f)(1-2p_f)(1-p_q) \end{array} \right)}. \tag{3.54}$$

A packet is transmitted only from the states $b_{i,0}$ where $i \in [0,m]$. Since τ denotes the probability of a transmission by a random node at a random time slot, the transmission probability is the sum of the steady state probabilities of such states.

$$\tau = \sum_{i=0}^{m} b_{i,0} = b_{0,0} \sum_{i=0}^{m} p_f^i = \frac{1-p_f^{m+1}}{1-p_f} b_{0,0}. \tag{3.55}$$

3.8.4 Expected Slot Length

Let t_σ, t_{sifs} and t_{difs} be the length of an idle slot, the SIFS and DIFS periods, respectively, and let t_δ be the propagation delay. Similarly, t_d, t_a, t_r and t_c are the transmission times of data, ACK, RTS and CTS frames, respectively.

A successful transmission is confirmed by the reception of an ACK packet. We define t_s as the time a sender has to wait for before it can be sure of the success of its last transmission and remove the HoQ packet from the queue. In the basic handshake method, this delay includes the initial DIFS, the time needed to transmit the data packets, the SIFS which separates these two transmissions and the propagation delay associated with the transmitted frames to reach their destinations. In the RTS/CTS based handshake, this delay additionally includes the time to transmit the RTS and CTS frames, the propagation delays and the two SIFS waiting periods that separate these additional transmissions. Taking all the above delay elements into account, t_s can be defined as

$$t_s = \begin{cases} t_{difs} + t_d + t_\delta + t_{sifs} + t_a + t_\delta & \text{for Basic DCF,} \\ t_{difs} + t_r + t_\delta + t_{sifs} + t_c + t_\delta & \\ \quad + t_{sifs} + t_d + t_\delta + t_{sifs} + t_a + t_\delta & \text{for RTS/CTS DCF.} \end{cases} \tag{3.56}$$

When a transmission fails, the sender can not detect either collision or reception with errors. A collision can occur on the data packet in the basic or the RTS packet in the RTS/CTS based handshake, respectively. In either case, the receiver does not transmit the ACK (basic) or CTS (RTS/CTS) frame. The sender waits for a timeout duration t_f before assuming the last transmission to be lost. No incoming acknowledgment (basic) or clear-to-send packet (RTS/CTS) at their expected time of arrival confirms that the last transmission failed. Considering the channel time used in the associated handshake mechanisms, the time t_f to detect the failure of the last transmission can be defined as

$$t_f = \begin{cases} t_{difs} + t_d + t_\delta + t_{sifs} + t_a + t_\delta & \text{for Basic DCF,} \\ t_{difs} + t_r + t_\delta + t_{sifs} + t_c + t_\delta & \text{for RTS/CTS DCF.} \end{cases} \tag{3.57}$$

The advantage of the RTS/CTS mechanism can be easily understood from the above equation. In both types of handshakes, the first packet in the sequence can collide with other transmissions. The transmissions that follow in the handshake usually do not suffer from collisions which is explained in the following. In the basic handshake mechanism, the data packet is followed by an ACK packet after an interval of a SIFS. Since every other node waits for the channel to be transmission-free for a DIFS, which is longer than a SIFS, the ACK transmission does not face a collision

(except due to hidden node and exposed node problems as addressed in the RTS/CTS mechanism). In a RTS/CTS based handshake, both the RTS and CTS frames carry the total duration of the handshake which is used by other nodes to set their NAV. As a result, the other nodes use virtual carrier sensing and refrain from accessing the channel during the handshake.

As mentioned above, the data frame in the basic handshake or RTS frame in the RTS/CTS handshake, respectively, can suffer from a collision. But the RTS frame being smaller than a typical data frame, makes t_f shorter in the RTS/CTS compared to the basic handshake. The performance gain is achieved from quickly detecting a failed transmission and resuming the following transmissions of other packets. However, the additional transmissions of the RTS and CTS frames do incur some delay overhead and the RTS/CTS based handshake only benefits in transmissions of long packets. VoIP packets are usually very small and carry only about 20∼160 bytes of payload and TCP/IP protocol headers. Therefore, the basic handshake is usually used. However, since some existing works in the literature used the RTS/CTS based handshake, we consider the both in this study.

DCF is a slotted time protocol where time is not computed in a continuous manner. We use an approach similar to [2, 19, 87, 92] where a slot can have a variable length. The delay before the counter is decremented and the time spent in detecting a transmission success or failure are all considered as slots. Using p_f and τ as defined in (3.38) and (3.55), respectively, we calculate the expected slot length t_{slot} as

$$t_{slot} = (1 - p_b)\, t_\sigma + p_b\, p_f\, t_f + p_b\, p_s\, t_s. \tag{3.58}$$

3.8.5 Medium Access Performance Measures

We derive the standard WLAN performance measures in this section which will be used in the VoIP capacity estimation model. Throughput is a standard measure of network performance but the delay and loss in channel access are more useful measures of voice performance. We discuss the MAC layer throughput and delay and loss in channel access using our proposed Markov model in the following section.

3.8.5.1 Throughput

The throughput ψ is the number of payload bits that the MAC layer can transmit per second. This is the same as the ratio of the number of bits transmitted in a slot time to the slot length which is given by

$$\psi = \frac{p_b\, p_s\, 8(l_d + l_r + l_u + l_i)}{t_{slot}}, \tag{3.59}$$

where l_d, l_r, l_u and l_i are the length of application data and the header sizes of the RTP, UDP and IP layers (all in bytes).

3.8.5.2 Medium Access Loss

A packet is dropped when it suffers from more than an m number of failed transmissions. Therefore, the packet loss e_c in medium access is the probability that a packet suffers from $(m+1)$ failed transmissions and is given by

$$e_c = p_f^{m+1}. \tag{3.60}$$

3.8.5.3 Channel Access Delay

The channel access delay d_c can be measured as the length of the period starting when a packet becomes the head of the queue (HoQ) and ending when an ACK frame confirms its successful reception. Delays faced by the dropped packets (because of exceeding the retry limit) do not contribute to it [90, 92]. The probability that a packet is successfully transmitted is $1 - e_c = 1 - p_f^{m+1}$. A packet may face delay at each retry stage and the probability that it reaches stage i and is not dropped (i.e., it is successfully transmitted at some later stage) is given by $\frac{p_f^i - p_f^{m+1}}{1-p_f^{m+1}}$. At retry stage i, a packet can face a $\frac{W_i+1}{2}$ number of slots on average. In total, the packet has to wait for an n_{slot} number of slots as HoQ which is presented as

$$n_{slot} = \sum_{i=0}^{m} \frac{(W_i+1)\left(p_f^i - p_f^{m+1}\right)}{2\left(1-p_f^{m+1}\right)}. \tag{3.61}$$

Using the above expression for the number of slots, the channel access delay can be estimated as

$$d_c = n_{slot}\, t_{slot}. \tag{3.62}$$

The delay and loss in the queue and medium access as derived in Sections 3.7 and 3.8 using the above steady state probabilities can now directly be fed into the E-model in Section 3.3 to determine voice quality.

3.9 Modified Assumptions on Idle Channel

Ziouva and Antonakopoulos [88] proposed an extension of the Markov model by conditioning the counter decrement procedure while the channel is idle. Although Bianchi and Tinnirello [87] mentioned it as only a modified assumption, we also explore this

extension in light of our model. According to [88], the counter values are decremented with probability $(1 - p_b)$ and

$$P\{i, k|i, k+1\} = 1 - p_b \quad \text{for } 0 \leqslant i \leqslant m \text{ and } 0 \leqslant k \leqslant (W_i - 2). \tag{3.63}$$

Applying this condition to our model, with consideration for power capture and unsaturated traffic, the transmission probability becomes τ' as shown in (3.64).

$$\tau' = \frac{2p_q(1-2p_f)(1-p_f^{m+1})(1-p_b)}{\left(\begin{array}{r} p_q W\left(1-p_f\right)\left(1-(2p_f)^{m'+1}\right)+2p_f^m\left(1-p_f\right)\left(1-2p_f\right)\left(1-p_q\right)\left(1-p_b\right) \\ +p_q p_f^{m'+1}\left(1-2p_f\right)\left(2^{m'}W+1\right)\left(1-p_f^{m-m'}\right)+p_q\left(1-2p_f\right)\left(1-p_f^{m'+1}\right) \\ +2\left(1-p_f\right)\left(1-2p_f\right)\left(1-p_q\right)\left(1-p_f^m\right)\left(1-p_b\right) \end{array}\right)}. \tag{3.64}$$

The modified idle channel condition is claimed to have achieved a greater accuracy by Ziouva and Antonakopoulos [88] in modeling the DCF mechanism. Additionally, they used a different state in their Markov model where a station transmits without contention when the channel is idle following a successful transmission. However, Bianchi's method is more popular (by use and citation) in the literature and is well proven using simulations. Although we showed that our model can adapt to the above modification through the use of (3.64), the analyses and simulations presented in this thesis are done using (3.55).

3.10 Markov Model Validation

To validate the proposed Markov model, we use both analytical and simulation based approaches in the following sections.

3.10.1 Analytical Validation

For validation purposes, in this section we show that using appropriate assumptions, our model reduces to the previous well accepted models. Our proposed model is closest to [90], the differences being provisions for power capture and unsaturated traffic condition which are considered in this thesis. If capture effect is ignored, i.e., $p_p = 0$, the transmission failure probability p_f in (3.38) reduces to $p_c + p_e - p_c\,p_e$. When a saturation load is applied, every station will always have some packets to send which gives, $b_E = 0$ and $p_q = 1$. With these assumptions, the transmission probability τ in (3.55) reduces to τ_1 for $m > m'$ which is shown in (3.65).

Table 3.5: Simulation parameters for performance analysis of 802.11 WLANs.

Parameter	Value	Parameter	Value
SIFS	10 μs	Idle slot	20 μs
DIFS	50 μs	Propagation delay	1 μs
PLCP header	144 bits	Preamble	48 bits
Data rate	1 Mbps	Basic rate	1 Mbps
PLCP rate	1 Mbps	ACK frame	14 bytes
RTS frame	20 bytes	CTS frame	14 bytes

$$\tau_1 = \lim_{\substack{p_q \to 1 \\ m > m'}} \tau = \frac{2(1-2p_f)(1-p_f^{m+1})}{\left(\begin{array}{r} W(1-p_f)\left(1-(2p_f)^{m'+1}\right) + (1-2p_f)\left(1-p_f^{m+1}\right) \\ +p_f^{m'+1} 2^{m'} W\,(1-2p_f)\left(1-p_f^{m-m'}\right) \end{array} \right)}. \tag{3.65}$$

The condition $m < m'$ is meaningless since in this case, effectively the retry limit becomes the new number of retry stages and m becomes equal to m'. We present the $m \leqslant m'$ condition, which is the same as $m = m'$, for analytical verification by using the model in [89]. For $m \leqslant m'$ we define the modified transmission probability τ_2 as

$$\tau_2 = \lim_{\substack{p_q \to 1 \\ m \leqslant m'}} \tau = \frac{2\,(1-2p_f)\left(1-p_f^{m+1}\right)}{W\,(1-p_f)\,(1-(2p_f)^{m+1}) + (1-2p_f)\left(1-p_f^{m+1}\right)}. \tag{3.66}$$

Both τ_1 and τ_2 match the expressions shown by Wu *et al.* [89].

When the backoff mechanism is not used, i.e., $m = m' = 0$, τ becomes independent of p_f. With the saturation traffic assumption as before, we find τ_3 as

$$\tau_3 = \lim_{\substack{m \to 0 \\ m' \to 0 \\ p_q \to 1}} \tau = \frac{2}{W+1}, \tag{3.67}$$

which matches the constant backoff window model by Bianchi *et al.* [85].

3.10.2 Validation using Simulation

We modeled the simulations in Network Simulator 2 (NS-2.28) which is widely used by network researchers. The details of simulations using NS-2 are presented in Appendix A.1. We found a very agreeing match, as shown in this section, between the simulation results and our theoretical analyses which validates our proposed Markov model. To carefully study the effect of network traffic load and packet arrival rate, a separate Poisson agent was also developed along with a corresponding Null agent.

These classes were made open source and posted in the NS-2 official mail group for comment and public use. We used static wireless nodes, deployed randomly and uniformly, each with a transmission range of 200m. The NOAH routing protocol was employed to update the routing tables. Each simulation result provided here is the average of 10 runs where each simulation run is 2000 seconds long. We tested the simulations using IEEE 802.11b WLAN nodes with DSSS based physical layers but the method is applicable to any 802.11 variant. The parameters used in the simulation are shown in Table 3.5 and are the same as in many other studies including [19, 89, 93]. All simulation results plotted in this section yielded $\rho < 10^{-8}$ at 99% confidence interval.

Figure 3.6 shows the throughput ψ as a function of the packet arrival rate λ for different payload sizes. We used a wide range of payload sizes (100∼5000 bytes) and a network of 10 stations. The simulation results match the theoretical data closely. The throughput increases almost linearly with increasing arrival rate up to a certain point, beyond which it remains almost unchanged. The first region indicates unsaturated network condition where $\lambda \ll \frac{1}{d_c}$ while the second region indicates a nearly saturated or saturated condition where $\lambda \geqslant \frac{1}{d_c}$. The smooth transition between the two regions is where a network switches into saturation and the packet arrival rate is almost equal to the service rate (e.g., at $\lambda = 20 \sim 30$ packets/sec for l_d=500B). This transition point is clearly a function of the payload size which occurs at a lower λ when the payload is larger.

For a larger payload, the transmission time becomes longer which makes the transmissions more susceptible to collisions and errors. Thus, the channel access delay becomes longer saturating the network even with a small packet arrival rate. A longer channel access delay incurs relatively higher degradation in the throughput since the channel is kept busy for each packet transmission. However, as the larger payload contributes to the total throughput, a larger payload still gives a higher throughput even though the channel access delay is longer in this case. This is why the rate of increase in the throughput under the unsaturated condition is higher for larger packets as the same throughput can be achieved by transmitting larger packets at a lower packet arrival rate. For the rest of this section, we use $l_d = 1000$ bytes as a representative data payload size.

The effect of packet error rate is presented in Fig. 3.7 for a 10-node WLAN. These results show that although the packet error rate decreases the saturation throughput considerably, its impact is negligible under the unsaturated condition. In our simulations, channel errors are introduced uniformly to the received packets to reflect a uniform random packet loss scenario. Since the number of transmissions under the unsaturated condition is quite low, fewer retransmissions for each packet take place due to channel errors. Therefore, the impact of an imperfect channel under the unsat-

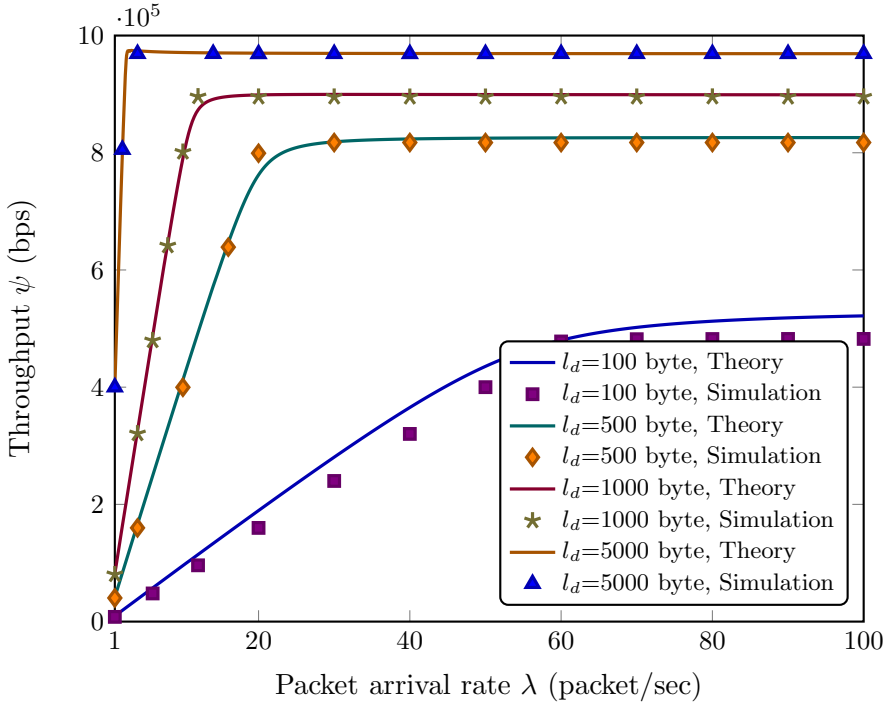


Figure 3.6: The effect of the packet size on throughput for different packet arrival rates with simulation results in a 10-node 802.11b WLAN using 1 Mbps data rate.

urated condition is insignificant. On the other hand, under the saturated condition, a large number of transmissions and retransmissions take place and, as a result, the impact of erroneous packets on overall throughput becomes greater.

Figure 3.8 illustrates the impact of capture threshold on the network throughput in a 10-node WLAN. The capture threshold defines the minimum acceptable ratio of the received power of the primary packet (the one received with maximum power) to the sum of received power of every other sensed packets. Therefore, a lower threshold gives a higher probability of a collided packet to be received correctly, i.e., captured. A higher threshold forces more of the collided packets to be rejected by the receiving station. Like the packet error rate, it has less impact in an unsaturated network. Since there are only a few collisions in this region, the number of captures is even smaller which reduces the impact of the capture effect.

We next study the effect of the network size (number of stations) which is shown in Fig. 3.9. Since the payload is kept constant at 1000 bytes, the saturation throughput is the same, i.e., 0.9 Mbps, in all cases. But under the unsaturated traffic condition, the rate of increase in throughput varies with the network size. The packet arrival rate and the number of nodes conjunctively define the traffic load of a network and give the rate of total packet generation (by all nodes). The number of nodes is also the number of competitors in a channel. Therefore, even if the packet generation rate and payload remain the same, a higher number of nodes may lead to more frequent collisions which,

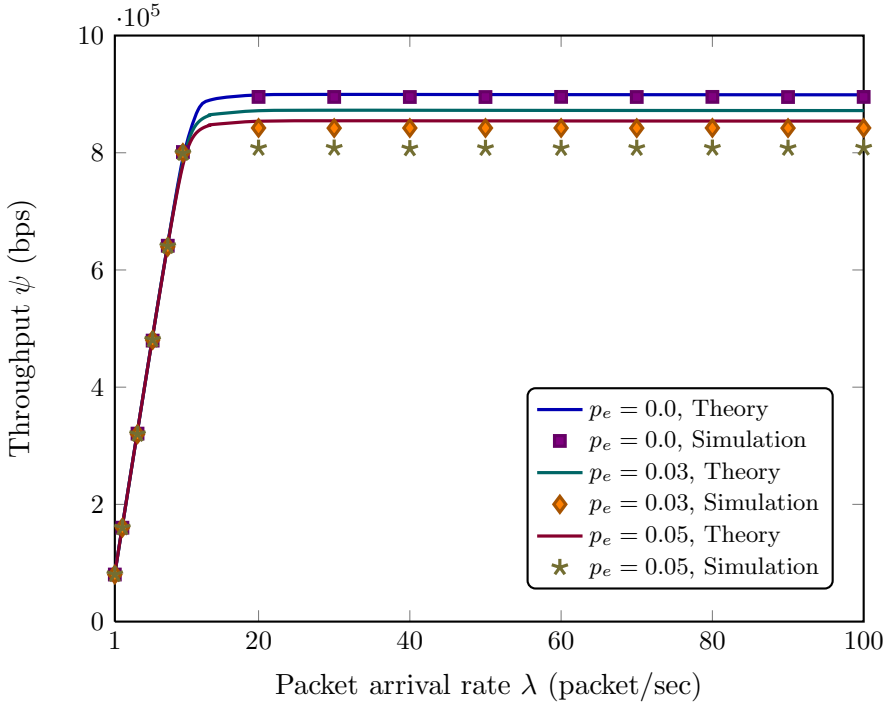


Figure 3.7: The effect of packet error rate p_e on the throughput at different packet arrival rates with simulation results in a 10-node WLAN using 1 Mbps data rate.

in turn, will result in a higher channel access delay and lower throughput. The results demonstrate this phenomenon where, everything else being unchanged, a network with a higher number of nodes become saturated even with a much lower packet arrival rate. Although the 4-node WLAN in Fig. 3.9 did not reach saturation until $\lambda \geqslant 32$, the 15-node network was saturated even at $\lambda = 9$. Therefore, the network size plays an important role in defining the saturation traffic load of a network, its impact being much higher than even the packet arrival rate of individual nodes.

These simulation results match the analytical results derived from our Markov model and substantiates our proposed model of the DCF based medium access mechanism. We now use the Markov model of DCF based medium access (Section 3.8) and the queuing model (Section 3.7) to formulate a voice capacity model in the following section.

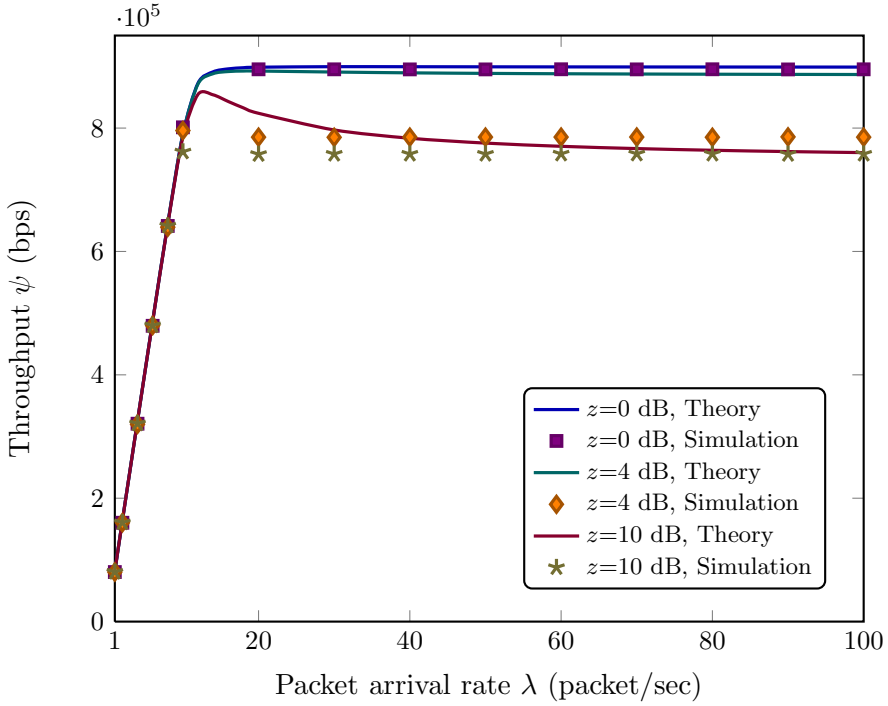


Figure 3.8: The effect of capture threshold z (dB) on the throughput for different packet arrival rates with simulation results in a 10-node WLAN using 1 Mbps data rate.

3.11 Call Capacity Model for Single Hop, Single Channel WLAN

Using the analyses in Sections 3.3∼3.8 and (3.1)∼ (3.62) we formulate the VoIP call capacity of a 802.11 WLAN by using following optimization problem.

$$\begin{aligned} &Max \quad n_s \\ &s.t. \quad I_d(d_e) + I_{e\_eff}(e_e) \leqslant \begin{cases} 28.3553 & \text{for medium quality,} \\ 13.1952 & \text{for high quality,} \end{cases} \end{aligned} \tag{3.68}$$

where

$$d_e = d_l + n_a d_f + d_i + d_q + d_c + d_j, \text{ and}$$
$$e_e = e_q + (1 - e_q)\, e_c + (1 - e_q - (1 - e_q)\, e_c)\, e_j.$$

Here, we try to maximize the number of calls (which is the same as the number of stations n_s), i.e., we add voice calls as long as the sum of the impairments due to the end-to-end delay and loss is less than the budget of total impairment as shown in (3.16). The impairments are estimated using the delay and loss in the queue and medium access. The delay and loss in the queue are functions of the queue length, arrival rate and departure rate. The departure rate is a function of the delay and

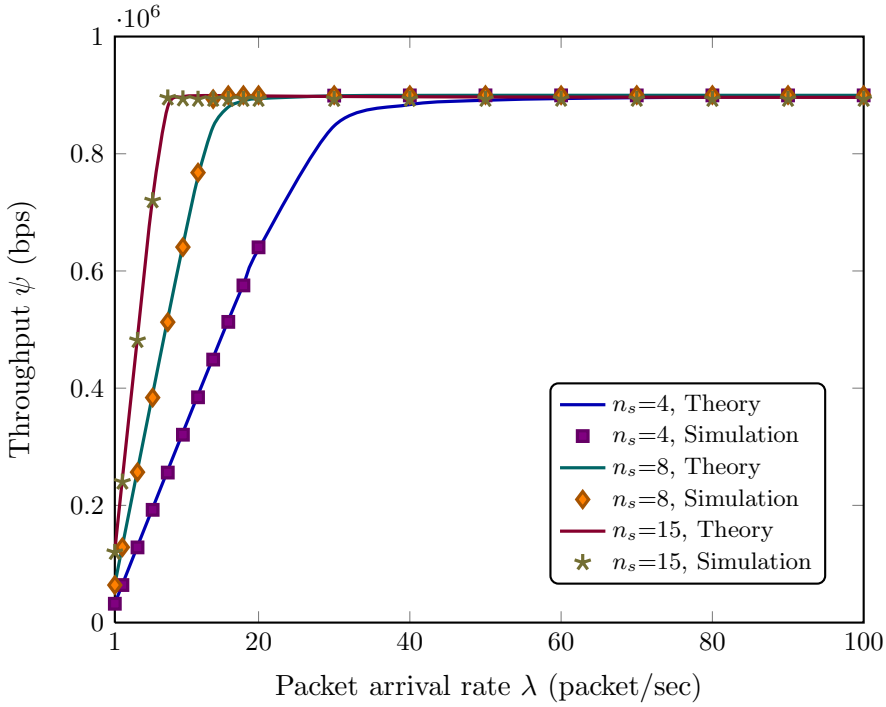


Figure 3.9: The effect of the number of stations n_s on the throughput at different packet arrival rates with simulation results in a 10-node WLAN using 1 Mbps data rate.

Table 3.6: Simulation parameters for voice capacity analysis in 802.11 WLANs.

Parameter	Value	Parameter	Value
UDP header	8 bytes	RTP header	12 bytes
IP header	20 bytes	MAC header	28 bytes
PLCP header	48 bits	Preamble length	144 bits
DIFS	128/50 μs	Queue length	50 packets
SIFS	28/10 μs	Propagation delay	1 μs
Idle slot	50/20 μs	Control data rate	1 Mbps
Data rate	1/11/54 Mbps	PLCP data rate	1 Mbps

loss in medium access which, in turn, depend on the network size, packet arrival rate, packet size and the DCF mechanism. The network can support the calls as long as the queue does not grow and the total voice quality impairment due to the end-to-end delay and loss remains less than or equal to the impairment budget.

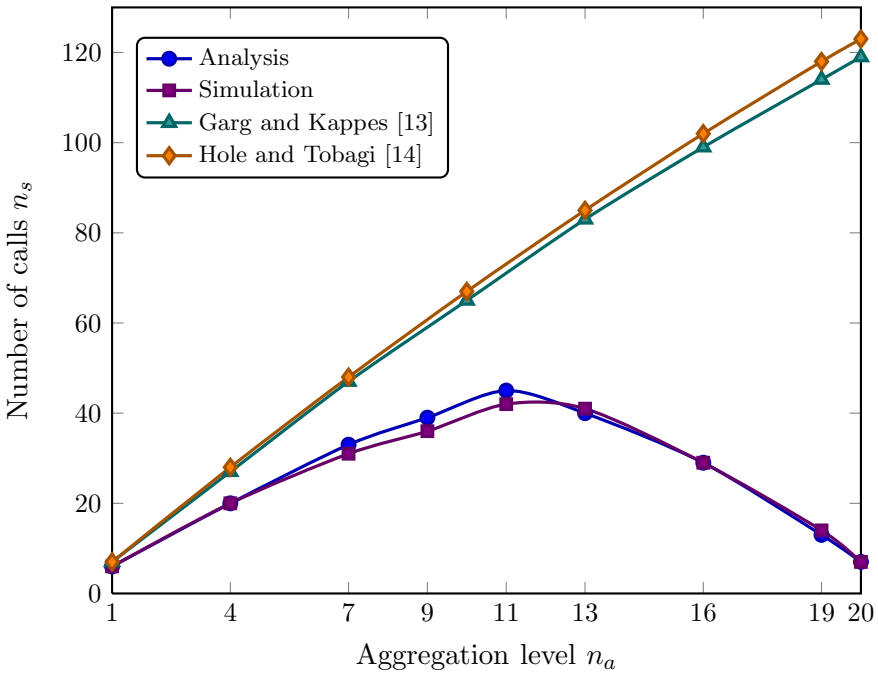


Figure 3.10: The VoIP call capacity of the G.729 codec at different aggregation levels in a 802.11b WLAN with 11 Mbps data rate.

3.12 Results and Analyses

We carried out extensive simulation in NS-2 (both 2.32 and 2.28) which is widely used by network researchers. In our simulations, IEEE 802.11 (802.11a and 802.11b) nodes were statically deployed randomly and uniformly. The NOAH routing protocol and a transmission range of 200 m were used. Each simulation result presented here is an average of 10 runs where each simulation run is 5000 seconds long. We tested the simulations for both 802.11a with FHSS and 802.11b with DSSS based PHY layer but our proposed method is applicable to any 802.11 variant. The parameters used in the simulation are shown in Table 3.6 and are the same as those used in many other studies including [19, 89, 93]. The parameters were chosen according to the 802.11 variant (802.11a or 802.11b) in use. All simulation results plotted in this section yielded $\rho < 10^{-8}$ at a 99% confidence interval.

3.12.1 VoIP Call Capacity for Different Codecs

The theoretical and simulation results for the call capacity of two way medium quality calls over a single hop WLAN using DSSS based 802.11b and FHSS based 802.11a are illustrated in Figs. 3.10∼3.12 and Figs. 3.13∼ 3.15, respectively. We used the G.729, G.711 and G.723.1 codecs to reflect the effect of codec characteristics. The call capacity is plotted as a function of the aggregation level n_a to emphasize its effect. In

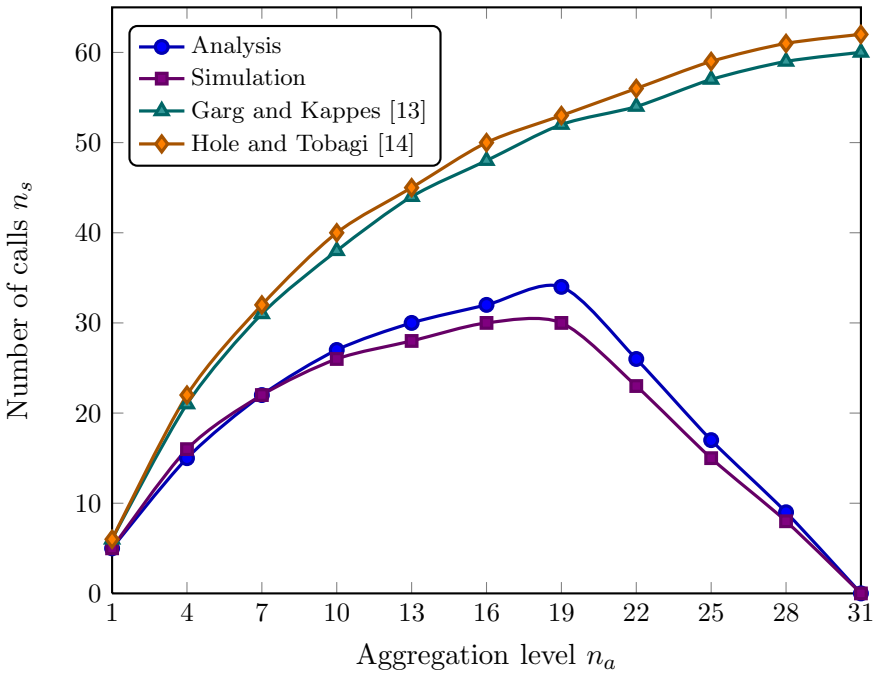


Figure 3.11: The VoIP call capacity of the G.711 codec at different aggregation levels in a 802.11b WLAN with 11 Mbps data rate.

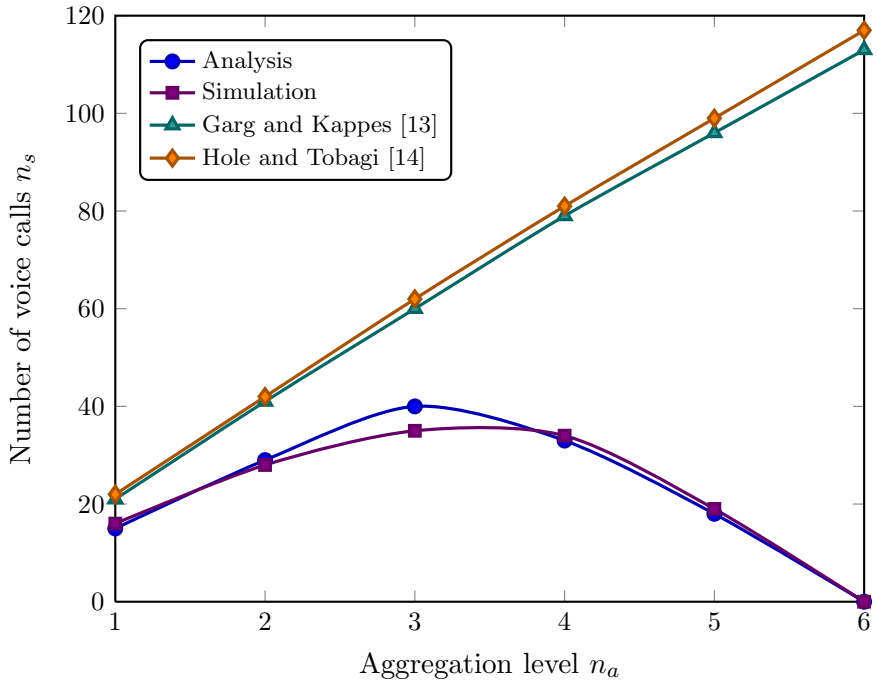


Figure 3.12: The VoIP call capacity of the G.723.1 codec at different aggregation levels in a 802.11b WLAN with 11 Mbps data rate.

all cases, our theoretical estimations show close matches with the simulation results which validates our approach in estimating the VoIP call capacity. For comparison purposes, the figures also present the call capacity estimated by Garg and Kappes [13] and Hole and Tobagi [14] which are the most cited works among the existing call capacity analyses. The results show the over-estimation in [13, 14] which is due to ignoring the effects of the queuing delay, queuing loss and dejitter buffer delay. Moreover, the call capacity in those models increase monotonically which is due to ignoring the effect of the packetization interval. Results from [104] are not compared as [104] mostly presented delay from a tested using only one codec (G.711) with $n_a = 2$. As a gross comparison, with similar configuration, we found the VoIP call capacity to be 9 (10 in our simulation) which is reported as 15, 12 and 12 by Shin and Schulzrinne [104], Garg and Kappes [13] and Hole and Tobagi [14], respectively. The overestimated capacity in [104] resulted from ignoring the delay and loss in queue and dejitter buffer.

When the aggregation level is smaller, the packetization interval is also smaller giving a higher packet arrival rate (packet arrival rate $\lambda \propto \frac{1}{n_a}$) leading to a lower call capacity. In this case, with even a smaller number of nodes, the queue usage λd_c becomes close to 1; and as soon as λd_c becomes greater than 1, the queue starts to grow. As a result, the queuing delay and queuing loss become very high resulting in a drastic degradation of the voice quality. With an increase in n_a (d_f is constant for a given codec), λ decreases, and the call capacity increases up to a certain point and then decreases again. The reason behind this later decrease is that the performance gain achieved by keeping λ low does not pay off for the increase in the end-to-end delay due to the higher packetization delay $n_a d_f$. When the quality impairment incurred by this increased packetization delay exceeds the benefit of keeping the packet arrival rate low (by keeping n_a high), the call capacity starts to decrease. One important point to note is that the user perceived quality (UPQ) is the same (i.e., the minimum acceptable rating factor $R = 70.07$ for medium quality calls) at any point of the plots representing our theoretical analysis. Therefore, network designers should try to attain the highest point in the curve where the voice call capacity is maximized.

3.12.1.1 IEEE 802.11b with DSSS based PHY

Comparison of Figs 3.10∼3.12 for IEEE 802.11b WLANs shows that G.729 supports more medium quality calls than G.711 (45 vs. 34) since G.711 generates larger packets and consumes greater bandwidth. Although G.723.1 uses higher compression than G.729, it supports fewer calls (only 40 calls are supported) due to its higher value of initial equipment impairment factor at zero loss (I_e is 15 for G.723.1 and 11 for G.729 as shown in Table 3.1).

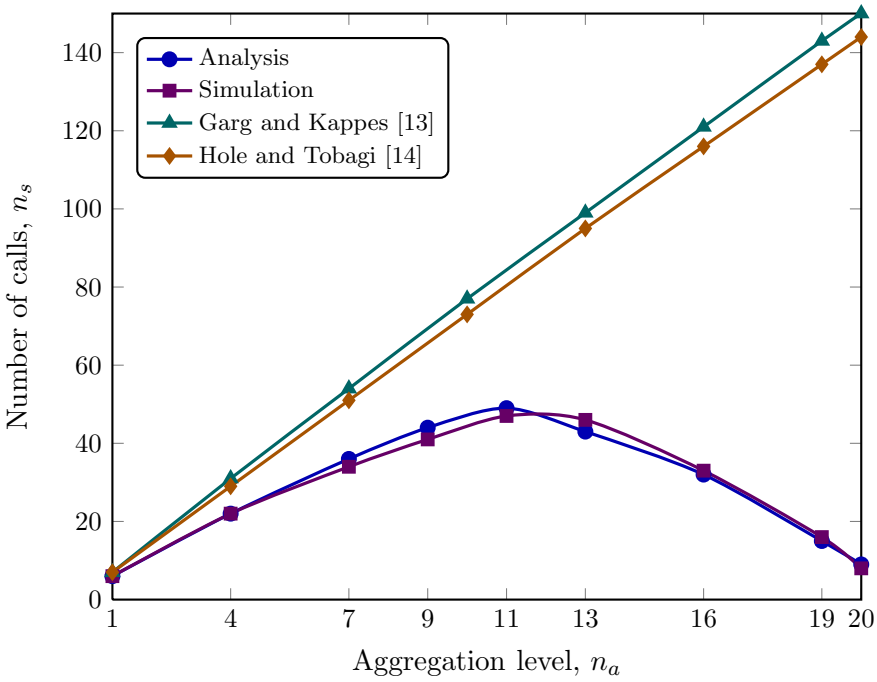


Figure 3.13: The VoIP call capacity of the G.729 codec at different aggregation levels in a 802.11a WLAN with 54 Mbps data rate.

3.12.1.2 IEEE 802.11a with FHSS based PHY

A similar trend is found in the FHSS based 802.11a (54 Mbps) WLANs except the peak capacity is now reached at a slightly lower value of n_a indicating its increased effect. In this case, the maximum numbers of supported two-way calls are 49, 58 and 44 for G.729, G.711 and G.723.1, respectively. A comparison of results for 11 Mbps (802.11b) and 54 Mbps (802.11a) indicates that, although the data rate is increased almost 5 times, the increase is capacity is minimal. Using the G.729 codec, only 4 additional calls can be supported when the data rate is increased from 11 Mbps to 54 Mbps (Fig. 3.10 vs. 3.13). Since the data rate is used to transmit data packets only, increasing the data rate does not improve the network performance proportionately for VoIP packets which are small in size. Control packets, preambles and PLCP headers are always transmitted at the lowest data rate. Moreover, the nodes spend a considerable amount of time in the backoff state which also does not change proportionately with an increase in the data rate. Only for G.711, which uses less compression and generates larger packets, the gain in capacity is relatively higher, i.e., 24 additional calls are now supported. Although G.729 performs better than G.711 in the 11 Mbps WLAN, it performs worse than G.711 in the 54 Mbps WLAN due to the high value of I_e (which is 11 for G.729 and 0 for G.711 as shown in Table 3.6).

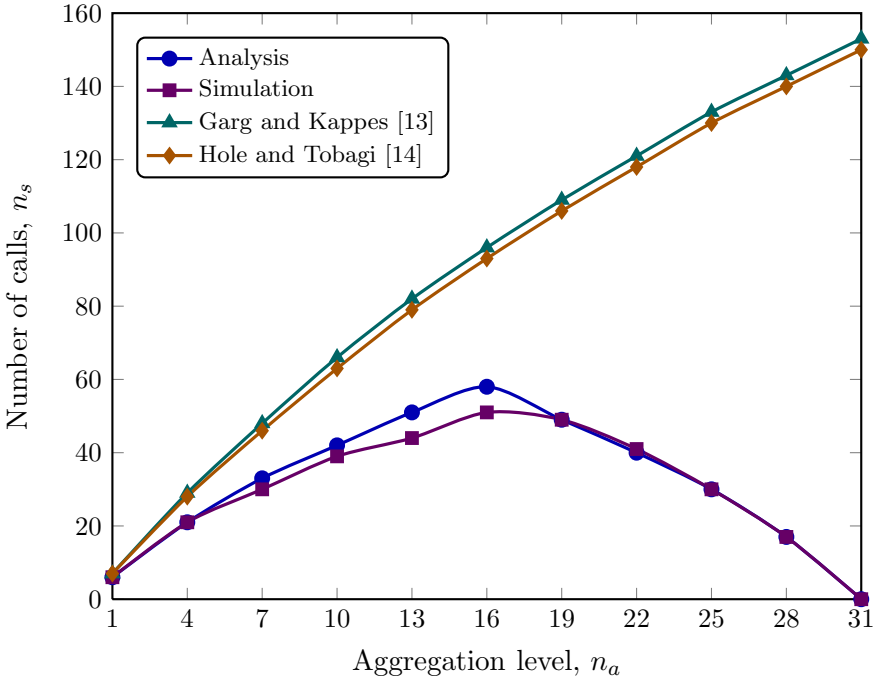


Figure 3.14: The VoIP call capacity of the G.711 codec at different aggregation levels in a 802.11a WLAN with 54 Mbps data rate.

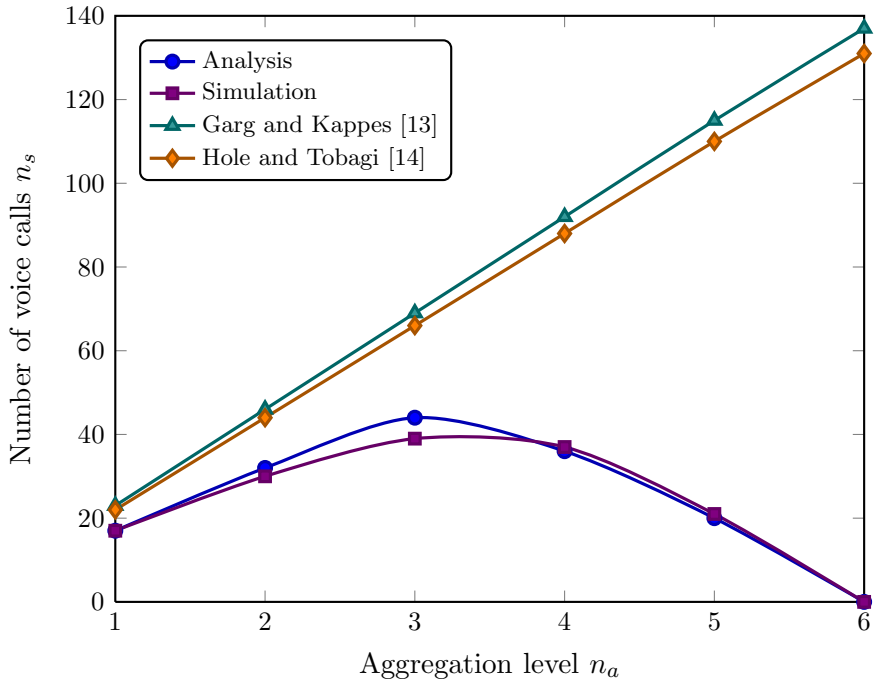


Figure 3.15: The VoIP call capacity of the G.723.1 codec at different aggregation levels in a 802.11a WLAN with 54 Mbps data rate.

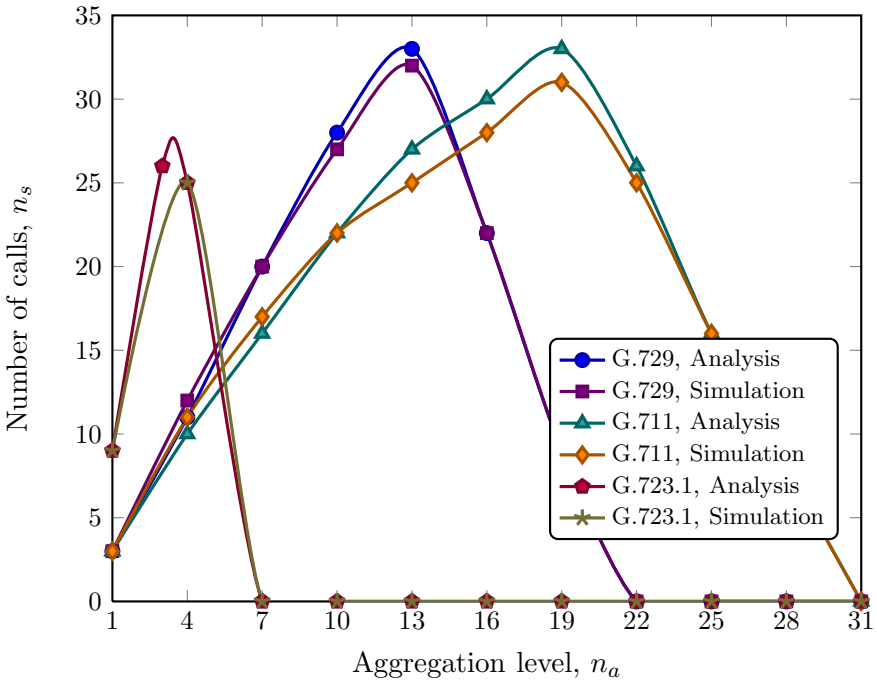


Figure 3.16: The VoIP call capacity of the G.729, G.711 and G.723.1 codecs at different aggregation levels using the RTS/CTS based handshake in a 802.11b WLAN with 11 Mbps data rate.

3.12.2 Capacity with RTC/CTS DCF

Figure 3.16 presents the performance of VoIP traffic employing the RTS/CTS mechanism over an 11 Mbps WLAN. Although the RTS/CTS mechanism minimizes the effect of the hidden node and exposed node problems resulting in fewer collisions of data packets, the overhead of RTS/CTS packet transmissions does not pay off for small VoIP packets. Therefore, RTS/CTS is found to support fewer voice calls compared to the basic DCF handshake for all codecs. The call capacity using the RTS/CTS and basic handshake mechanisms are found to be 33 vs. 45, 33 vs. 34 and 26 vs. 40 for G.729, G.711 and G.723.1, respectively. Only G.711 shows less degradation in capacity as it generates larger packets keeping the impact of the overhead of transmitting the RTS and CTS packets relatively lower.

3.12.3 Effect of Call Quality

A comparison of high quality versus medium quality calls for different codecs is shown in Fig. 3.17. A moderate number of high quality calls is supported by the G.711 codec (28 calls) whereas the G.729 and G.723.1 codecs support nearly zero high quality calls. The impairment budget for high quality calls is 13.1952 as shown in (3.16) but the

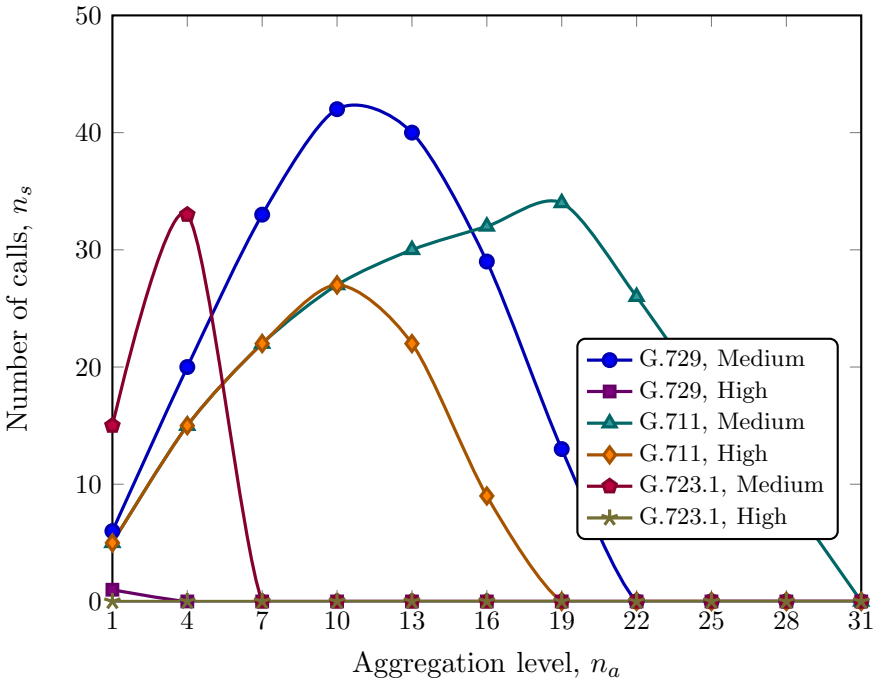


Figure 3.17: The effect of voice call quality requirement (impairment budget for high and medium quality calls) on the VoIP call capacity of the G.729, G.711 and G.723.1 codecs at different aggregation levels.

initial equipment impairments of the G.729 and G.723.1 codecs are very high even with zero packet loss which abnegates any high quality calls.

3.12.4 Voice Quality in Unsaturated WLAN

Figure 3.18 illustrates the voice quality in an unsaturated WLAN. In this case too, we find a close match between our analyses and simulations. To ensure that the network does not become saturated, the WLAN size is kept to 5 stations. As the aggregation level increases, the packet arrival rate decreases. The MAC layer delay d_c decreases only slightly with any decrease in the packet arrival rate under the unsaturated condition but the packetization delay increases with an increase in the aggregation level leading to a higher end-to-end delay. As a result, the UPQ rating R decreases nonlinearly with increasing n_a. Since e_c, e_q and e_j can be expected to be insignificant in an unsaturated network, the effective equipment impairment factor should have a fixed value (due to the initial equipment impairment I_e). Therefore, the UPQ, in this case, becomes a representative of the delay impairment only. A similar magnitude of R for G.711 at $n_a = 2$ is reported from simulations in [104] although we observed a slightly higher R for due to a smaller network size used in our results.

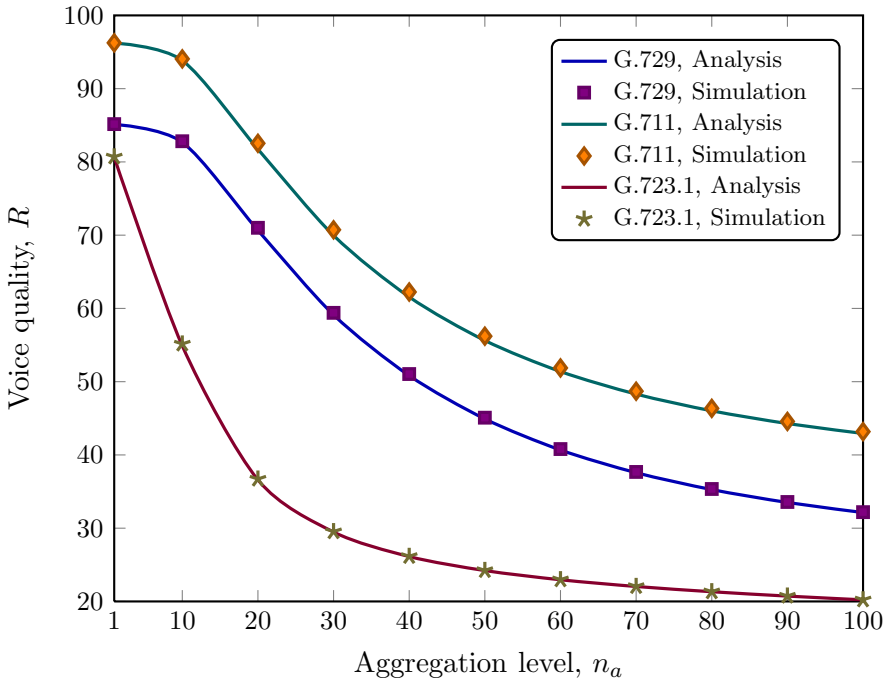


Figure 3.18: The voice quality rating R score (determined using the ITU-T E-model) variation with different aggregation levels for the G.729, G.711 and G.723.1 codecs in an unsaturated WLAN of 5 nodes.

3.13 Key Contributions

In this chapter, we introduced a novel call capacity estimation model for VoIP over 802.11 WLANs. The followings are the key contributions that differentiate our model from other relevant works.

- Despite the ease of estimating call capacity using the bandwidth usage per call or channel access delay, such methods are inadequate in giving a true estimate of the voice quality since voice quality is degraded by the conjunctive effect of delay and loss in the end-to-end path. To ensure an acceptable level of voice quality, we utilized the ITU-T E-model which considers codec characteristics and the conjunctive effect of the end-to-end delay and loss and, thereby, ensure that voice quality does not degrade when a WLAN is designed using our call capacity estimation model.

- The existing analytical works in the literature used over simplified assumptions in determining the delay and loss in the channel access and, thereby, overestimated the VoIP call capacity. The IEEE 802.11 DCF based medium access mechanism uses a complex CSMA/CA mechanism that results in non-linear changes in the

network performance measures with respect to the offered load. This renders the results presented in the existing works to be unreliable since they assumed fixed delay in channel access. We designed and utilized a Markov model to model the finite retry BEB based DCF mechanism and determined the exact changes in the delay and loss in channel access. Extensive simulations show a close match with our numerical analyses which validates our Markov model. The Markov model allows us to determine the voice call capacity with reasonable accuracy by providing exact estimates of the delay and loss in DCF based medium access.

- The delay and loss in the queue degrade the voice quality severely which is ignored in the existing studies. We modeled the queue considering its service and arrival rates and buffer size, and determined the queuing delay and queuing loss as experienced by the voice packets. Considering the impact of queue performance on the voice quality allows us to avoid any overestimation in determining the voice capacity.

- Most existing works assumed an ideal error-free channel which is possible only in theory. Every real channel suffers from noise and interference and, therefore, considering an ideal channel in determining voice capacity results in an overestimation. We considered the packet error rate in estimating the probability of a transmission failure to reflect the effect of an imperfect channel.

- Most WLAN devices support power capture today which can improve the network performance and increase the voice capacity. But power capture is ignored in the existing analytical models which resulted in an under-estimation of the call capacity. We considered power capture in a Rayleigh fading channel in our capacity model where the capture probability is parameterized to allow the use of any other fading channels as well.

- Most WLANs operate under the unsaturated traffic condition for a considerable amount of time but the voice quality in an unsaturated WLAN was never investigated. Our proposed Markov model considers both unsaturated and saturated traffic conditions and simulation results show a close agreement to our numerical analysis for voice quality in unsaturated WLANs.

3.14 Summary

In this chapter, we presented an analytical model to estimate the VoIP call capacity of IEEE 802.11 WLANs. The ITU-T E-model is used to determine the voice quality impairments in the end-to-end path. An impairment budget was formulated

for medium and high quality calls using the ITU-T E-model guidelines. We used a Markov model to determine the delay and loss incurred by the finite retry BEB based DCF mechanism. In conjunction to the Markov model, we used a $M/M/1/s_q$ system to determine the delay and loss the queue. We presented the call capacity using an optimization problem that can be used in network design and planning so that voice quality is maintained in such networks. Most public WLANs are multihop where users can be more than one hop away from the access point. The end-to-end delay and loss in such networks are completely different from those in a single hop WLAN since the intermediate nodes forward traffic on behalf of nodes in the outer areas and suffer from a higher traffic load. Additionally, we recall from the previous chapter (Section 2.3) that the IEEE 802.11 standards support multiple channels and up to three channels (four in Europe) can be utilized simultaneously without significant cross-channel interference. The utilization of multiple channels can increase the voice call capacity significantly and needs to be considered carefully. Additionally, most future WLAN devices are expected to bear more than one network interfaces [21] which should also be considered in the light of the voice capacity. In the following chapter, we model the VoIP call capacity of a multihop, multi-channel WLAN where the client nodes are equipped with multiple network interfaces.

Chapter 4

Voice Capacity of Multi-channel, Multihop WLANs

4.1 Introduction

In the previous chapter, we modeled the VoIP call capacity of a single channel, single hop WLAN considering the voice quality requirements according to the ITU-T E-model guidelines. We designed a Markov model to determine the exact delay and loss in the DCF based medium access. In conjunction to the Markov model, a $M/M/1/s_q$ queuing system was used to model the queue of a WLAN node that had only a single network interface. We recall from Section 2.3 that the IEEE 802.11 standards support the use of multiple channels and up to three channels (four in Europe) can be used simultaneously without significant inter-channel interference. When a large number of nodes compete for medium access in a single channel, the DCF based medium access incurs high delay and loss due to frequent collisions and high backoff delay. But when multiple channels are utilized, the nodes are divided into a number of channel-wise groups where fewer nodes compete in each channel and the number of collisions decreases. As a result, the delay and loss in the medium access decrease which improves the voice performance substantially. Collisions can also be reduced by the use of multihop WLANs where the nodes are distributed over a number of localized areas or collision domains so that a node competes only within its own domain with fewer nodes. Additionally, a multihop WLAN offers a larger coverage than a single hop WLAN using the same number of access points. Furthermore, future mobile devices are likely to be equipped with multiple network interfaces [21] which allows a node to operate in multiple channels simultaneously. Considering all the above mentioned factors in determining the voice capacity is of utmost importance and should be carefully analyzed during the network design and planning. But no existing analytical model of the VoIP call capacity in the

current literature addressed any of the above three factors. In this chapter, we model the voice call capacity of a multi-channel, multihop WLAN where the client nodes are equipped with multiple network interfaces.

4.2 Key Aspects in Multi-channel, Multihop WLAN with Multi-interface Nodes

As mentioned in the previous section, no existing *analytical* model in the current literature on VoIP call capacity of the IEEE 802.11 WLANs considered any of the following three factors, namely, multiple channels, multiple network interfaces and multihop scenario. Each of these factors has drastic impact on voice capacity and, therefore, should be considered carefully in any call capacity analysis.

Multiple Channels A number of simulation based works [107, 108, 110, 111] investigated the call capacity of wireless mesh networks which show that considerable performance improvement can be achieved by employing multiple channels. The IEEE 802.11 standards define 13 channels (and 14 in some countries) of which three are non-overlapping (four in Europe) and can be used simultaneously without significant inter-channel interference. When multiple channels are utilized simultaneously, the contention for channel access among the nodes is reduced since the nodes are divided into smaller groups as they are distributed among the available channels and, therefore, fewer nodes now compete in each channel. As a result, the throughput increases and the delay and loss in the medium access decrease. Therefore, a substantial increase in the voice capacity can be expected by utilizing multiple channels.

Multiple Radio Network Interfaces (NI) Availability of multiple network interfaces can improve reliability since wireless networks are prone to temporary outages due to traffic overload and noise. The most popular (and commercially successful) version among the IEEE 802.11 standard is 802.11b which operates in the 2.4 GHz spectrum. Although both the 5.5 GHz (used in 802.11a) and the 2.4 GHz spectrums are free, the 2.4 GHz ISM spectrum is heavily used by many home appliances (including baby monitors, security cameras, microwave ovens, cordless phones, etc.) due to the low manufacturing cost and low current drain of such devices. As a result, packet transmissions in the 2.4 GHz band suffer from high noise and interference. Moreover, due to the ad hoc nature of operation, any roaming node can join a public Wi-Fi network without warning which causes increase in the traffic load. These factors can result in the temporary outage of a crowded channel. Client nodes equipped with multiple interfaces can be

connected to multiple channels simultaneously and avoid connection drop of any ongoing voice call in such a case. But the utilization of multiple NIs increases the number of competitors in a channel since each NI tries to access the available channels individually and transmit packets. This can result in increased contention and, therefore, the effect of multiple NIs on the call capacity should be carefully analyzed.

Multihop Routes When wireless coverage is needed for a large area, some areas may fall outside the transmission range of the AP. However, the DCF based medium access mechanism allows the nodes to compete among themselves and access the channel without any coordination of the AP. The Ad hoc routing protocols set up multihop routes and the nodes which are not within the transmission range of the AP can still send data packets to and receive from the AP via other nodes on respective routes. In this case, the intermediate nodes forward the data packets (along with their own packets) to and receive from the AP on behalf of the nodes that are located far from the AP. A multihop WLAN with a single AP can provide the wireless coverage (and the Internet connectivity) to a larger area than a single hop WLAN. For instance, in a corporate WLAN or home network of a large house the node density is low (compared to public hotspots) but the nodes are located over a wide area. In these cases, multihop WLANs will be useful. Since fewer access points are required in the multihop scenarios to provide the Internet connection to all nodes, the use of multihop WLANs can reduce the operational costs substantially.

To assist in the design and planning of multihop, multi-channel WLANs with multi-NI nodes, we design an analytical call capacity model in the following sections considering the voice quality requirements. All real network aspects introduced in the previous chapter are also considered.

4.3 Modeling a Multihop, Multi-Channel WLAN

In this section, we model the offered traffic load and the collision domain size of a multihop, multi-channel WLAN so that the Markov model introduced in the previous chapter can be used to determine the delay and loss in channel access. We also introduce a different queuing model for the multi-interface nodes in Section 4.4.

We denote the number of channels in the WLAN and the number of network interfaces per node by n_c and n_t, respectively. The number of nodes in the WLAN is denoted by n_s, as in the previous chapter. Client nodes try to be nearer to the AP for a better signal strength and, additionally, due to the circular coverage provided by the

regular omni-directional antennas, public hotspots usually take a circular shape with the AP at the center unless obstacles, like walls, buildings, etc., shape the network otherwise. Therefore, most works in the literature considered a circular shaped multihop WLAN. Some works considered a grid layout which is primarily applicable to mesh networks. In this thesis, we focus on WLANs and without any loss of generality consider a circular shaped WLAN as done in previous works including [150].

For tractability of analysis, we split the network into a ϕ number of concentric and annulus rings as shown in Fig. 4.1. Ring 1 is at the center (around the AP) of the WLAN and ring ϕ is the outermost ring. Since the AP acts as the gateway to the Internet, all data packets must pass through it, and the nodes in the ring i relay (or forward) data packets to and receive from the AP for the nodes in all outer rings $i+1, i+2, \ldots, \phi$. Each ring is ω meters wide and the value of ω can be chosen by the network designers considering the coverage versus interference trade off which is addressed many times in the literature [151, 152].

As a signal propagates through the air, wall and earth, it is attenuated and its power level decreases exponentially with the traveled distance. A signal can be differentiated from the background noise depending on whether the received signal strength is sufficiently higher than the noise and interference. A transmitted frame retains sufficient signal strength to be received by a node that is located within the transmission range of the transmitter. However, a transmitted frame can collide with another frame whose transmitter is located within the interference range of the receiver, i.e., a transmitted signal retains sufficient signal strength to collide with another frame within the interference range. To ensure connectivity throughout the network, the ring width ω should be smaller than or equal to the transmission range so that the nodes in adjacent rings can receive each other's transmissions and every node throughout the network can maintain connectivity with the AP. On the other hand, to reduce the number of collisions, ω should be greater than half of the interference range so that nodes in non-adjacent rings can not overhear each other's transmissions, and unnecessary collisions can be avoided. Denoting the transmission and interference ranges by r_t and r_i (both in meters), respectively, the ring width should be chosen so that $r_t \geqslant \omega > \frac{r_i}{2}$. The total area under the i-th ring is given by

$$\Lambda_i = \pi\omega^2 \left(2i-1\right). \tag{4.1}$$

While a uniform node distribution can be expected in organized networks like office spaces, class rooms or conferences, hotspots are often formed in public networks like coffee shops, train stations and airports. In a hotspot, the node density is much higher in the rings nearer to the AP than in the outer rings, much like an exponential distribution. Therefore, we present analysis for both uniform and exponential node

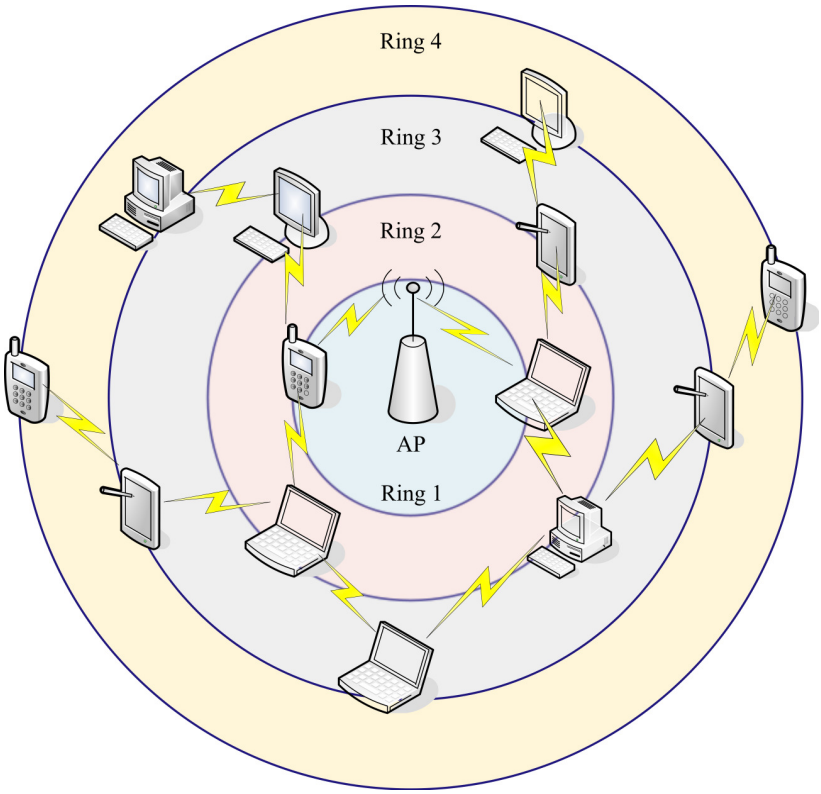


Figure 4.1: A Multihop 802.11 WLAN with the AP at the center. Intermediate nodes (nearer to the AP) forward packets to and receive from the nodes in outer rings.

distributions although our model can accommodate the use of any standard distribution. If the node distribution of a hotspot, which is very specific to a given network, can be approximated more accurately using some other distribution, the new distribution can be used instead of the exponential distribution. We denote the node density of the i-th ring with Ω_i where

$$\Omega_i = \begin{cases} \frac{n_s}{\pi(\phi\omega)^2} & \text{for a uniform node distribution,} \\ \theta e^{-\theta i} & \text{for an exponential node distribution.} \end{cases} \tag{4.2}$$

In the latter case, the total number of nodes, determined by summing over all the rings, is given by $\sum_{i=1}^{\phi} \Lambda_i \Omega_i$ and the value of θ can be obtained by solving (4.3).

$$n_s - \sum_{i=1}^{\phi} \Lambda_i \Omega_i = 0. \tag{4.3}$$

4.3.1 Collision Domain Size

The delay and loss in the medium access and queue are functions of the traffic load which is defined by a transmitting node's packet arrival rate and its collision domain

size. Collision domain of a node includes all nodes within its interference range, i.e., the nodes whose transmissions may collide with the frames received by the original node. The number of nodes within the interference range of a node in the ring i is $\pi r_i^2 \Omega_i$. While each node is a contender in a single channel network, in a multi-channel scenario each node with multiple NIs can compete in a number of channels. Each NI of a node can be tuned to a different channel and, therefore, each NI acts as a contender in the channel it is tuned to. As a result, one station can compete in n_t channels simultaneously and each NI tuned to a given channel's frequency is a contender on that channel. Since a node in the ring i has $\pi r_i^2 \Omega_i$ nodes within its interference range (as mentioned before) and each node has n_t interfaces which are connected to n_t different channels chosen from the available n_c channels. Therefore, the average number of contenders ν_i in a collision domain in the ring i is given by

$$\nu_i = \left\lceil \frac{\pi r_i^2 \Omega_i n_t}{n_c} \right\rceil . \tag{4.4}$$

The primary advantage of using multiple channels is gained from splitting a large collision domain into a number of multiple smaller collision domains since ν_i decreases with increasing n_c (i.e., $\nu_i \propto \frac{1}{n_c}$).

4.3.2 Effective Packet Arrival Rate

Let n_v denote the number of VoIP applications or connections active in each node. Most existing works considered only one two-way call per node where $n_v = 1$. But we also consider applications like teleconferencing where multiple parties can talk to each other simultaneously and $n_v > 1$. In this case, each VoIP application communicates with n_v similar peers, i.e., each node maintains n_v connections, and the encoders at both ends of each connection generate $\frac{1}{n_a d_f}$ packets per second (the notations carry their usual meanings as in Chapter 3). Since the nodes in the i-th ring forward packets to and receive from the nodes in all the outer rings, in a steady state all the packets generated in rings $i+1, i+2, \ldots, \phi$ in unit time have to be relayed by the nodes in the ring i in unit time. Therefore, the nodes in the intermediate rings need to transmit more packets than the nodes in the outer rings. We term the total packet arrival rate of the queue of a node as the effective packet arrival rate which is the sum of the packets the associated node generates itself or forwards to and receives from other nodes per second. Summing over the ring index, the total number of nodes in rings $i, i+1, \ldots, \phi$ is given by $\sum_{j=i}^{\phi} \Lambda_i \Omega_i$. The $\Lambda_i \Omega_i$ number of nodes in the i-th ring need to forward packets to and receive from all the nodes in rings $i+1, i+2, \ldots, \phi$ in addition to their own packets. Therefore, on-average a node in the i-th ring transmits all the voice packets generated by $\frac{\sum_{j=i}^{\phi} \Lambda_j \Omega_j}{\Lambda_i \Omega_i}$ nodes. For n_v number of two-way voice calls

per client node, each node generates $\frac{2n_v}{n_a d_f}$ packets per second which gives an effective packet arrival rate λ_i of a node in the i-th ring as

$$\lambda_i = \frac{2n_v \sum_{j=i}^{\phi} \Lambda_j \Omega_j}{n_a d_f \Lambda_i \Omega_i}. \tag{4.5}$$

4.3.3 Delay and Loss in Medium Access

The attributes of a collision domain are the same as those of a single hop, single channel WLAN where all the nodes can hear each other's transmissions. Therefore, we can now determine the delay and loss in the medium access for a collision domain of a multihop, multi-channel WLAN using the Markov model presented in the previous chapter. However, the channel access delay and loss are not uniform across the network. With a lower ring index (of a ring nearer to the AP), the effective packet arrival rate and collision domain size (for the exponential node distribution) increase. Therefore, the delay and loss in channel access are much higher in rings nearer to the AP than the outer rings. We denote the delay and loss in channel access for a node in ring i by $d_c(\nu_i, \lambda_i)$ and $e_c(\nu_i, \lambda_i)$, respectively, which can be determined in a similar way it is done in a single channel, single hop WLAN with ν_i nodes and λ_i packet arrival rate using (3.62) and (3.60) (Section 3.8.5), respectively, where ν_i and λ_i are defined in (4.4) and (4.5), respectively. We use ν_i and λ_i to model the queue of a node in ring i in the following section.

4.4 Modeling the Queue of a Multi-interface Node

Although the Markov model presented in Section 3.8 can be utilized in determining the delay and loss in channel access considering the collision domain size and packet arrival rate at each hop of WLAN, the queue must be remodeled due to the availability of multiple interfaces. In a multi-NI node, each NI picks the next available packet in the queue and attempts to transmit it. Thus, each radio interface (NI) acts like a server for the queue. To utilize a better conditioned channel more and provide all VoIP applications a fair share of the available resources, we model the queue of such a node as a multi-server shared queuing system. Each VoIP application of the node generates voice packets and inserts them into the queue. Additionally, whenever an NI receives a packet that needs to be forwarded, it inserts the packet into the shared queue. A schematic diagram of the proposed queuing system is illustrated in Fig. 4.2 where client node A has two network interfaces *NI 1* and *NI 2*. Whenever a network interface becomes free, either by successfully transmitting the current packet or by

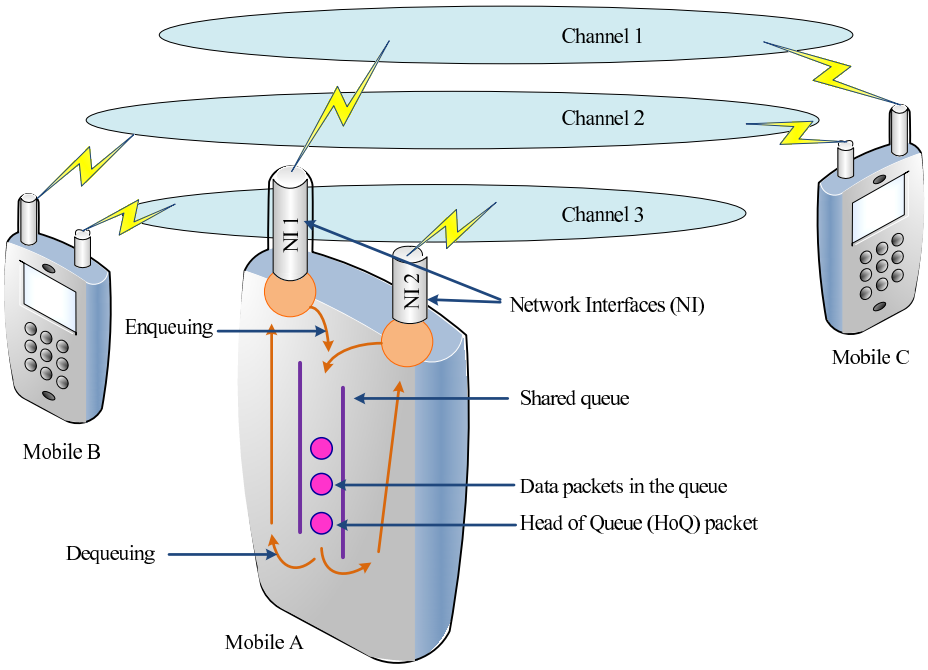


Figure 4.2: A multi-interface node (A) in a multi-channel WLAN. Node A has two NIs (NI 1 and NI 2) which are connected to two different channels (channels 1 and 3, respectively) and act like servers to the shared queue.

dropping it, the NI picks the next available packet in the queue. The two network interfaces *NI 1* and *NI 2* of node A are connected to channels 1 and 3, respectively, and can operate independently. We model the queue as an $M/M/n_t/s_q$ queuing system with n_t servers and a maximum queue size of s_q. The service rate of each NI is $\frac{1}{d_c}$ (we use d_c and λ interchangeably with $d_c\,(\nu_i, \lambda_i)$ and λ_i, respectively, in this chapter) which gives the maximum queue utilization ratio ρ as

$$\rho = \frac{\lambda}{n_t \frac{1}{d_c}} = \frac{2n_v d_c}{n_a d_f n_t}. \tag{4.6}$$

The probability of exactly i packets being in the queue is denoted by P_i. When $i < n_t$, only i NIs will be accessing the channels but when $i \geqslant n_t$, all n_t NIs will be actively accessing the channels. Assuming that a steady state exists, we define P_i as

$$P_i = \begin{cases} \frac{P_0(\lambda d_c)^i}{i!} & \text{for } 1 \leqslant i < n_t, \\ \frac{P_0(\lambda d_c)^i}{n_t{}^{(i-n_t)} n_t!} & \text{for } n_t \leqslant i \leqslant s_q. \end{cases} \tag{4.7}$$

The sum of the probabilities $P_0, P_i, \ldots, P_{s_q}$ is equal to 1 which gives the empty

queue probability P_0, i.e., the probability of having zero packets in the queue, as

$$P_0 = \begin{cases} \sum_{i=0}^{n_t-1} \frac{(n_t\rho)^i}{i!} + \frac{(n_t\rho)^{n_t}(1-\rho)^{s_q-n_t+1}}{n_t!(1-\rho)} & \text{for } \rho \neq 1, \\ \sum_{i=0}^{n_t-1} \frac{(n_t\rho)^i}{i!} + \frac{(n_t\rho)^{n_t}}{n_t!}(s_q - n_t + 1) & \text{for } \rho = 1. \end{cases} \tag{4.8}$$

In the following, we determine the delay $d_q(\nu_i, \lambda_i)$ and loss $e_q(\nu_i, \lambda_i)$ in the queue of a multi-NI WLAN node with a collision domain size of ν_i nodes and an effective packet arrival rate of λ_i using standard queuing analysis for multi-server queues. For brevity, we use d_q and e_q interchangeably with $d_q(\nu_i, \lambda_i)$ and $e_q(\nu_i, \lambda_i)$, respectively, in this chapter.

4.4.1 Queuing Loss at Each Hop

The queue becomes full when exactly s_q packets are in the queue, and the probability of its occurrence is denoted by P_{s_q}. All the packets which arrive at an already full queue are dropped and lost. Therefore, the queuing loss e_q is given by

$$e_q(\nu_i, \lambda_i) = P\{\text{Queue is full}\} = P_{s_q} = \frac{n_t^{n_t}}{n_t!}\rho^{s_q}. \tag{4.9}$$

4.4.2 Queuing Delay at Each Hop

With an effective packet arrival rate of λ, packets are dropped from the queue at a rate of λe_q. Therefore, the effective packet arrival rate to the queue is $\lambda(1 - e_q)$. Little's theorem dictates that the average number of customers over time in a stable system is equal to the average arrival rate multiplied by the average time a customer spends in the system. It can also be shown from Little's theorem that the average number of customers over time waiting in a stable system is equal to the average arrival rate multiplied by the average waiting time. We recall that the queuing delay d_q is defined as the delay from the time a packet is enqueued to the time it is fetched by the MAC layer of a network interface. When the number of packets in the queue is less than or equal to n_t, each packet is being served by some NI. A packet faces queuing delay only when the number of packets in the queue is greater than the number of network interfaces and the packet has to wait until an interface becomes free. Therefore, the average number of packets in the waiting state is given by $\sum_{i=n_t+1}^{s_q}(i - n_t)P_i$, and applying Little's theorem we find

$$d_q\lambda(1 - e_q) = \sum_{i=n_t+1}^{s_q}(i - n_t)P_i. \tag{4.10}$$

Using (4.7), (4.8) and (4.9) in (4.10) and after further simplification, the queuing delay can be defined as

$$d_q\left(\nu_i,\lambda_i\right)=\begin{cases}\dfrac{P_0 n_t^{n_t}\rho^{n_t+1}\begin{pmatrix}1-\left(s_q-n_t+1\right)^{s_q-n_t}\\ +\left(s_q-n_t\right)^{s_q-n_t+1}\end{pmatrix}}{\lambda(1-\rho)^2\left(n_t!-n_t^{n_t}\rho^{s_q}\right)} & \text{for } \rho\neq 1,\\ \dfrac{P_0 n_t^{n_t}(s_q-n_t)(s_q-n_t+1)}{2\lambda\left(n_t!-n_t^{n_t}\rho^{s_q}\right)} & \text{for } \rho=1.\end{cases} \tag{4.11}$$

4.5 End-to-end Delay and Loss

Voice quality is affected by the delay and loss in the end-to-end path of the voice packet where each hop contributes its share to the final delay and loss. Testing the impairments to the voice quality caused by the end-to-end delay and loss for a packet generated in the ϕ-th ring (the outer most ring) suffices the same for the whole network, since the delay and loss for packets generated in any other ring are smaller. Let the cumulative delay and loss faced by a packet generated by a node in the ring ϕ as the packet hops through the rings $\phi,\phi-1,\ldots,i$ (since a packet is forwarded by single hops, i.e., one ring at a time) be denoted with Δ_i and Γ_i, respectively. The delay experienced by a packet in medium access and queue at each hop is summed to give the total cumulative delay Δ_i when the packet from the outermost ring traverses ring i and is given by

$$\Delta_i=\sum_{j=i}^{\phi}\left(d_c\left(\nu_j,\lambda_j\right)+d_q\left(\nu_j,\lambda_j\right)\right). \tag{4.12}$$

Packet loss in the queue and medium access at each hop is only applicable to the fraction of packets which have not been dropped yet and are multiplicative over successive hops. The total packet loss suffered by the packets generated in ring ϕ as they traverse up to and including ring i is given by

$$\Gamma_i=\begin{cases}e_q\left(\nu_\phi,\lambda_\phi\right)+\left(1-e_q\left(\nu_\phi,\lambda_\phi\right)\right)e_c\left(\nu_\phi,\lambda_\phi\right) & \text{for } i=\phi,\\ \Gamma_{i+1}+\left(1-\Gamma_{i+1}\right)e_q\left(\nu_i,\lambda_i\right)+e_c\left(\nu_i,\lambda_i\right)\\ \quad-\left(\Gamma_{i+1}+\left(1-\Gamma_{i+1}\right)e_q\left(\nu_i,\lambda_i\right)\right)\ e_c\left(\nu_i,\lambda_i\right) & \text{otherwise.}\end{cases} \tag{4.13}$$

Here, Δ_i and Γ_i monotonically increase with a decrease in i. Therefore, the end-to-end delay and loss as experienced by a VoIP packet generated in the outermost ring become maximum while the packet is traversing the ring 1 (the innermost ring) which are denoted by Δ_1 and Γ_1, respectively, and can be determined using $i=1$ in (4.12) and (4.13), respectively.

4.6 VoIP Call Capacity Model for Multi-channel, Multihop WLANs with Multi-interface Nodes

As mentioned before, the effective packet arrival rate is much higher for nodes in the lower indexed rings. The gradually increasing traffic load (as ring index i decreases) forms a critical zone in the immediate vicinity of the AP which limits the call capacity of the entire network. When the effective packet arrival rate of any node nearer to the AP becomes greater than its maximum service rate, its queue starts to grow and the queuing loss increases substantially. This problem is most severe in the ring 1 (with the lowest i) and to avoid this scenario, we must have $n_t n_a d_f \Lambda_1 \Omega_1 \geqslant 2 n_s n_v d_c(\nu_1, \lambda_1)$ which gives the upper limit of the channel access delay in ring 1 as

$$d_c(\nu_1, \lambda_1) \leqslant \frac{n_a d_f n_t \Lambda_1 \Omega_1}{2 n_s n_v}. \tag{4.14}$$

Since the number of the actual VoIP calls is given by $n_s n_v$, using the analysis presented in Sections 4.3 $\sim$ 4.6 and using (4.1) $\sim$ (4.14) we formulate the voice capacity of a multichannel, multihop WLAN with multi-interface nodes as the following optimization problem

$$\begin{aligned} Max \quad & n_s n_v \\ s.t. \quad & I_d(d_l + n_a d_f + \Delta_1 + d_j) + I_{e\_eff}(\Gamma_1 + (1 - \Gamma_1) e_j) \leqslant I_b, \text{ and} \\ & d_c(\nu_1, \lambda_1) \leqslant \tfrac{n_a d_f n_t \Lambda_1 \Omega_1}{2 n_s n_v}. \end{aligned} \tag{4.15}$$

The above formulation maximizes the number of voice calls while maintaining the delay and loss impairments within their total budget so that voice quality does not degrade. The concept of an impairment budget for the voice quality using the ITU-T E-model was introduced in Chapter 3. Since the packets generated by the nodes in the outermost ring face the maximum end-to-end delay and loss, the voice quality is the worst for the voice packets generated in ring ϕ. Therefore, the above optimization problem ensures that the lowest voice quality of any ongoing call, despite the location of the node initiating it, remains above an acceptable level.

4.7 Results and Analyses

In the last chapter, we verified our proposed Markov model and voice call capacity model using simulations in NS2. The simulation results followed our proposed Markov model for DCF and the analytical call capacity model closely. In the current chapter, we use similar logical arguments and analytical methods to formulate the voice call capacity of a multichannel, multihop WLAN. Moreover, to determine the delay and loss in the medium access, the same Markov model is used while standard queuing

analysis is employed to model the performance of the queue . Therefore, similar agreement with simulations can be expected for the call capacity model presented in this chapter. In the following, we discuss some analytical results depicting the impacts of multiple channels, network interfaces and multihop routes on the VoIP capacity.

4.7.1 Effect of Multiple Channels

The benefit of utilizing multiple channels is elaborated in Figs. 4.3, 4.4, 4.5 and 4.6. To investigate the potential performance gain from employing multiple channels, we use a single hop WLAN with 50 client nodes which are uniformly distributed. Each node uses an 802.11b interface to transmit data packets at 11 Mbps and has a queue size of 10 packets. Each node initiates a voice call using the G.729 codec and an aggregation level of 1 to reflect a scenario with a high traffic load and substantial queuing loss.

Figure 4.3 illustrates the change in network conditions in terms of the transmission probability and transmission failure probability when the number of channels is increased. With an increasing number of channels, a collision domain is split into smaller domains and fewer nodes now compete in each channel. Therefore, the number of collisions decreases giving a smaller failure probability p_f. A lower failure probability and collision frequency indicate more successful transmissions as the probability of a successful transmission (not shown in figure) increases with a decreasing collision probability. As a result, the nodes operate mostly in the early retry stages (with smaller contention window sizes) and spend less time in the idle state leading to an increase in the transmission probability. For the above WLAN, the transmission probability increases more than three folds and the transmission failure probability decreases to one third when the number of channels is increased from 1 to 10.

A lower transmission failure probability and a higher transmission probability indicate that the packets are now (with a higher number number of channels) served at a higher rate by the MAC layer. Therefore, the packets wait in the queue for a shorter period and, as a result, the probability of the queue being non-empty decreases. Figure 4.4 illustrates the probability of the queue being non-empty and packet loss in the queue and channel access. Due to the high traffic arrival rate of this WLAN, the non-empty queue probability p_q decreases slowly until $n_c = 4$ and then decreases steadily as further channels are added. Although the decrease in p_q is minimal for $n_c \leqslant 4$ (where $p_q \rightarrow 1$), the packet loss in the queue decreases quickly due to the higher service rate for an increase in n_c. This is one major advantage of utilizing multiple channels as the delay and loss in the queue are identified to be major determining factors of the voice quality [23, 25, 26]. The queuing loss e_q decreases sharply and almost linearly

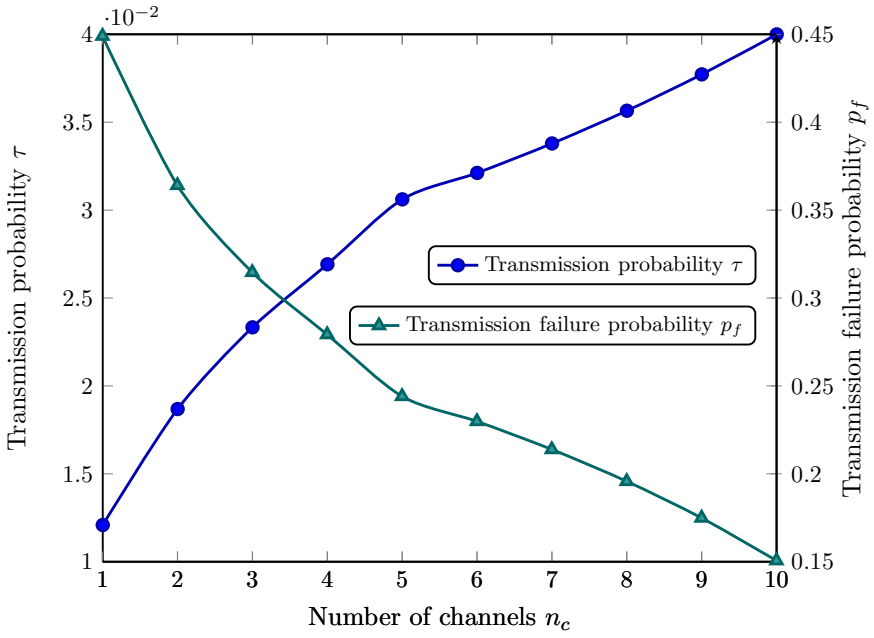


Figure 4.3: The effect of multiple channels on transmission probability and transmission failure probability in a 50-node WLAN with 11 Mbps data rate using the G.729 codec and $n_a = 1$.

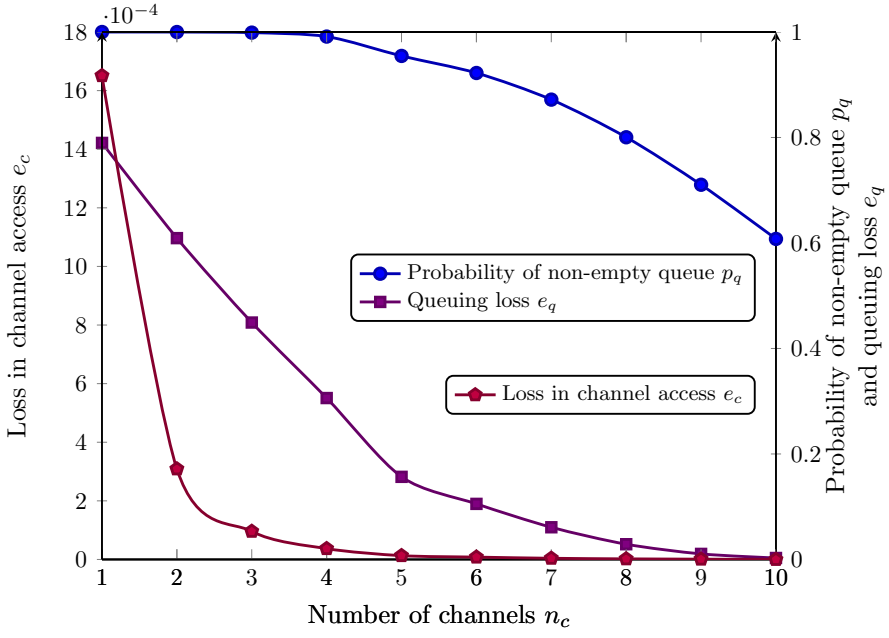


Figure 4.4: The effect of multiple channels on the non-empty queue probability and packet loss in the queue and channel access in a 50-node WLAN with 11 Mbps data rate using the G.729 codec and $n_a = 1$.

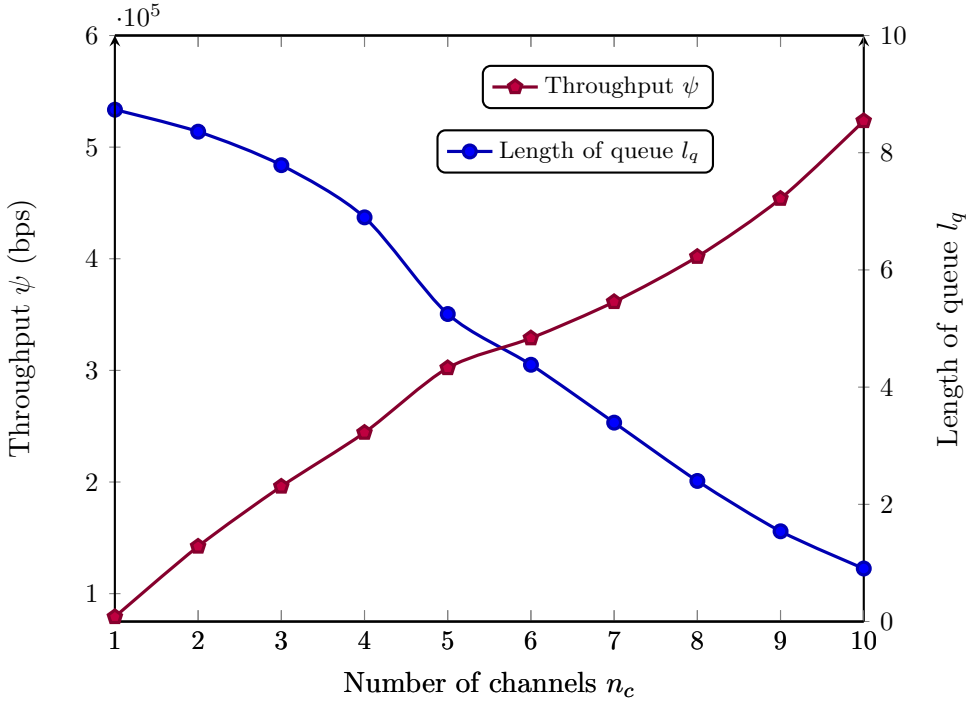


Figure 4.5: The effect of multiple channels on the queue length and throughput in a 50-node WLAN with 11 Mbps data rate using the G.729 codec and $n_a = 1$.

as additional channels are added until $n_c = 5$ where e_q becomes 0.16 compared to $e_q = 0.79$ for $n_c = 1$. Beyond this point, increasing the number of channels leads to a non-empty queue probability of much less than 1 and the queuing loss decreases at a slower rate. The channel access loss, however, decreases sharply and almost linearly when the second channel is added. The channel access loss e_c is 0.0016 and 0.0003 for $n_c = 1$ and $n_c = 2$, respectively, which emphasizes the benefit of splitting a collision domain by adding more channels. For $n_c \geqslant 5$, the channel access loss remains minimal (close to zero) which is due to the lower frequency of collisions and lower transmission failure rate.

Figure 4.5 shows the average queue length and MAC layer throughput for a node in the above mentioned WLAN. The average queue length is a good measure of the required queue buffer size in order to minimize packet loss. With an increasing number of channels, the queue occupancy decreases as shown in the figure. Therefore, the required buffer size also decreases. In particular, only one packet is expected in the queue in the long term when $n_c = 10$. The expected throughput is also shown in Fig. 4.5 which gradually increases with the increasing number of channels. However, as mentioned in Chapter 2, the throughput is an inadequate measure to estimate voice quality. Therefore, we investigate expected voice quality using the E-model next.

Figure 4.6 presents the delay in queue and channel access and the resulting voice quality for the network configuration we considered. As a lower number of contenders compete in each channel when the number of channels is increased, the probability of a collision decreases, as discussed earlier. As a result, the nodes mostly operate in the early retry stages and seldom reach the final retry stage, and this behavior reduces the packet loss in channel access. Additionally, the delay in channel access also becomes lower, as reflected in Fig. 4.6, since the expected delay in a retry stage decreases exponentially with a decrease in retry stage (as a delay of $2^{i-1}Wt_\sigma$ is incurred in retry stage i). This lower delay in the channel access offers a higher service rate of the queue which results in a lower queuing delay. Figure 4.6 shows that the magnitude of decrease in the queuing delay is much higher than the corresponding decrease in the channel access delay. The cumulative effect of a lower delay and loss in both the queue and medium access lowers the total impairment to the voice quality and, as a result, the voice capacity increases with an increasing number of channels. In the 50-node WLAN we considered here, the voice quality rating R is 62.53 and 85.01 for $n_c = 1$ and $n_c = 10$, respectively. Since we need a minimum R of 70.07 and 80.16 to maintain medium and high voice quality (please refer to (3.15)), we find that adding a second channel allows all 50 nodes to initiate a medium quality call each since $R = 77.79$ at $n_c = 2$. Similarly, every node can maintain a high quality call when the third channel is added since $R = 81.42$ at $n_c = 3$.

Figure 4.7 shows the call capacity of a multichannel WLAN with 1, 2 and 3 channels in two different scenarios. An *unlimited* radio band scenario is presented where each channel has a regular radio spectrum and can transmit data at 11 Mbps (which is used in all other figures). But as some countries and local authorities restrict the use of specific frequency bands, we also consider a resource *limited* scenario where an existing 11 Mbps frequency band is split into 1, 2 and 3 sub-channels. As a result, the data transmission rate also decreases proportionately in these sub-channels [153]. Although adding a resource-unlimited channel gives a greater increase in voice capacity, the capacity increase in the limited case is still significant considering that the spectrum resource usage is the same for $n_c = 1 \sim 3$. (The additional increase in capacity compared to the previous chapter is attributed to a larger queue size. A queue size of 100 packets is used in this chapter except where specified otherwise). The call capacity is presented for different aggregation levels, and the maximum capacity is found at $n_a = 10$ for all the above cases. For the G.729 codec, the maximum call capacity is found to be 130 and 195 in the unlimited case and 110 and 141 in the limited case with an n_c of 2 and 3 $n_c = 3$, respectively.

As the capture threshold z decreases, the received power ratio requirement $\left(10^{\frac{z}{10}}\right)$ for a successful capture decreases. Therefore, more collisions satisfy the power re-

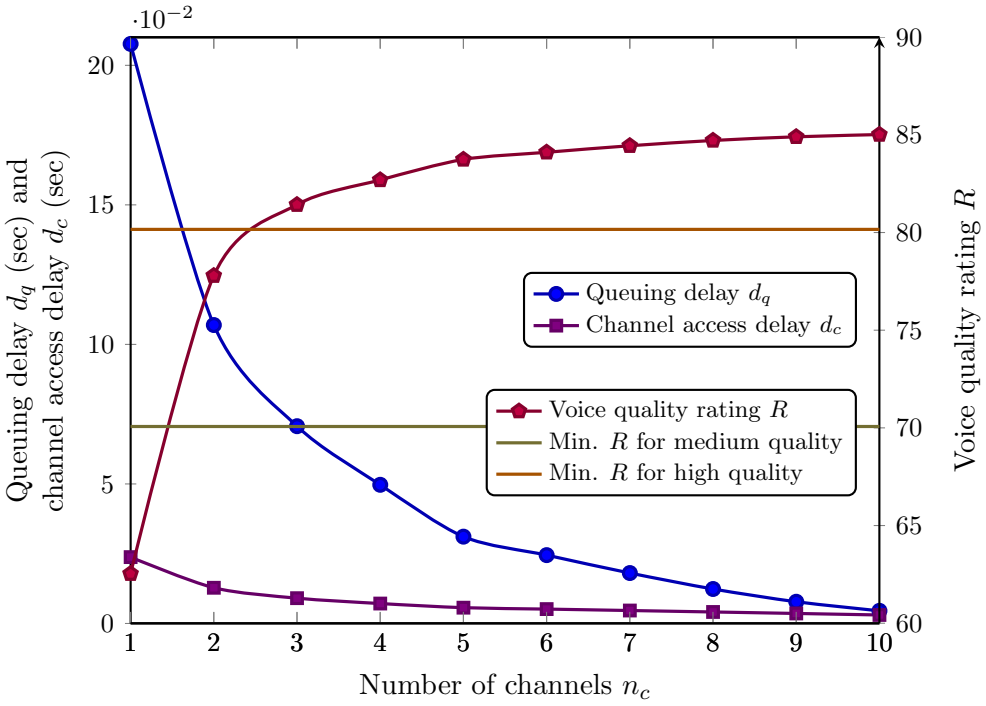


Figure 4.6: The effect of multiple channels on the delay in queue and channel access and voice quality in a 50-node WLAN with 11 Mbps data rate using the G.729 codec and $n_a = 1$.

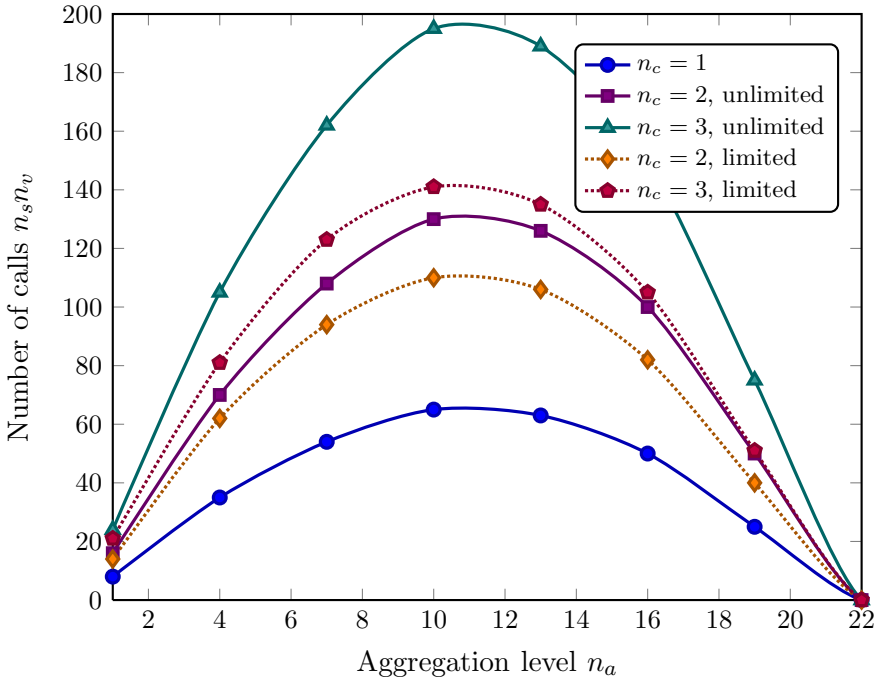


Figure 4.7: The effect of multiple channels and spectrum availability on the voice call capacity using the G.729 codec at different aggregation levels.

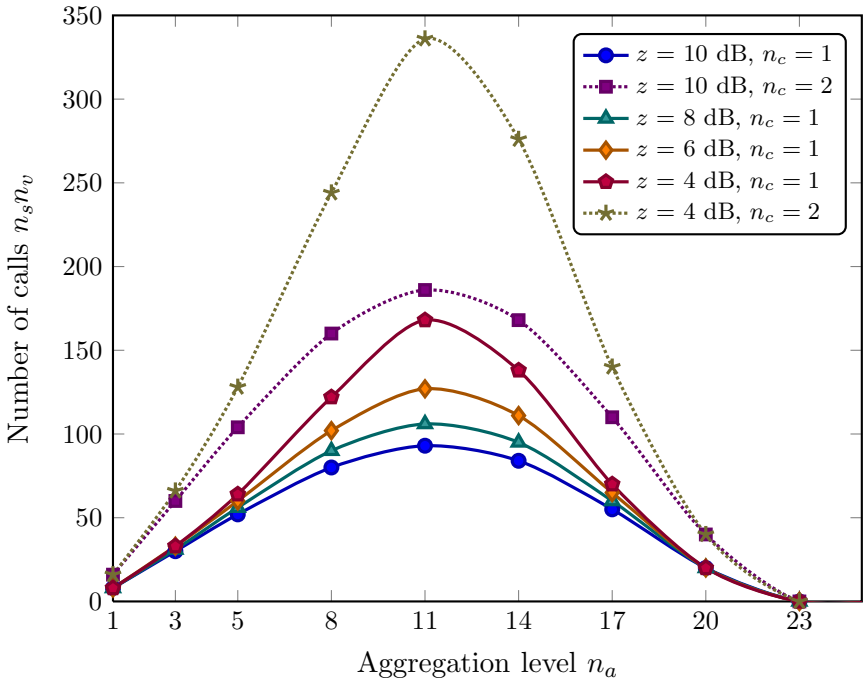


Figure 4.8: The effect of the capture threshold and multiple channels on voice capacity in WLAN with 11 Mbps using the G.729 codec at different aggregation levels.

quirement since the probability that the power ratio of the highest received power to the sum of all other received powers is higher than $\left(10^{\frac{z}{10}}\right)$ increases. As a result, a higher number of collided packets can now be recovered using power capture and the probability of transmission failure decreases. This factor boosts the network performance and the throughput increases while the channel access delay decreases. As a result, more calls can be supported with a lower capture threshold. The increase in call capacity by decreasing the power capture threshold is shown in Fig. 4.8. A call capacity increase by 75 calls is found when the capture threshold is decreased from 10 dB to 4 dB in a single channel WLAN. The benefit of power capture is even more prominent when the number of channels is increased ($z = 4$ dB vs. $z = 10$ dB for $n_c = 2$) and a higher capacity increase can be attained by increasing the number of channels rather than by decreasing the capture threshold.

4.7.2 Effect of Multiple Network Interfaces

The impact of multiple network interfaces on voice capacity in a multi-channel WLAN is presented in Fig. 4.9. We use the G.729 codec with different aggregation levels to determine the call capacity with different packet generation rates (by modifying n_a). To reflect the impact of multiple network interfaces, we use n_t=1, 2, 3 and 4 where

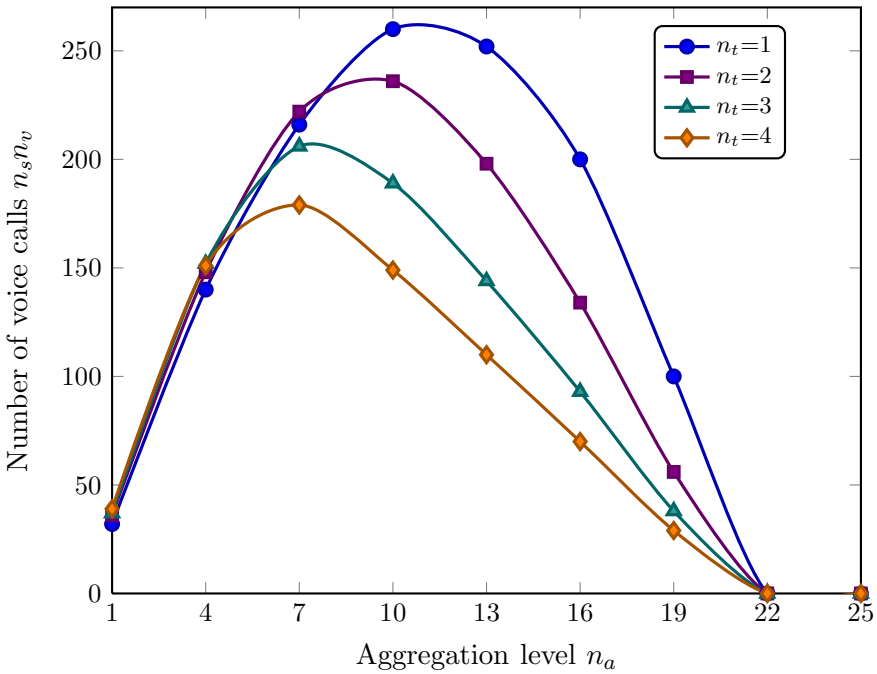


Figure 4.9: The effect of the number of network interfaces on call capacity in a multi-channel WLAN using the G.729 codec at different aggregation levels in a WLAN with 11 Mbps data rate.

$n_c = 4$. The use of multiple NIs has two opposing effects on the voice performance depending on the offered traffic load. With a low aggregation level ($n_a = 1 \sim 7$), the packet generation rate is high and the nodes in every channel operate under the saturated or near saturated conditions. Although adding additional NIs increases the number of contenders in each channel resulting in a higher channel access delay and loss, it also increases the total service rate of the queue since each NI acts as a server to the queue. As a result, the delay and loss in the queue decrease and a slightly higher call capacity is attained. But when the aggregation level is high, the packet generation rate is low; and adding more NIs only incurs a higher delay and loss in the channel access due to the higher number of contenders. The increase in the service rate of the queue is too small in this case to gain any benefit, and the voice capacity decreases. In the 4-channel WLAN mentioned above, the call capacity with $n_a = 4$ for n_t=1, 2, 3 and 4 are 140, 148, 152 and 151, respectively. At $n_a = 10$ the call capacity with the same number of NIs become 260, 236, 189 and 149, respectively. Note that even at $n_a = 4$, adding the fourth NI results in a slightly decreased capacity (compared to $n_t = 3$) due to the high number of contenders in each collision domain. These results suggest that multiple NIs should be used carefully and only to avoid a saturated traffic condition with high queuing loss or frequent network outages in some of the channels.

4.7.3 Voice Performance in Multihop WLAN

In the following, we investigate the performance of voice traffic in multihop WLANs. In addition to multiple hops in the end-to-end path, we also consider the impacts of multiple channels, aggregation level, interference range and node distribution.

4.7.3.1 Effect of Multiple Channels

For a fixed size network, increasing the maximum number of hops ϕ has a varied impact on call capacity as shown in Fig. 4.10, mainly due to two opposing effects explained as follows. As the number of rings increases, the packets generated by nodes at the outermost ring have to pass through a higher number of hops where at each hop the delay and loss in medium access and queue accumulate. As a result, the voice packets generated at the outer rings suffer from a higher end-to-end delay and loss compared to the inner rings. However, since the nodes are now spread dispersedly over a wider area (due to the higher ϕ), the contenders in each channel are also split into several smaller localized collision domains and higher spatial reuse occurs which results in much lower delay and loss in the medium access and queue at each hop. Due to these two opposing effects and their variations with the maximum number of hops, the change in call capacity can be distinguished in two regions. When the aggregation level is low ($n_a = 1 \sim 10$), the packet arrival rate is high, and the accumulated delay and loss over multiple hops result in poor voice quality (i.e., a lower R score). Therefore, the call capacity decreases with an increase in ϕ, and networks with a lower number of hops perform better. However, with a higher aggregation level (e.g., for $n_a \geqslant 11$), the delay and loss at each hop are minimal due to the low traffic arrival rate. Moreover, due to spatial reuse, the same channel frequency is simultaneously used by distant nodes in much smaller collision domains (each containing fewer nodes) leading to further decrease in the delay and loss at each hop. As the end-to-end delay and loss are decreased, the voice quality improves (i.e., a higher R score is attained) and more calls can be supported with a higher number of hops. Apart from attaining a greater coverage, these results highlight the increase in the voice capacity in a multihop network achieved through smaller collision domains and greater spatial reuse.

Figure 4.11 presents the maximum end-to-end delay Δ_1 and loss Γ_1 for a 20-node network using the G.729 codec with $n_a = 2$ for $n_c = 1$ and 2. For $n_c = 1$, Γ_1 quickly increases with an increasing number of hops since the packet loss in the queue and channel access over successive hops are multiplicative. The rate of increase in Δ_1 is high for $\phi \geqslant 5$ and the voice quality quickly degrades for a higher number of hops resulting in a quick decrease of the voice capacity. We find a similar change for $n_c = 2$

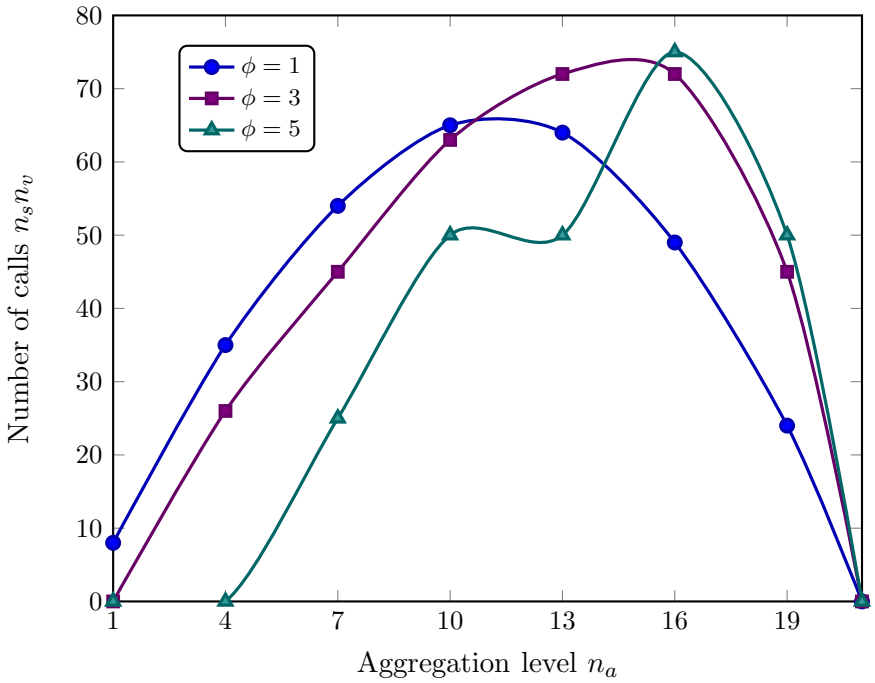


Figure 4.10: The effect of the maximum number of hops ϕ on the voice capacity using the G.729 codec and different aggregation levels in an 802.11b WLAN with 11 Mbps data rate.

but both Δ_1 and Γ_1 are much lower for $\phi < 5$ with $n_c = 2$ which emphasizes the use of multiple channels in the multihop networks.

4.7.3.2 Effect of Interference Range

Figure 4.12 presents the effect of the interference range r_i and the maximum number of hops ϕ on the call capacity where the ring width is the same as the transmission range, i.e., $\omega = r_t = 10$ m. As the interference range increases, the size of the collision domains increases and more stations compete in each collision domain resulting in less spatial reuse and, as a result, the call capacity decreases. On the other hand, as the number of hops increases, a greater spatial reuse takes place but the delay and loss accumulate over multiple hops in the end-to-end path to give a high end-to-end delay and loss. As a result, we find that the call capacity continuously decreases with increasing r_i. At ϕ=4 and 5, the call capacity remains almost the same. For $\phi > 5$, Δ_1 and Γ_1 are very high (also in Fig. 4.11) and no call can be supported for $\phi > 7$.

The cumulative effect of multiple hops and multiple channels on the voice quality is illustrated in Fig. 4.13 for $\phi = 1 \sim 10$ and $n_c = 1 \sim 4$. We use the G.729 codec with $n_a = 2$ in a 70-node WLAN (dense network with high traffic load) where the data

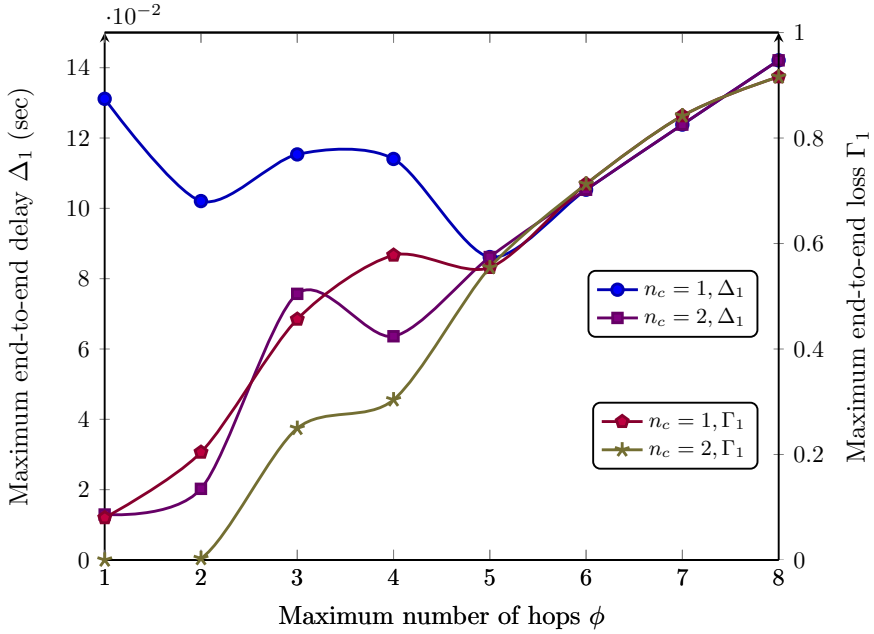


Figure 4.11: The effect of the maximum number of hops ϕ on the end-to-end network delay Δ_1 and loss Γ_1 using the G.729 codec and $n_a = 2$ in a 20-node 802.11b WLAN with 11 Mbps data rate.

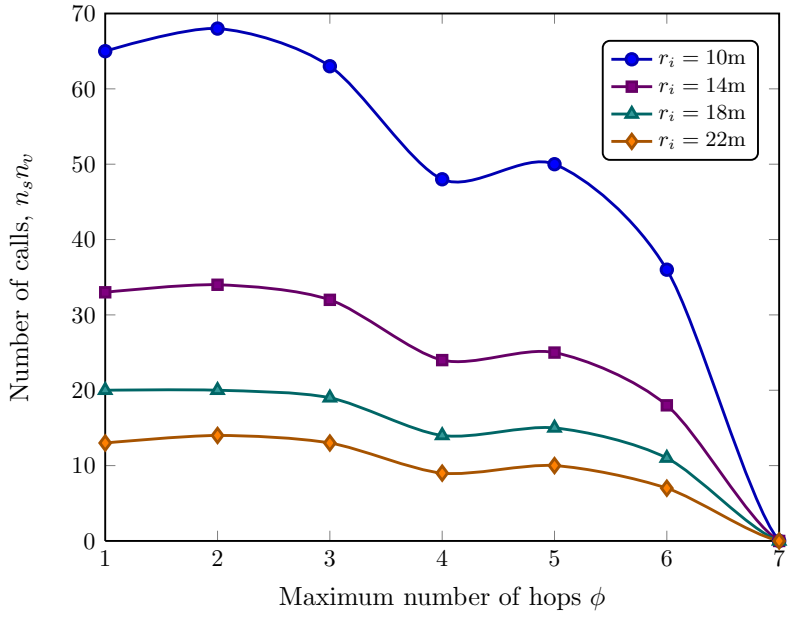


Figure 4.12: The effect of the interference range r_i on the voice capacity using the G.729 codec and $n_a = 10$ in an 802.11b WLAN with 11 Mbps data rate, $r_t = 10$ m and $\omega = 10$ m.

packets are transmitted at 11 Mbps. With an increasing number of hops, the WLAN is split into multiple localized collision domains where the transmissions received by a node do not collide with the transmissions by other nodes located in a different collision domain. Each collision domain now contains fewer nodes and the number of collisions decreases yielding a lower delay and loss in channel access. A lower delay in channel access also reduces the delay and loss in the queue. Although the delay and loss are accumulated over a higher number of hops, the voice quality rating R increases as the delay and loss in channel access and queue at each hop are smaller. However, if the number of hops become too high, the end-to-end delay and loss become high giving a lower R despite the lower delay and loss at each hop. In the single channel WLAN of Fig. 4.13, R increases from 36.05 at $\phi = 1$ to 67.97 at $\phi = 6$ and then starts to decrease with increasing ϕ. Incorporating multiple channels also splits the collision domain at each hop as the NIs are divided into different spectrum bands and, thereby, increases the voice quality considerably. However, after splitting a collision domain few times into multiple small collision domains, the number of contenders in each collision domain becomes minimal and, due to the non-linear relationships between the delay or loss in channel access and the collision domain size, no further benefit can be achieved by splitting it. In Fig. 4.13 we see that for $\phi \geqslant 9$ the voice quality remains almost the same for both $n_c = 1$, 2, 3 and 4. Similarly, for $\phi \geqslant 6$ and $\phi \geqslant 4$, the WLANs with $n_c = 2$, 3 or 4 and the WLANs with $n_c = 3$ or 4, respectively, exhibit the same level of the voice quality. The minimum voice quality requirements for medium and high quality calls are also illustrated in Fig. 4.13 which show that the increase in the voice quality by incorporating additional channels allows each of the 70-nodes to initiate medium and high quality calls for $n_c \geqslant 2$ and $n_c \geqslant 4$ ($n_c > 4$ WLANs are not shown), respectively, for a given range of ϕ. In the 2-channel scenario, medium quality calls are supported when the nodes are spread over $3 \sim 8$ hops, however, for $\phi \geqslant 9$, the end-to-end delay and loss becomes too high and no call can be supported. Similarly, high quality calls are supported for $\phi = 1 \sim 3$ with $n_c = 4$. In the 4-channel scenario, no calls are supported for $\phi \geqslant 4$ due to a high end-to-end delay and loss.

4.7.3.3 Effect of Node Distribution

The effect of exponential and uniform node distributions on the voice quality is illustrated in Fig. 4.14. Here, data packets are transmitted at 11 Mbps and the G.729 codec is used with $n_a = 1$ to reflect a high traffic load. We use ϕ=2, 4 and 6 and $\theta = 0.1 \sim 5.0$ for the exponential node distribution, and determine the collision domain size and effective packet arrival rate at each hop which define the delay and loss in the channel access and queue. The maximum end-to-end-delay and loss are com-

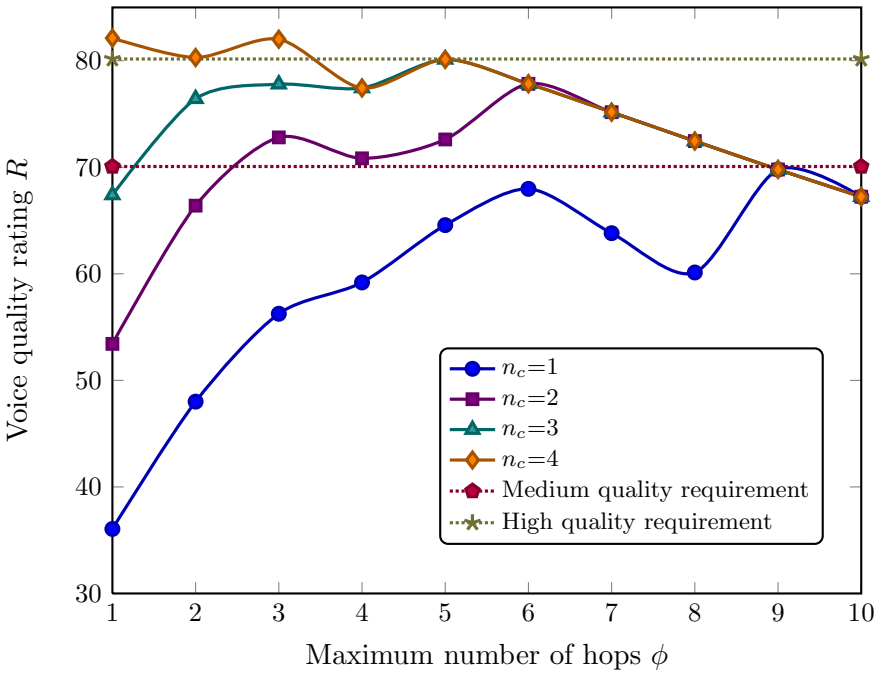


Figure 4.13: The effect of multiple hops and channels on the voice quality in a 70-node WLAN using the G.729 codec and $n_a = 2$.

puted and the minimum voice quality R, which is incurred to voice streams generated by nodes in the outer-most ring, is determined using the ITU-T E-model. Although the density parameter θ is not applicable to the uniform distribution, we use the same number of nodes and the number of hops, as used in an exponential scenario, in an equivalent WLAN with uniform node density and present the voice quality for comparison purposes. The node density of ring 1 is the maximum for $\theta = 1$ which forms a bottleneck for the WLAN and all packets suffer from high delay and loss in ring 1. For $\theta > 1$, with increasing θ, more nodes are moved to the outer rings and the bottleneck becomes less rigid since the collision domain in the 1-st ring becomes smaller. The effective packet arrival rate is highest in the ring 1 and a lower collision domain size can increase the voice quality significantly. While it means that the delay and loss in ring 1 is reduced, it also means that each outer hop will be populated with a higher number of nodes and delay and loss at each outer hop will be higher now compared to $\theta \to 1$. Due to these two opposing effects, the voice quality initially increases with increasing θ for $\theta = 1 \sim 4.5$ and remains almost the same for $\theta > 4.5$. (Although, $\theta < 1$ also improve voice quality, it restricts the total network size to only a few nodes and can be ignored). The voice quality with the exponential distribution remains almost unaffected for changes in ϕ although with the uniform distribution considerable increase in R is observed for any increase in ϕ. The voice quality in a WLAN

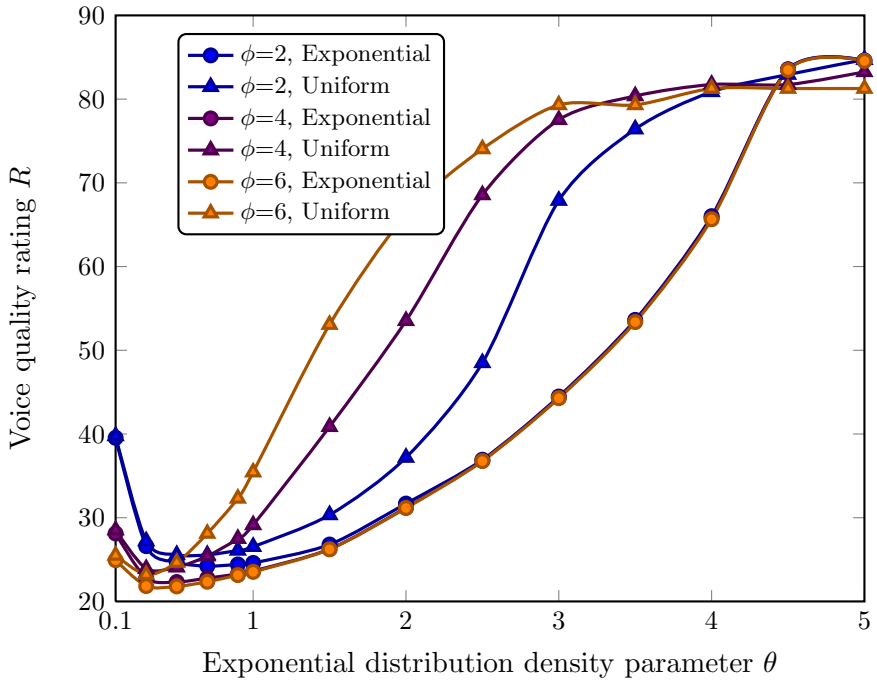


Figure 4.14: The effect of exponential and uniform node distributions on voice quality in multihop WLANs with 11 Mbps data rate using the G.729 codec and $n_a = 1$.

with uniform node distribution and the same number of nodes show that the uniform distribution offers a higher voice quality compared to the corresponding exponential case. However, for $\theta \geqslant 4.5$, the exponential scenario limits the node density of the 1-st ring to only few nodes which shows a higher voice quality than the corresponding uniform distribution scenario. In the 6-hop WLAN, the voice quality rating R with exponential and uniform distributions are 31.16 and 66.20 at $\theta = 2.0$ which become 84.56 and 81.26 at $\theta = 5.0$, respectively.

4.7.3.4 Effect of Maximum Number of Hops

The effect of the maximum number of hops ϕ on voice capacity is shown in Fig. 4.15. We use the G.729 codec with aggregation levels of 4, 6 and 10 and present the call capacity for $\phi = 1 \sim 8$. Data packets are transmitted at 11 Mbps and the results for n_c=1 and 2 are illustrated. When ϕ is increased, the WLAN occupies a larger area and, with the same number of nodes in the WLAN, the client nodes are now located more dispersedly. As a result, each localized collision domain contains fewer nodes and the delay and loss in both medium access and queue decreases in each collision domain. Therefore, the end-to-end delay and loss decrease with an increasing ϕ, and the call capacity initially increases. But when ϕ becomes considerably high, the end-to-end

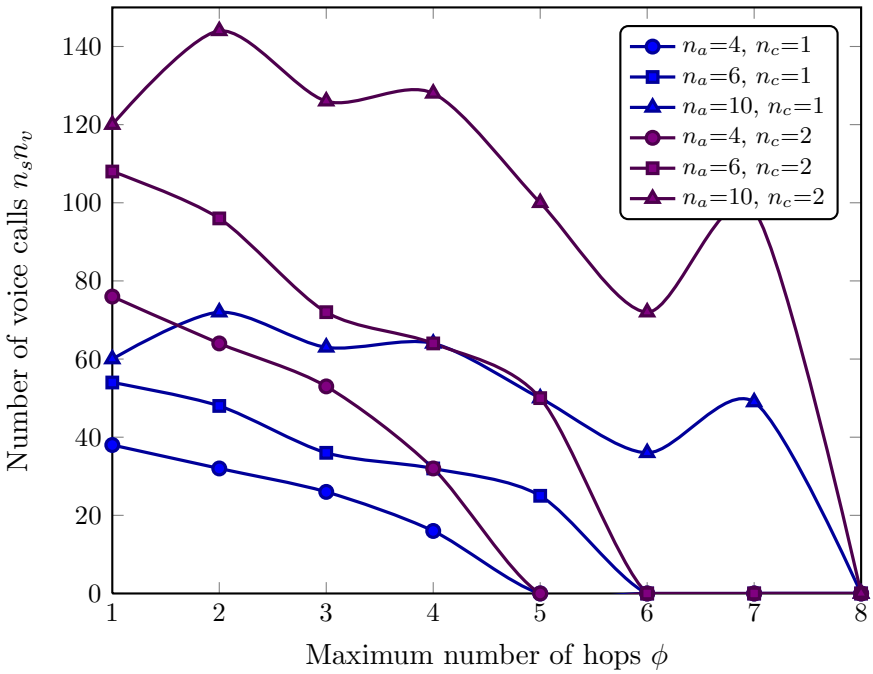


Figure 4.15: The voice capacity versus the maximum number of hops and multiple channels for the G.729 codec at 11 Mbps data rate.

delay and loss start to increase as the delay and loss at each hop are accumulated, and the voice capacity decreases. But considering the wider area $\pi\omega^2\phi^2$ (which increases quadratically with ϕ) covered by one AP, the use of a higher ϕ might be beneficial for public networks. In the results illustrated in Fig. 4.15, the voice capacity increases when ϕ is increased from 1-hop to 2-hop using an aggregation level of 10. In the n_a=4 and 6 cases, the packet arrival rates remain too high to offer a lower end-to-end delay and loss even with the slightly lower collision domain sizes. For $\phi > 2$, the voice capacity decreases slowly and becomes 0 at ϕ=5, 6 and 8 for n_a=4, 6 and 10, respectively. However, utilizing multiple channels in addition to multiple hops can offer a higher voice capacity while maintaining the wide coverage. For instance, with n_a=10, the call capacity in the single channel scenarios are 60 and 72 for ϕ=1 and 2, respectively, which are increased in the 2-channel scenario to 120 and 144, respectively. But despite the increase in call capacity, the maximum coverage remains the same even in the multichannel scenario, i.e., no calls are supported with n_c=1 or 2 for a ϕ higher than 5, 6 and 8 with n_a=4, 6 and 10, respectively.

4.8 Key Contributions

As mentioned in Section 2.5, no existing analytical call capacity model for the 802.11 WLANs in the current literature considered multiple hops, channels or network interfaces. Although a three dimensional modeling *attempt* was mentioned in [83] for multihop WLANs, using the queue size as the third dimension in addition to backoff stage and retry limit, the impacts of multiple channels and network interfaces were ignored and the collision domain size or the effective packet arrival rate which vary at each hop of a multihop WLAN were not incorporated. Most importantly, voice traffic and voice quality guidelines were not considered and the model was incomplete since no estimation of the performance measures was derived. In this regard, the key contributions in this chapter are as follows—

- Most public Wi-Fi networks are multihop in nature which offer wider coverage per access point. Moreover, the IEEE 802.11 standards support the simultaneous use of multiple channels that can increase the voice capacity significantly. In this chapter, we presented a call capacity model for multihop, multi-channel WLANs. We used an impairment budget (introduced in the previous chapter) using the ITU-T E-model guidelines to determine the voice quality requirement and used codec parameters in conjunction with the delay and loss in the medium access and queue to determine the voice capacity of such networks. We used the Markov model presented in the previous chapter which allows us to model the medium access mechanism with considerable accuracy. The collision domain size and the effective packet arrival rate due to packet forwarding over multihop routes were determined for each hop. We formulated the maximum end-to-end delay and loss and determined the minimum voice quality to ensure that the voice quality does not degrade in an attempt to allow any additional call. The maximum voice capacity of such networks under various conditions were demonstrated.

- Most future WLAN devices are expected to bear multiple network interfaces [21], therefore, the effect of multiple NIs on the voice capacity should be carefully considered. But no existing analytical call capacity model considered the effect of NIs. The call capacity model proposed in this chapter considers the effect of multiple NIs in determining the number of contenders in each collision domain. Additionally, we used a multi-server queuing system to model the delay and loss in the queue of a multi-NI node.

- While a uniform node distribution can be expected in planned networks, public WLANs often suffer from bottlenecks formed due to a non-uniform node distribution which results in a higher node density in the closest proximity of the

AP. We used the exponential distribution, which can be replaced with any other non-uniform distribution easily, as a representative node distribution for such a scenario and compared the voice quality degradation to a uniform distribution.

- We demonstrated and analyzed the effect of multiple hops, channels, NIs and interference range on the voice capacity. The effect of the node distribution on voice quality was also compared between exponential and uniform node distributions. A detailed analysis of the changes in the transmission probability, voice quality and delay and loss in both the medium access and queue were carried out for variations in the number of channels, capture effect or maximum number of hops to identify the true performance of a multi-channel, multihop WLAN.

4.9 Summary

In this chapter, we developed an analytical model to estimate the VoIP call capacity of the IEEE 802.11 based multi-channel, multihop networks. The model ensures the call quality by determining the impairments to the voice quality using the ITU-T E-model. Considering the possible client node density (which can be exponential or uniform), the number of channels and the number of hops, we determined the collision domain size and effective packet arrival rate at each hop. The Markov model presented in the previous chapter was used to determine the delay and loss in the channel access for each collision domain. A multi-server queuing system was also used to model a multi-interface node. The delay and loss in the queue and channel access were then used to determine the maximum end-to-end delay and loss which, together with the impairment budget, gave the voice call capacity in the form of an optimization problem. Our analyses showed that the incorporation of multiple channels can increase the voice capacity significantly. Additionally, utilizing a multihop WLAN can also increase the voice capacity to some extent (within a given range of hop sizes). But increasing the number of network interfaces decreases voice performance and can be used to increase voice capacity only when the packet arrival rate is high. We also showed that the bottleneck formation due to non-uniform node distribution (we presented results for exponential distribution) degrade the voice performance significantly which emphasizes the proper planning of WLANs so that a nearly uniform node distribution can be deployed.

The DCF based medium access wastes a considerable amount of channel time in the backoff state and suffers from a high channel access delay. Moreover, the frequent collisions in a dense collision domain incur a high channel access loss. The high channel access delay also results in high delay and loss in the queue. The IEEE 802.11 standards offer a time synchronized medium access mechanism called the Point Coordination

Function (PCF) that can reduce the channel access delay as collisions can not occur and backoff is not needed. But the existing call capacity models ignored the relative channel time usage in uplink (from the client nodes to the AP) and downlink (from the AP to the client nodes) transmissions and provided inadequate analyses with imperfect channels. We present a call capacity estimation model for the PCF based medium access mechanism in the following chapter.

Chapter 5

Voice Capacity of PCF based WLANs

5.1 Introduction

In the previous two chapters (Chapters 3 and 4), we investigated the voice capacity of DCF based single hop and multihop WLANs. Despite the utilization of a carrier sensing with collision avoidance technique in the DCF based medium access mechanism (based on CSMA/CA, please see Section 2.3.2.1), collisions can still occur. A Binary Exponential Backoff algorithm (the backoff mechanisms are outlined in Section 2.3.4) is used in the DCF mechanism which enables the WLAN nodes to adapt to changes in the traffic load and reduce collisions at the cost of a high delay in the channel access. However, collisions still occur giving rise to channel access loss, especially under a high traffic load. The high channel access delay, in turn, incurs high delay and loss in the queue as shown in Chapters 3 and 4. In addition to DCF, the IEEE 802.11 standards also define a time-synchronized medium access mechanism called the Point Coordination Function (PCF) (explained in Section 2.3.2.2) where the client nodes access the channel in a time synchronized manner under the coordination of a Point Coordinator (PC) located within the AP. Since the PC allows only one node at a time to transmit, collisions do not occur when the PCF based medium access is used and, therefore, backoff is not required. The channel remains idle only for a SIFS between two consequent transmissions which is required to identify separate transmissions. As a result, the channel time used to transmit a packet is low in the PCF based medium access which also reduces the delay and loss in the queue. In the absence of collisions, packet loss occurs only due to channel errors which can be kept to a minimum using the multi-rate MAC mechanism [154]. The lower delay and loss in the channel access and queue with PCF than DCF can substantially increase the voice capacity and should be

thoroughly investigated. In this chapter, we model the call capacity of a PCF based WLAN using the voice quality impairment budget introduced in Chapter 3.

5.2 Motivation to Investigate Call Capacity using PCF

The low delay and loss with the PCF based medium access mechanism inspired a number of researchers, and considerable research effort is invested in analyzing the voice performance in PCF based WLANs. However, most of the existing works primarily focused on improving the round robin polling mechanism proposed for PCF in the IEEE 802.11 standards which is beyond the scope of this thesis. Some of the investigations, including [3, 20, 114, 118], also estimated the voice capacity which were reviewed in Section 2.5.5. In determining the call capacity with the PCF based medium access, most of the above works [20, 114, 118] considered an ideal channel only which makes their findings unreliable in real networks since every real channel exhibits noise and interference to some extent. Although [3] considered an imperfect channel, the standard voice quality measures and codec characteristics were ignored. The delay and loss in the queue incur severe voice quality degradation which were also not considered in [3]. The capacity model presented in [3] used only the number of retransmissions due to channel errors ignoring the difference in channel time usage due to channel errors in either the uplink (from a client node to the AP) or the downlink frames (from the AP to a client node); and thereby, provided inadequate call capacity analyses. With packet error rate p_e, the probability of failure and success of a transmission are given by p_e and $(1-p_e)$, respectively. The probability that a packet requires i transmissions to be successfully transmitted is given by $p_e^{i-1}(1-p_e)$. A packet will be dropped if the number of retransmissions (one less than the total number of transmissions) exceeds the retry limit m. For a given node, all the uplink and downlink transmissions that take part in a super-frame are conjunctively termed as a *transaction* in this dissertation for the tractability of analysis as mentioned in Section 2.3.2.2. Transactions with separate client nodes are independent of each other. Using the method in [3], the length of a transaction t'_t is given by

$$t'_t = \sum_{i=0}^{m} 2i\, p_e^i\, (1-p_e)(t_d + t_{sifs}), \tag{5.1}$$

where t_d and t_{sifs} are the time taken to transmit a data packet and the length of a SIFS period, respectively, and 0^0 (for $i = 0$ and $p_e = 0$, i.e., with an ideal channel) is 1. The above expression in (5.1), although simplistic, is correct only if we assume that channel errors occur only in the uplink transmissions, i.e., channel errors never

occur in the downlink transmissions. This assumption is unjustified as channel error can occur in both uplink and downlink transmissions.

When a downlink frame contains channel errors, the frame is discarded by the receiving node. The receiving node, therefore, does not transmit any uplink data frame as it remains unaware that it has been polled. It can not acknowledge the last downlink frame either. As a result, the channel remains idle and the point coordinator continues with its following transmission after a PIFS. The channel time wastage in this case is $(t_d + t_{sifs} + t_{pifs})$, where t_{pifs} is the length of a PIFS period, compared to the $2(t_d + t_{sifs})$ time that is wasted when an uplink frame contains channel error. The maximum amount of error introduced in determining the time to poll a node once using the above mentioned assumption [3] is $(t_d + t_{sifs} - t_{pifs})$. Denoting the expected length of a transaction with t_t, the maximum amount of error in estimation of the expected transaction length is given by

$$\begin{aligned} t'_t - t_t &= \sum_{i=0}^{m} p_e^i (1 - p_e)\, i\, (t_d + t_{sifs} - t_{pifs}) \\ &= \begin{cases} \frac{(t_d + t_{sifs} - t_{pifs})\left(p_e + p_e^{m+1}(mp_e - m - 1)\right)}{1 - p_e} & \text{for } 0 < p_e < 1, \\ 0 & \text{for } p_e = 1. \end{cases} \end{aligned} \tag{5.2}$$

The magnitude of this error is depicted in Fig. 5.1 for different aggregation levels and data rates in an 802.11b WLAN using the G.729 codec. A similar error pattern, with different magnitudes, holds for other codecs and 802.11a/802.11g standards.

As an example, consider a typical VoIP connection using the G.729 codec with an aggregation level of 2 which generates a payload of 20 bytes per packet after each 20 ms interval. The RTP, UDP and IP layers add another 40 bytes of header where the MAC header can be 28 bytes (used in NS2), 34 bytes including optional fields (used in OPNET) or 36 bytes if FEC (Frame Error Correction) is used. In an 802.11b WLAN, the data part of a packet is transmitted at a maximum of 11 Mbps while the PLCP and preamble parts of the data and control packets are transmitted at 1 Mbps. Therefore, the transmission of a data packet takes 256 μs. When there are no channel errors, a transaction contains two such transmissions for the uplink and downlink packets which are separated by two SIFS. Therefore, a single handshake containing a downlink and an uplink frame using an ideal channel takes 532 μs to complete. If the uplink packet contains channel error, the channel is still kept busy for the same amount of time. However, if the downlink frame contains channel error, the channel remains idle in the absence of an uplink transmission; and the PC proceeds with its following transmissions after $256 + 10 = 266$ μs, thereby, reducing the wastage of the channel time. The channel access delay, as explained later in this chapter, is a function of the super-frame length which is again a function of the transaction length.

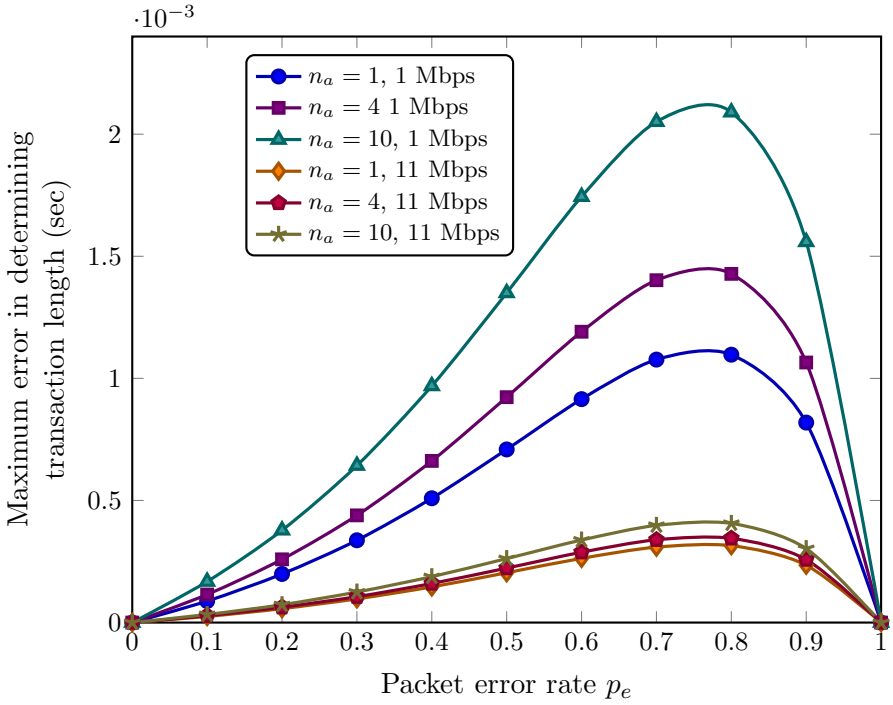


Figure 5.1: The maximum error in the previous work in [3] due to considering the number of retries only and ignoring the difference in channel time wastage due to channel errors in the uplink and downlink transmissions. The G.729 codec is used in an 802.11b WLAN.

Therefore, considering only the number of retries is not sufficient to correctly model the PCF based channel access. In fact, it gives an over-estimation of the channel access delay; and consequently, the voice quality impairment is overestimated leading to an under-estimation of the voice capacity. To correctly model the PCF based channel access mechanism, we introduce a Markov model in the following section.

5.3 Markov Model for the PCF based Channel Access

We model a transaction of the PCF based medium access mechanism using our proposed Markov model which is shown in Fig. 5.2. The oval shapes denote different states of a transaction. The "Begin" and "End" states are the start and end points of a transaction where the "RS i" states (for $i = 1, 2, \ldots, m$ where m is the retry limit) represent the i-th retry state. Being in the i-th state means that the AP is going to poll the client node in this transaction for the i-th time next. With a retry limit m, the AP can retry a failed transmission in the ongoing transaction for a maximum of m times. The arrows represent all possible transitions from a state to other possible states and are labeled with a *"transition probability : transition delay"* pair for easier

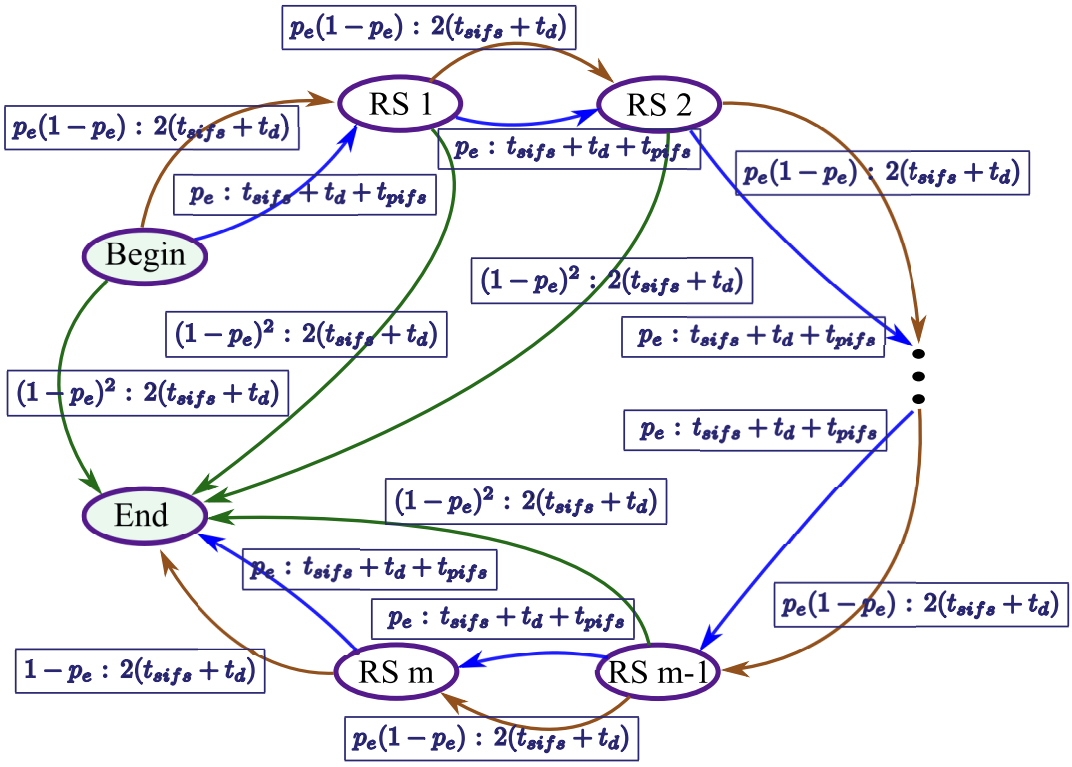


Figure 5.2: Markov model for a transaction in the IEEE 802.11 PCF based medium access mechanism representing the retry states "RS 1", "RS 2",...,"RS m" with oval shapes. The green arrows represent successes in both the downlink and uplink transmissions while blue and brown arrows represent channel error in either the downlink or the uplink frames, respectively. Transitions are marked with a *"transition probability : transition delay"* pair.

analysis. In the following, we first develop a model for two way traffic where each node has some packets to send to the AP and the AP also has some packets to send to every node at any given time. We present appropriate modifications to incorporate the uplink-only and downlink-only traffic conditions in Section 5.5.

In an error-free scenario, the AP transmits a downlink frame to a node after waiting for a SIFS (following the last transmission). The node waits for a SIFS and replies with an uplink frame. Such a transaction can occur with a probability of $(1-p_e)^2$ and takes a total of $2(t_{sifs}+t_d)$ time to complete. These transitions from the retry states to the "End" state are shown with green arrows.

If a downlink frame transmission contains channel error (marked with the blue arrows), the destination node can not recover its content and does not respond. As a result, the channel remains idle and the AP proceeds with the next transmission after a PIFS. Thus, a total of $(t_{sifs}+t_d+t_{pifs})$ time is wasted in this case. However, if the downlink frame is received correctly but the uplink frame contains channel error

(marked with brown arrows), the channel is kept busy during the transmission of the uplink frame and a channel time of $2(t_{sifs} + t_d)$ is wasted. A transmission failure due to channel error in the downlink or uplink frame can take place with probabilities of p_e and $p_e(1 - p_e)$, respectively. Channel error in either the downlink or the uplink frame will require retransmissions of both the frames and the system moves to the next retry state with probability $p_e(2 - p_e)$. Once the AP has retried polling the node in this transaction for m times, the transaction is forced to finish. From the "RS m" state, the transaction finishes either with probability $(1 - p_e)$ requiring $2(t_{sifs} + t_d)$ time or with probability p_e requiring $(t_{sifs} + t_d + t_{pifs})$ time.

A retry state can be reached from its preceding retry state or from the "Begin" state (for retry state "RS 1") with probability $\mathbb{P}_r$. Using the analysis above, $\mathbb{P}_r$ can be defined as

$$\mathbb{P}_r = p_e(2 - p_e). \tag{5.3}$$

A retry state can be reached from its immediate prior state if any channel error occurs in either the uplink or downlink frames as shown using two transitions to each retry state in Fig. 5.2. Between two such transitions to a retry state, the transition (brown arrow) labeled as $p_e\,(1 - p_e) : 2\,(t_{sifs} + t_d)$ corresponds to channel error in an uplink frame; and the other transition (blue arrow) labeled as $p_e : t_{sifs} + t_d + t_{pifs}$ corresponds to channel error in a downlink frame. Considering the probabilistic delay using either of the transitions, the expected delay $\mathbb{D}_r$ in the transition to a retry state from its immediately prior state is given by

$$\mathbb{D}_r = p_e(t_{sifs} + t_d + t_{pifs}) + 2p_e(1 - p_e)(t_{sifs} + t_d). \tag{5.4}$$

A transaction can succeed with probability $\mathbb{P}_s$ from any of the retry states "RS 1", "RS 2", ...,"RS m-1" or the "Begin" state (green arrows). Such a transition requires two consecutive successful transmissions for the downlink and the uplink frames. We define $\mathbb{P}_s$ as

$$\mathbb{P}_s = (1 - p_e)^2. \tag{5.5}$$

The time taken by two such consecutive successful transmissions is denoted with $\mathbb{D}_s$ and is given by

$$\mathbb{D}_s = 2(t_{sifs} + t_d). \tag{5.6}$$

The probability of reaching the "End" state from the state "RS m" is denoted with $\mathbb{P}_e$. With a finite retry limit, a transaction always ends after m retries and

$$\mathbb{P}_e = 1. \tag{5.7}$$

However, depending on whether the transaction ends, i.e., reaches the "End" state from the "RS m" state, with a success or a failure in the downlink transmission, the

time taken in this transition $\mathbb{D}_e$ is defined as

$$\mathbb{D}_e = p_e \left(t_{sifs} + t_d + t_{pifs}\right) + 2\left(1 - p_e\right)\left(t_{sifs} + t_d\right). \tag{5.8}$$

To determine the length of a transaction, we calculate the weighted sum of the delays in every possible path from the "Begin" state to the "End" state where the transitional probability is the weight of an associated transitional delay in our proposed Markov chain. If a transaction requires i retransmissions where $i < m$, the transaction reaches the "End" state finally from the "RS i" state. Such a transaction can take place with probability $\mathbb{P}_r^i\,\mathbb{P}_s$ and requires $(i\,\mathbb{D}_r + \mathbb{D}_s)$ time to finish. If a transaction requires m retransmissions, it reaches the "End" state from the "RS m" state. Such a transaction can occur with probability $\mathbb{P}_r^m\,\mathbb{P}_e$ and takes $(m\,\mathbb{D}_r + \mathbb{D}_e)$ time to finish. Therefore, the total length t_t of a transaction can be shown to be

$$t_t = \sum_{i=0}^{m-1} \mathbb{P}_r^i\,\mathbb{P}_s\left(i\,\mathbb{D}_r + \mathbb{D}_s\right) + \mathbb{P}_r^m\,\mathbb{P}_e\left(m\,\mathbb{D}_r + \mathbb{D}_s\right). \tag{5.9}$$

Using (5.7) in (5.9) and after further simplification, we present the length t_t of a PCF transaction as

$$\underset{p_e \neq 1}{t_t} = \mathbb{P}_s\mathbb{D}_r\frac{\mathbb{P}_r^m\left(m\mathbb{P}_r - m - \mathbb{P}_r\right) + \mathbb{P}_r}{\left(\mathbb{P}_r - 1\right)^2} + \mathbb{P}_s\mathbb{D}_s\frac{\mathbb{P}_r^m - 1}{\mathbb{P}_r - 1} + p_e^{m-1}\left(m\mathbb{D}_r + \mathbb{D}_e\right), \tag{5.10}$$

which is undefined for $p_e = 1$, i.e., when every transmission contains channel errors. Applying L'Hôpital's rule twice and finding limit at $p_e \to 1$, we derive t_t for $p_e = 1$ as

$$\underset{p_e = 1}{t_t} = \mathbb{P}_s\mathbb{D}_r(m^2 + m - 2) + \mathbb{P}_s\mathbb{D}_s m + p_e^{m-1}\left(m\mathbb{D}_r + \mathbb{D}_e\right). \tag{5.11}$$

In the 802.11 PCF mechanism, every node is polled once during a CFP. In a WLAN of n_s nodes, the length of the contention free period is, therefore, $(t_b + n_s t_t + t_{sifs} + t_{ce})$ where t_b and t_{ce} are the transmission times of a Beacon and a CF-End frame, respectively. In a super-frame, a Contention Period (CP) and a Contention Free Period (CFP) alternate, and the Beacon frame is transmitted after the channel remains idle for a PIFS period following the contention period. The length of the contention period is denoted by t_{cp} and is explained in Section 5.7. Therefore, the length of the super-frame t_{sf} is given by

$$t_{sf} = t_b + n_s t_t + t_{sifs} + t_{ce} + t_{cp} + t_{pifs}. \tag{5.12}$$

5.4 Delay and Loss in Medium Access and Queue

In this section, we determine the delay and loss in the PCF based medium access mechanism. The delay and loss in the queue are also discussed.

5.4.1 Channel Access Loss

The MAC layer drops a packet after retransmitting it for $(m+1)$ times. Therefore, the packet loss e_c in channel access incurred to a voice stream is the probability of $(m+1)$ successive failed transmissions and is given by

$$e_c = \mathbb{P}_r^{m+1}. \tag{5.13}$$

5.4.2 Channel Access Delay

Unlike DCF, the uplink and downlink packets experience different channel access delay when PCF is used. A client node transmits one uplink packet to the AP in each super-frame. Therefore, the upper bound of channel access delay is given by the length of the super-frame t_{sf}. The length of the super-frame was used as the channel access delay in [3] due to the assumption that a voice packet becomes available exactly at the beginning of each super-frame (in [3]). However, depending on the distribution of the time when a packet is available at a node, the packet can face different lengths of delay in the medium access. If the packet becomes available just after a node has completed its transaction, it will have to wait for another full length of a super-frame, i.e., until the next time the node is polled. On the other hand, if the packet becomes available just before the node is polled, the delay will be close to zero. CBR applications like VoIP generate the voice packets after regular intervals. For such applications, the packet availability (or generation) time is a function of the packetization delay and the distribution of the application starting time. For VoIP applications, the packetization delay can be derived from the aggregation level and voice frame length but the application starting time distribution can be very specific to the voice call usage pattern of a WLAN. Such a distribution can be estimated using statistical methods. Without such information, it is logical to assume a uniform distribution of packet availability time as used in [155], which gives an expected channel access delay d_c of $\frac{t_{sf}}{2}$ for the uplink packets. However, the AP transmits n_s downlink frames in each super-frame, i.e., the queue of the AP attains a service rate of $\frac{n_s}{t_{sf}}$ (packet/sec). Since the AP maintains a voice connection to each of the client nodes and n_s can be large, the packet availability rate is very high for the queue of the AP. Therefore, we do not half the above delay to avoid any over-estimation of call capacity. Using the above analysis, the channel access delay d_c for uplink and downlink packets are given by

$$d_c = \begin{cases} \frac{t_{sf}}{2} & \text{for uplink packets,} \\ \frac{t_{sf}}{n_s} & \text{for downlink packets.} \end{cases} \tag{5.14}$$

To estimate the call capacity of a WLAN, we need to consider the worst case voice quality impairments in the network so that every call can maintain an acceptable level

of voice quality. The channel access delay for the uplink packets being longer than that for the downlink packets, it suffices to consider only the uplink packet delay as channel access delay in the capacity model. For this reason, we use $d_c = \frac{t_{sf}}{2}$ to determine the channel access delay in our call capacity estimation model.

5.4.3 Delay and Loss in the Queue

It is evident from the Chapters 3 and 4 that the delay and loss in the queue are determined by the packet arrival rate, service rate, queue size and the number of network interfaces. The packet arrival rate can be determined using the codec configuration while the service rate is defined by the delay in channel access. Therefore, we use the queuing models explained in Section 3.7 and Section 4.4 for singe-interface and multi-interface WLAN nodes, respectively.

5.5 Traffic Variation— One Way Traffic

Each VoIP call requires two voice streams running in opposite directions (each stream runs from the mouth-piece of one end to the speaker at the other). To incorporate one way voice streams (used in IP-radio and IPTv), we additionally consider one way traffic. In our previous chapters (Chapter 3 and 4), we considered two way traffic only since the incorporation of one way traffic model with DCF is straightforward and does not require any major change. With the DCF based medium access, the channel access delay is a function of the number of nodes in the collision domain, i.e., the network size in a single hop network, and the traffic arrival rate of the nodes. Therefore, using one way traffic instead of two way traffic requires us to only adjust the traffic arrival rate appropriately. The Markov model that we proposed in Section 3.8 can be used with varying packet arrival rates and network sizes. In order to model the network performance with one way traffic in a DCF based WLAN, we need to determine the modified packet arrival rate only, which is half of that with two way traffic for a CBR traffic model like VoIP, and use the Markov model for the DCF based medium access mechanism without any further change.

In the PCF based medium access, the use of one way traffic shortens the length of a transaction and requires additional considerations. As every transaction is shortened, the total length of a super-frame becomes shorter. As a result, the channel access delay and the end-to-end delay become shorter giving a lower delay impairment. Furthermore, as explained in this section, the channel time usage for the uplink and downlink traffic are also different which results in different super-frame lengths for uplink-only and downlink-only traffic. Therefore, we need to carefully consider these special cases.

5.5.1 Downlink-only Traffic

If traffic is downlink-only, e.g., streaming video, Internet radio, IPTv, etc., the downlink packets carry the associated payload and the uplink packets contain acknowledgment information only. We redefine the transitional delays for the downlink–only traffic as follows.

$$\mathbb{D}_r = p_e(t_{sifs} + t_d + t_{pifs}) + p_e(1 - p_e)(2t_{sifs} + t_d + t_a), \tag{5.15a}$$

$$\mathbb{D}_s = 2t_{sifs} + t_d + t_a, \tag{5.15b}$$

$$\mathbb{D}_e = p_e(t_{sifs} + t_d + t_{pifs}) + (1 - p_e)(2t_{sifs} + t_d + t_a), \tag{5.15c}$$

where t_a is the time to transmit an acknowledgment-only frame. An acknowledgment-only frame is much smaller in size since it carries no payload but control information only. For instance, with the G.711 codec and an aggregation level of 10, the transmission time of a VoIP frame and an acknowledgment frame over a 1 Mbps channel are 7100 μs and 304 μs, respectively.

5.5.2 Uplink-only Traffic

On the other hand, when the traffic is uplink-only, e.g., P2P seeding, web hosting, etc., the downlink packets contain polling information only while the uplink packets contain the associated payload. In this case, we redefine the transitional delays for the uplink-only traffic as follows.

$$\mathbb{D}_r = p_e(t_{sifs} + t_{poll} + t_{pifs}) + p_e(1 - p_e)(2t_{sifs} + t_{poll} + t_d), \tag{5.16a}$$

$$\mathbb{D}_s = 2t_{sifs} + t_{poll} + t_d, \tag{5.16b}$$

$$\mathbb{D}_e = p_e(t_{sifs} + t_{poll} + t_{pifs}) + (1 - p_e)(2t_{sifs} + t_{poll} + t_d), \tag{5.16c}$$

where t_{poll} is the time to transmit a polling-only frame. Similar to an acknowledgment-only frame, a polling-only frame is also very small giving a short transmission time. For the example scenario mentioned above for data frame vs. ACK frame transmission times, a polling packet takes only 352 μs to transmit.

5.6 Rigid versus Variable CFP Repetition Interval

We recall from Section 2.3.2.2 that every CFP starts with the transmission of a Beacon frame. We also recall that the Beacon frames are transmitted repetitiously at intervals of the Beacon Interval. The Target Beacon Transmission Time (TBTT) is carried by every Beacon and control frames, like an association confirmation frame, so that the client nodes know when to expect the next Beacon frame. The CFP repetition interval

or the length of a super-frame is an integer multiple of the Beacon interval which is also broadcasted to the client nodes through these control frames.

The length of a super-frame is a function of the number of nodes (which is equal to the number of transactions) in a WLAN as shown in (5.12). In a properly designed WLAN, the number of nodes or the expected number of simultaneous voice calls can be estimated statistically. Thus, it is possible to configure the access point to use a CFP repetition interval and Beacon interval so that every active node, i.e., a node with an ongoing voice call, is polled once in every super-frame and the transmission of every Beacon frame starts a CFP. Alternatively, if a steady number of calls can not be expected, a call admission control module can be used instead in the AP, which will keep records of any ongoing calls and adjust the CFP repetition interval and Beacon interval accordingly in a dynamic manner. In a WLAN where all the nodes are PCF compatible, the time spent in the contention period is essentially wasted. In either of the above two cases, i.e., adjusting these intervals using statistical or CAC based methods, the channel time wastage in the contention period can be minimized by adjusting the length of the super-frame t_{sf} that can be determined using (5.12).

However, if static configuration or dynamic adjustment of the CFP repetition interval becomes inapplicable for some reason, the AP can be configured with a CFP repetition interval t_{cri} so that every Beacon still starts a CFP to reduce the number of unnecessary Beacon transmissions. If this new CFP repetition interval becomes greater than the required super-frame length to poll each of the active nodes once, the CFP will be shortened by an early transmission of the CF-End frame. Therefore, despite the value of t_{cri}, the super-frame length can still be determined using (5.12).

On the other hand, if the new CFP repetition interval is shorter than the required super frame length to poll all the client nodes, only a fraction of all the nodes can be polled in each super-frame. Let us denote the number of nodes that can be polled in each super-frame by n_p. The length of such a super-frame t'_{sf} can be shown to be

$$t'_{sf} = t_{cp} + t_{pifs} + t_b + t_{sifs} + t_{ce} + n_p t_t. \tag{5.17}$$

The number of nodes that is polled in each super-frame is given by

$$n_p = \left\lfloor \frac{t'_{sf} - t_{cp} - t_{pifs} - t_b - t_{sifs} - t_{ce}}{t_t} \right\rfloor. \tag{5.18}$$

It will take a total of $\left\lceil \frac{n_s}{n_p} \right\rceil$ super-frames to poll each of the nodes once and the expected channel access delay is given by

$$d'_{sf} = \frac{n_s}{2} \frac{t_{cp} + t_{pifs} + t_b + t_{sifs} + t_{ce} + t_t \left\lfloor \frac{t'_{sf} - t_{cp} - t_{pifs} - t_b - t_{sifs} - t_{ce}}{t_t} \right\rfloor}{\left\lfloor \frac{t'_{sf} - t_{cp} - t_{pifs} - t_b - t_{sifs} - t_{ce}}{t_t} \right\rfloor}. \tag{5.19}$$

The channel access delay and other performance measures can be determined using these changed factors.

5.7 Length of the Contention Period (CP)

The length of the contention period t_{cp} is a function of the number of contending nodes. Since the PCF-compatible nodes are served during the contention free period, only PCF-incompliant nodes are expected to compete for channel access during the contention period. Additionally, a number of works, including [3], suggested using the PCF mechanism during a CFP to transmit voice packets and to use the DCF mechanism during a CP to transmit background data packets. Therefore, PCF-compatible nodes with lower-priority data can also be expected to compete during the contention period. We discuss the expected length of the contention period in this section.

5.7.1 Variable Length CP Based on Number of Nodes

If there are PCF-incompliant nodes or nodes with data packets that need to be transmitted during the contention period, we can use the Markov model presented in Chapter 3 to determine the contention period length. The number of contending nodes is to be used as the network size, however, the packet arrival rate needs to be adjusted appropriately. With a packet arrival rate of λ packet/second, the total number of packets generated by a node during a super-frame is λt_{sf}. But the node gets only t_{cp} time to deliver these packets to the access points. Therefore, the effective packet arrival rate, in this case, becomes $\lambda \frac{t_{sf}}{t_{cp}}$. Considering these changes, an appropriate queuing model (for single NI or multi-NI nodes) can be used to determine the delay and loss in the queue.

5.7.2 Constant CP to Transmit a Maximum Size Packet

The original IEEE 802.11 standard suggests that the minimum length of the contention period should be such that a maximum size packet can be transmitted and acknowledged. This suggestion ensures that the PCF-incompatible nodes are not starved. However, when all nodes support PCF based medium access, such a mechanism is wasteful in terms of the channel access delay and loss. Although the IP layer supports a maximum packet size of 1500 bytes, the 802.11 MAC layer allows a maximum packet size of 2304 bytes. The 1500 byte restriction is a popular MTU (Maximum Transmission Unit) choice for the Internet Protocol (IP) due to inherent limitations of the Ethernet protocol. As a result, if an application tries to send a packet that is larger than 1500 bytes, the IP layer splits it into multiple fragments and sends them

separately. The IP layer at the receiving end then merges these fragments to recreate the original packet. This mechanism is called fragmentation. However, custom applications can skip the IP layer and send a packet that is larger than 1500 bytes. If the packet is larger than 2304 bytes, the MAC layer will use fragmentation and transmit it in parts. Denoting the payload size of a maximum size packet with l_{max}, the length of the contention period can be presented as

$$t_{cp} = t_{difs} + 8\left(\frac{l_{max} + l_m}{r_d} + \frac{l_p}{r_b}\right) + t_{sifs} + t_a, \tag{5.20}$$

where r_d and r_b are the data and basic transmission rates, respectively, t_{difs} is the length of a DIFS period and $l_{max} = 2304$ if no encryption mechanism is used. If MAC level encryption is used, the payload size increases after the encryption process. The 2304 bytes of payload in a maximum size packet become 2312, 2324 or 2320 bytes after encryption using WEP (Wired Equivalent Privacy)[1], TKIP-WPA (Temporary Key Integrity Protocol—Wi-Fi Protected Access)[2] or CCMP-WPA2 (CCM mode Protocol—Wi-Fi Protected Access 2)[3], respectively.

5.7.3 Arbitrary Length CP for Idle Waiting

During the contention free period, the access point controls access to the channel. As no collisions are expected, neither the access point nor the client nodes perform physical carrier sensing (Section 2.3.3) before a transmission. As a result, if there are multiple access points within the interference range of each other, successive collisions are expected. The IEEE 802.11 standards keep a provision for an additional backoff in such a case. If an access point finds the channel busy when a new CFP is supposed to start, it delays the Beacon transmission for a random number of idle slots. The random number is chosen uniformly from the range $[0, 1, \ldots, W]$ where W is the initial contention window. This is the reason some authors used the following expression for the length of the contention period

$$t_{cp} = t_{idle}\frac{W}{2} \tag{5.21}$$

5.8 Call Capacity Model for PCF based WLANs

Due to the Constant Bit Rate (CBR) nature of the VoIP applications under consideration, the packet arrival rate of the queue is deterministic. On the other hand, in a network where all nodes are PCF-compliant, the lengths of both the CFP and CP

[1] http://en.wikipedia.org/wiki/Wired_Equivalent_Privacy
[2] http://en.wikipedia.org/wiki/Wi-Fi_Protected_Access
[3] http://en.wikipedia.org/wiki/CCMP

are deterministic. As a result, the length of the super-frame is predictable and the channel access delay and the service rate of the queue are also deterministic. In such a queuing system, if the packet arrival rate becomes greater than the service rate, the queue grows quickly and becomes full. This results in a high queuing loss which adversely affects the voice quality. To avoid such a scenario, the service rate should be greater than or equal to the arrival rate of the queue. With an aggregation level of n_a and a voice frame length of d_f, this requirement can be represented as

$$\frac{1}{n_a d_f} \leqslant \frac{1}{d_c}. \tag{5.22}$$

We use (5.14) and (5.13) to determine the delay and loss in the PCF based medium access mechanism and an appropriate queuing model for single NI or multi-NI nodes to determine the delay and loss in the queue. Using (5.22) to avoid a high queuing loss, and using end-to-end delay (3.17) and loss (3.20) in the voice quality impairment budget (3.16), we formulate the VoIP call capacity of a PCF based WLAN using the following optimization problem.

$$\begin{aligned} Max \quad & n_s n_v \\ s.t. \quad & I_d(d_e) + I_{e\_eff}(e_e) \leqslant \begin{cases} 28.3553 & \text{for medium quality,} \\ 13.1952 & \text{for high quality,} \end{cases} \\ \text{and} \quad & d_c \leqslant n_a d_f. \end{aligned} \tag{5.23}$$

As in the previous chapters, in this optimization problem, we try to maximize the total number of voice calls while maintaining the voice quality impairments within their total budget. Therefore, the optimization problem essentially gives the maximum number of quality voice calls that can be attained in the WLAN under consideration.

5.9 Throughput of a PCF based WLAN

Although we do not use the throughput to determine voice capacity in this thesis, we discuss the throughput of a PCF based WLAN in this section for a better understanding of the WLAN performance. In the following, we determine the MAC layer throughput for two way and one way traffic.

5.9.1 Throughput— Two Way Traffic

With two way traffic, each transaction adds $2(n_a l_f + l_r + l_u + l_i + l_m)$ bytes of transmitted payload per super-frame where n_a is the aggregation level and l_f is the frame size in bytes. Similarly, l_r, l_u, l_i and l_m are the headers (in bytes) added by the RTP, UDP, IP and MAC layers. In t_{sf} time, the total transmitted data bits give the

throughput ψ (in bps) which is contributed by n_s transactions. However, the e_c portion of all the transmitted packets are ultimately dropped due to exceeding the retry limit. Therefore, the throughput ψ for a PCF based WLAN can be defined as

$$\psi = \frac{16n_s\left(n_a l_f + l_r + l_u + l_i + l_m\right)\left(1 - e_c\right)}{t_{sf}}, \tag{5.24}$$

where t_{sf} is determined using (5.12).

5.9.2 Throughput— One Way traffic

With one way traffic, the throughput decreases substantially since payloads are carried by either the uplink or the downlink frame only. We redefine the throughput as

$$\psi' = \frac{8n_s\left(n_a l_f + l_r + l_u + l_i + l_m\right)\left(1 - e_c\right)}{t''_{sf}}, \tag{5.25}$$

where t''_{sf} is the associated length of a super-frame considering the modified delay elements for one way traffic as explained in Section 5.5.

5.10 Results and Analyses

In this section, we investigate different aspects of a PCF based WLAN and their impacts on voice traffic. We use the analytical model presented in this chapter to determine the call capacity of a PCF based WLAN. The analytical results derived from our model are also verified by extensive simulations.

5.10.1 Effect of Traffic Load

The traffic load is defined by the network size and the packet arrival rate. The aggregation level n_a and the voice frame length d_f of the codec give a packet arrival rate of $\lambda = \frac{1}{n_a d_f}$. The traffic load closely dictates the performance of a PCF based WLAN. In the following, we analyze some performance measures including queue length, queuing delay, queuing loss and voice quality rating for different traffic loads in a PCF based WLAN.

5.10.1.1 Queue Length l_q

As mentioned before, the delay and loss in the queue are crucial performance measures of the voice traffic. A small queue is prone to becoming full which results in frequent packet drop and high queuing loss. On the other hand, a large queue reduces the packet loss; but as the packets tend to linger in the queue for longer intervals, the

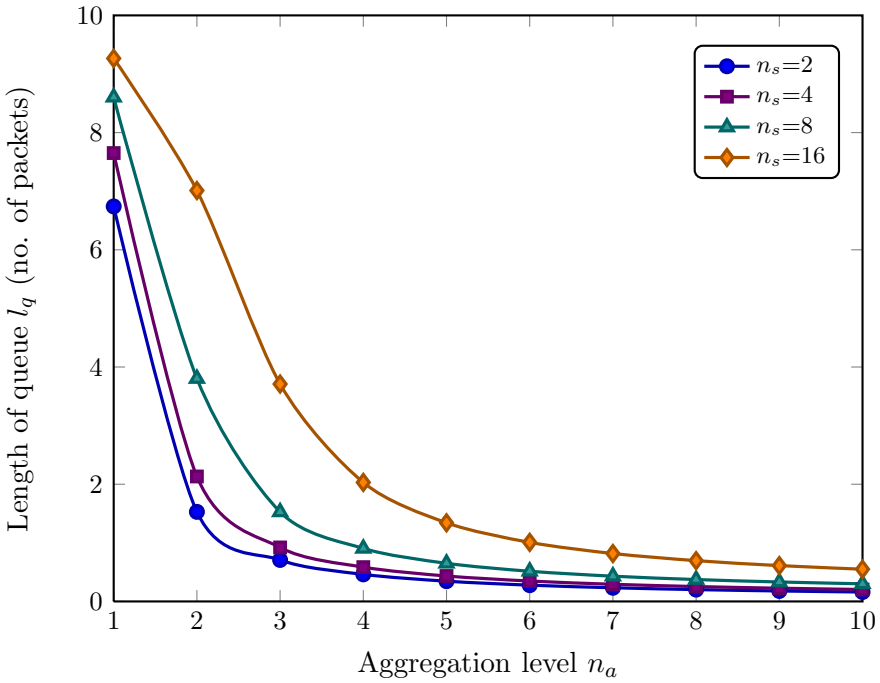


Figure 5.3: The queue length with traffic load variations for the G.729 codec with different aggregation levels in PCF based WLANs with 2, 4, 8 and 10 nodes, 10 packet queue size and 1 Mbps data rate showing shorter queue length for lower traffic arrival rate and smaller WLAN.

queuing delay becomes high. Determining the optimal queue size is, therefore, an important network design issue. For a given queue size, the average queue length gives an estimate of the queue utilization in terms of the number of packets that can be expected statistically over a long period. We present the expected queue length in Fig. 5.3 for the G.729 codec in a PCF based WLAN with a data transmission rate of 1 Mbps. To reflect a high queuing loss scenario, we use a queue size of 10 packets. As the aggregation level n_a increases, the packetization delay increases and the packet arrival rate decreases. Thus, for a given network size n_s, the offered traffic load decreases with increasing n_a. As a result, the queue occupancy decreases. In the presented scenario, the expected queue length l_q decreases sharply as n_a increases from 1 to 4, beyond which the decrement becomes slower. With $n_s = 2$, the queue length is 6.74, 0.46 and 0.16 for aggregation levels of 1, 4 and 10, respectively. On the other hand, with increasing network size, the super-frame length increases since every node is polled once during the CFP. As a result, the channel access delay increases which reduces the service rate of the queue. For a given aggregation level, packet arrival rate to the queue remains constant, and therefore, a decreased service rate of the queue results in a higher packet availability in the queue and a longer queue length. With aggregation

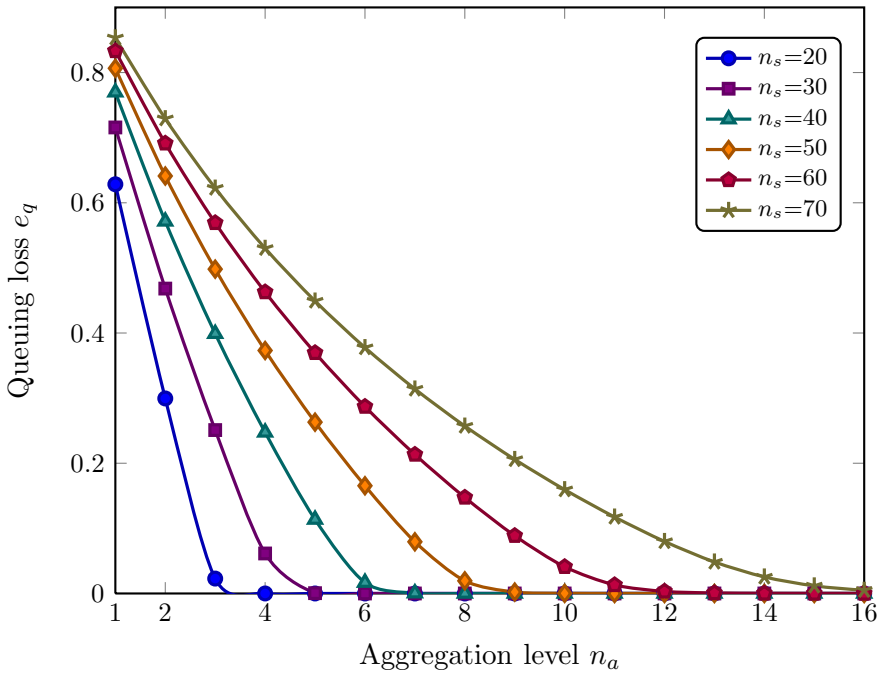


Figure 5.4: The queuing loss versus traffic load relationship for the G.729 codec with different aggregation levels in a PCF based WLAN with a variable number of nodes and 1 Mbps data rate illustrating the increase in queuing loss for a lower aggregation level and a larger network size.

level $n_a = 1$, the queue length is found to be 6.74, 7.65, 8.60 and 9.26 for network sizes of 2, 4, 8 and 16 nodes, respectively. Although the network size is doubled every time, the queue length increased at a relatively lower rate.

5.10.1.2 Queuing Loss e_q

Any packet that arrives at an already full queue is dropped. Therefore, queue occupancy has a direct relationship with queuing loss. As mentioned in the last section, traffic arrival rate and queue occupancy decreases with increasing aggregation level. Similarly, queuing loss also continuously decreases with increasing aggregation level, but it increases with a higher number of nodes which is presented in Fig. 5.4. Similar to the previous section, we use a queue size of 10 packets to reflect a high queuing loss scenario. To demonstrate the nature of queuing loss variation, we use six network sizes (of 20, 30, 40, 50, 60 and 70 nodes) with the aggregation level varying from 1 to 16. The decreasing curve of queuing loss can be split into two regions. In the first region, queuing loss decreases almost linearly with the aggregation level until its level and the network size conjunctively offer a sufficiently low traffic load that keeps queuing loss nearly zero in the second region for any further increase in the aggregation level. For

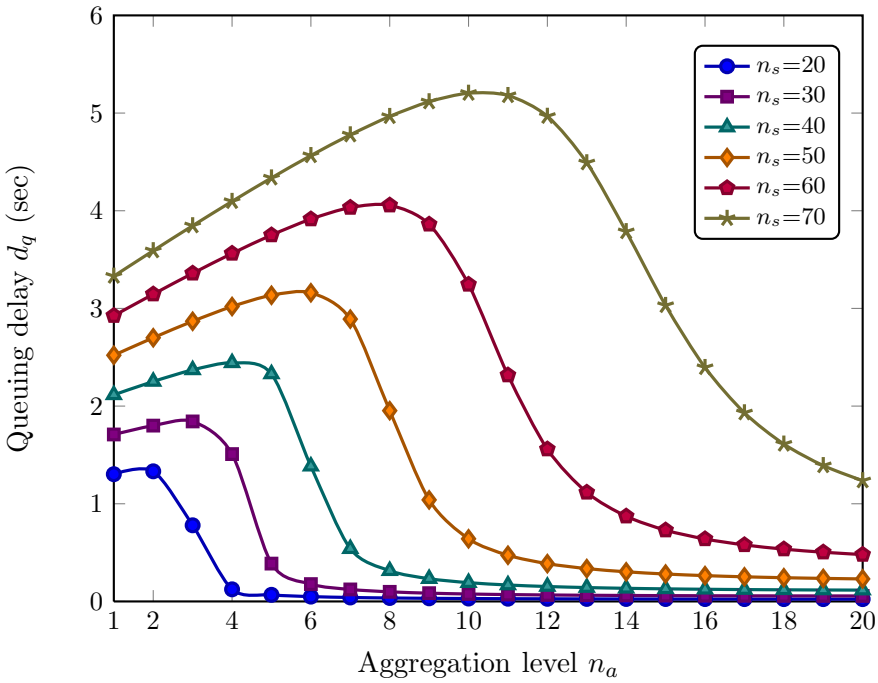


Figure 5.5: The queuing delay for the G.729 codec with different aggregation levels and network sizes in a PCF based WLAN with 1 Mbps data rate and a queue size of 10 packets.

the 40-node WLAN, queuing loss is 0.77 at $n_a = 1$ which quickly decreases to 0.016 at $n_a = 6$ and remains close to zero for $n_a \geqslant 7$.

5.10.1.3 Queuing Delay d_q

We demonstrate queuing delay for the same network conditions used in the last section for queuing delay. The results are illustrated in Fig. 5.5. Queuing delay variation can be split into three regions.

1. When the aggregation level is low, the packetization interval remains short and offered traffic load becomes high. As a result, the channel access delay becomes long and packets linger in the queue for a long time incurring a higher queuing delay. The queue, in this region, remains prone to becoming full and the queuing loss (Fig. 5.4) attains a high value. As the aggregation level increases, the queuing loss also decreases which means more and more packets will contribute to give the expected queuing delay (since only the packets that are not dropped contribute to queuing delay). Therefore, despite the decrease in the offered load, queuing delay increases and queuing loss decreases. For the $n_s = 40$ WLAN scenario in Fig. 5.5, this region is identified as $n_a = 1 \sim 4$.

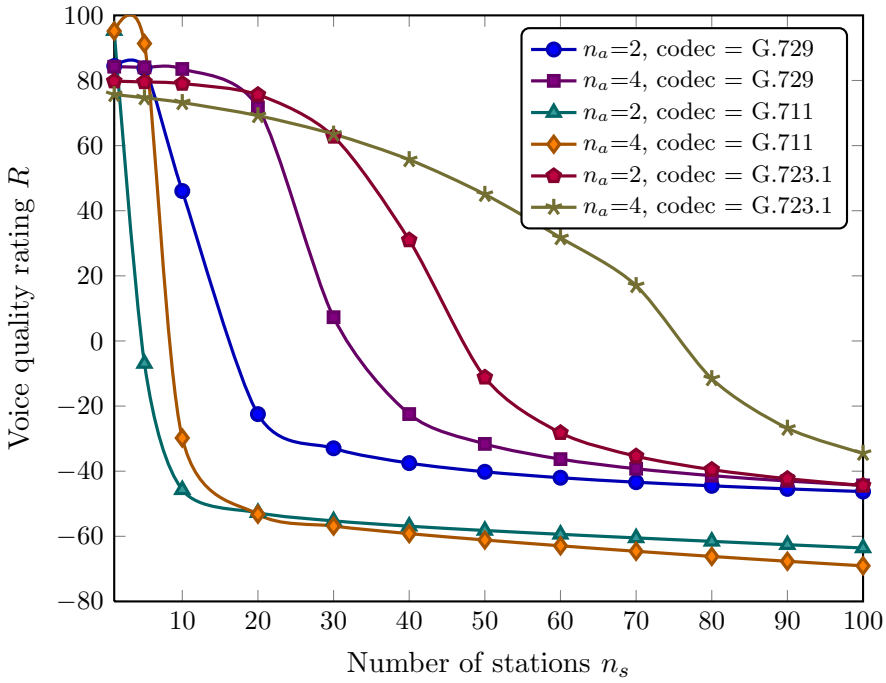


Figure 5.6: Voice quality rating R for different network sizes and codecs with 1 Mbps data rate showing voice quality degradation with increasing WLAN size.

2. When the queuing loss becomes insignificant (Fig. 5.4) for a traffic load, the second region starts. The decrease in the queuing loss becomes minimal with increasing aggregation level but the queuing delay (Fig. 5.5) starts to decrease sharply. We find this region to be $n_a = 5 \sim 7$ for the 40-node WLAN in Fig. 5.5.

3. In the third region, the offered load is minimal which gives a low queue occupancy and nearly zero queuing loss. Queuing delay is also insignificant in this region which is a good indication of under utilized network resources. This region is $n_a \geqslant 8$ for the 40 node WLAN in Fig. 5.5.

5.10.1.4 Voice Quality Rating

Figure 5.6 depicts the resulting voice quality for G.729, G.711 and G.723.1 codecs in terms of the E-model rating R. Here we present results for only two aggregation levels but we observed similar trends for other aggregation levels as well. (Although R-score is shown in Fig. 5.6 for a wider range to better understand the trend of voice quality, an R score is considered to be 0 whenever its value is negative). The delay and loss impairments result in a decreased voice quality with any increase in the network size. Additionally, the R score decreases for a decrease in the aggregation level due to the higher packet generation rate. Therefore, the voice quality is adversely affected by the

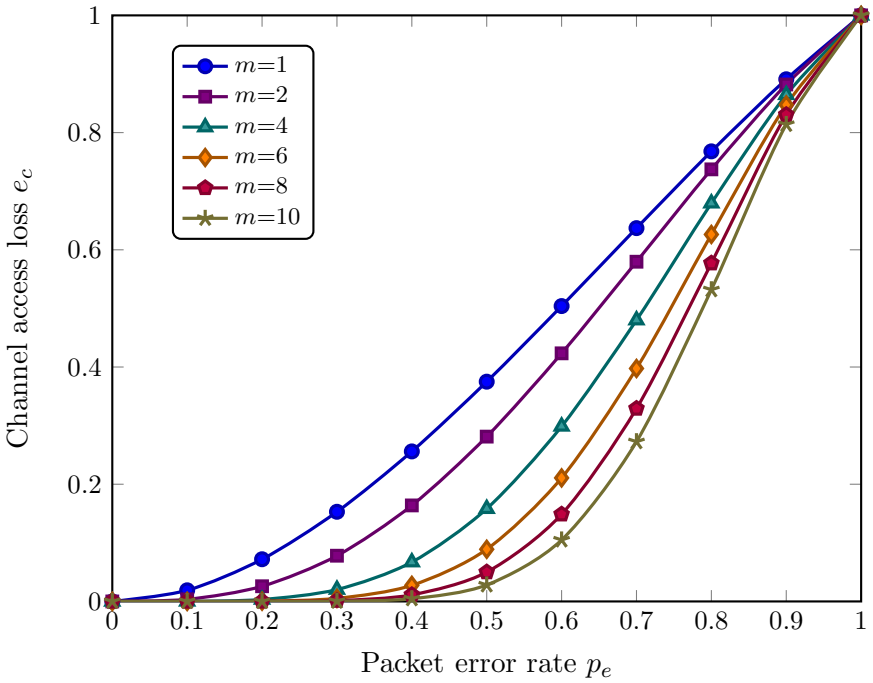


Figure 5.7: The channel access loss versus the packet error rate and retry limit in a 20-node WLAN with 1 Mbps data rate using the G.729 codec and $n_a = 2$ showing a higher channel access loss for a lower retry limit and a higher packet error rate.

offered traffic load which increases with both increasing network size and decreasing aggregation level. G.711 and G.723.1 show sharper and slower decrease, respectively, in the R score than G.729. This is primarily due to the higher and lower bandwidth consumption (compared to G.729) of G.711 and G.723.1 codec, respectively.

5.10.2 Effect of Imperfect Channels

We investigate the effect of an imperfect channel and the various features closely related to packet loss, e.g., retry limit, queue size, etc., in this section. As mentioned before, the effect of an imperfect channel is considered as packet error rate which is defined as the ratio of the number of transmissions affected by channel errors to the total number of transmissions.

5.10.2.1 Channel Access Loss e_c

With an imperfect channel, a transmission that is exposed to channel error is dropped by the receiving station. The same packet is retransmitted by the sending station until a pre-configured retry limit is exceeded. Once the retry limit exceeds, the packet will be dropped and considered as channel access loss. Channel error adversely affects the

network performance due to the increased channel access loss which is illustrated in Fig. 5.7. We use a 20-node WLAN where data is transmitted at 1 Mbps. Voice traffic is generated using the G.729 codec with an aggregation level of 2. When the packet error rate is low, fewer transmissions are impaired by channel errors and channel access loss e_c remains low. Once the packet error rate attains a considerable value, channel access loss exponentially increases, and at $p_e = 1$, e_c becomes 1 which means that every transmission is affected by channel error and all packets will be dropped eventually by the medium access control layer. From (5.3) and (5.13) we can rewrite the channel access loss as

$$e_c = \left(2p_e - p_e^2\right)^{m+1},$$

which has a quadratically decreasing part $(-p_e^2)$ before the exponential operation. While $2p_e$ increases linearly with p_e, p_e^2 increases quadratically, and therefore, the increase in the channel access loss slows down for very high values of p_e. We observed that for $p_e > 0.8$, the rate of increase in channel access loss slightly decreases. We consider retry limits of 1, 2, 4, 6, 8 and 10 to illustrate their effect on the channel access loss. With an increasing retry limit, a packet is retransmitted a higher number of times. As a result, the packet attains a higher probability of being received without error, and packet loss decreases.

5.10.2.2 Queuing Loss e_q

Figure 5.8 presents queuing loss with an imperfect channel. As expected, queuing loss e_q increases with increasing packet error rate. However, when the packet error rate increases beyond a threshold of $p_e = 0.6$, the queuing loss decreases until another threshold of $p_e = 0.8$ and then continues to increase with increasing packet error rate. This sudden decrease in queuing loss does not necessarily mean network performance improvement. Actually, when packet error rate enters this region, most of the transmissions are affected by channel error. As a result, the retry limit expires quickly and the packets are dropped by the MAC layer at a high frequency (as shown in Fig. 5.7). The queue can not differentiate between the success and failure of a transmission; and when a packet is fetched and dropped by the MAC layer, it is considered to be served from the queue point of view. Therefore, the service rate of the queue becomes high and queuing loss decreases. However, for $p_e > 0.8$, the rate of increase in channel access loss slightly slows down (Fig. 5.7), compared to that for $0.6 \geqslant p_e \geqslant 0.8$. As a result, the increase in the service rate of the queue also slows down leading to an increased queuing loss.

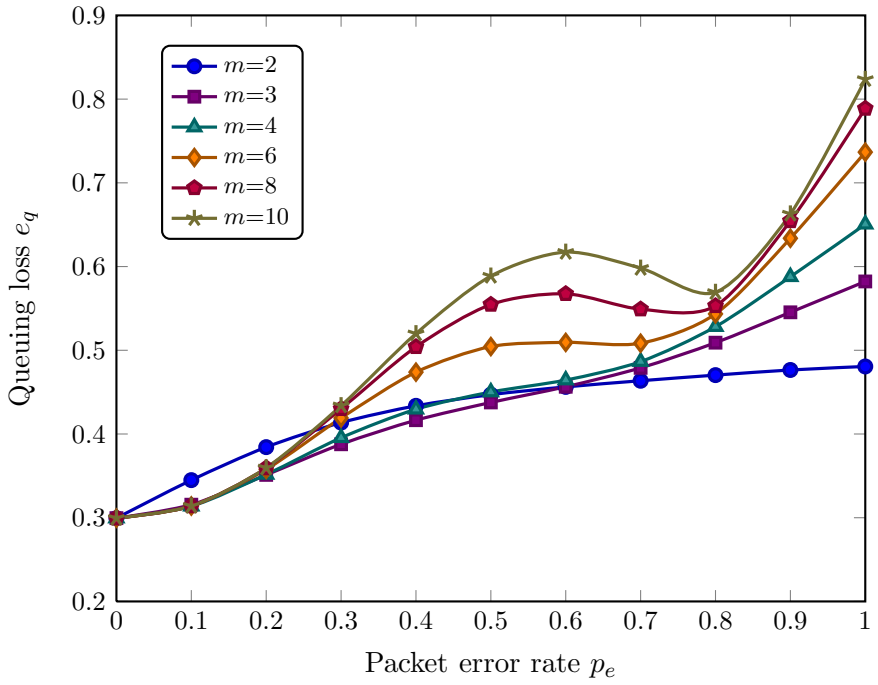


Figure 5.8: The queuing loss versus the packet error rate and retry limit in a 20-node WLAN with 1 Mbps data rate showing increase in queuing loss for higher retry limit and packet error rate.

5.10.2.3 Queuing Delay d_q

Queuing delay also shows a similar trend compared to queuing loss. Figure 5.9 illustrates the queuing delay for the same WLAN. With increasing packet error rate, the queuing delay increases. The queuing delay decreases slightly in the range $p_e = 0.6 \sim 0.8$ and then continues to increase again. This trend of queuing delay can be explained using the justifications presented in the last section.

5.10.3 Effect of Retry Limit

In order to reduce packet loss in both channel access and queue, transmission retries play a crucial role. In this section, we investigate the impact of varying retry limit on channel access loss. With increasing retry limit, a packet is retransmitted a higher number of times. As a result, the packet attains a higher probability to be transmitted with success and the channel access loss decreases as shown in Fig. 5.10. As retry limit increases, channel access loss decreases exponentially due to the exponential nature of e_c. Therefore, channel access loss can be kept close to zero, i.e., insignificant, by using a particular retry limit depending on the packet error rate. In our given illustration

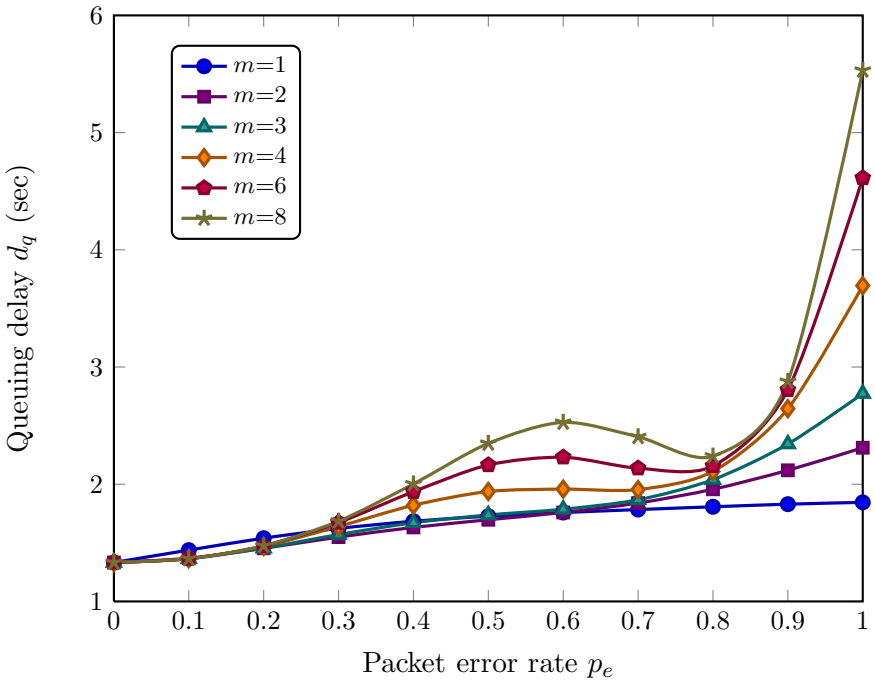


Figure 5.9: The queuing delay versus the packet error rate and retry limit for the G.729 codec with an aggregation level of 2 in a 20-node WLAN with 1 Mbps data rate illustrating the increase in queuing delay for higher retry limit and packet error rate.

the channel access loss becomes close to zero (less than 0.005) at retry limits of 2, 4, 7 and 10 when the packet error rate p_e are 0.1, 0.2, 0.3 and 0.4, respectively.

5.10.4 Effect of Queue Size

Queue size is a crucial factor in determining voice quality. The primary benefit by increasing the queue size is achieved from a reduced queuing loss. With a larger queuing buffer, a higher number of packets can be accommodated in the queue. As a result, the queue becomes less prone to becoming full and queuing loss decreases. Use of a smaller queue has the opposite effect as shown in Fig. 5.11. With an increasing queue size, queuing loss quickly decreases initially in the range $s_q = 1 \sim 7$. Further increase in the queue size decreases the queuing loss slowly and for the ideal channel, i.e., where $p_e = 0$, the queuing loss attains a small value of 0.01 at $s_q = 50$. With increasing packet error rate, a higher number of retransmissions is required for each packet. Therefore, the packets wait in the queue for a longer period and a larger queue is required to attain a lower queuing loss.

Figure 5.12 illustrates queuing delay for different queue sizes and packet error rates in a 1 Mbps WLAN with 20 nodes. We use the G.729 codec with an aggregation level

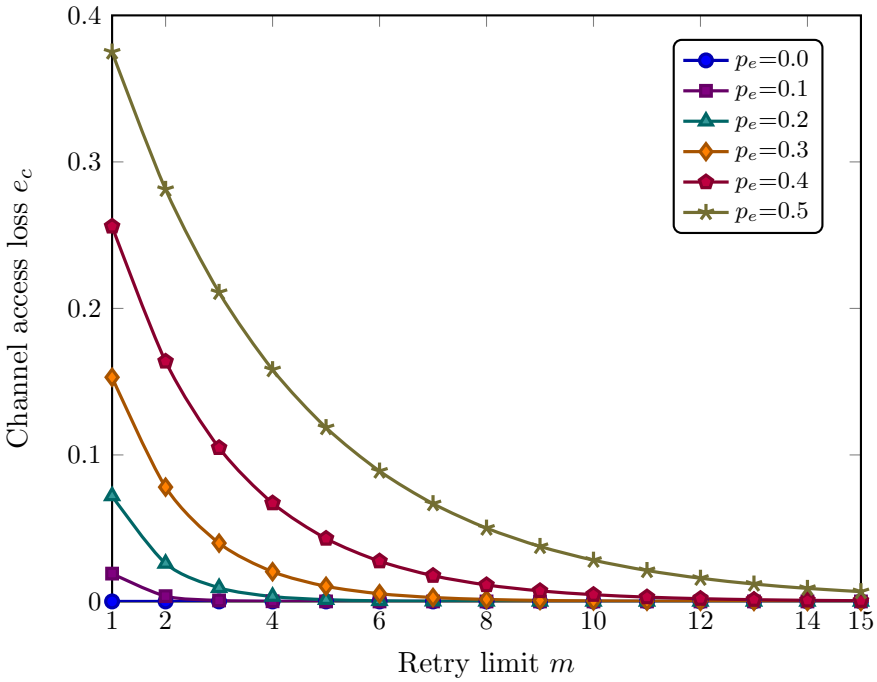


Figure 5.10: Channel access loss versus retry limit and packet error rate for G.729 codec with an aggregation level of 2 in a 20-node WLAN with 1 Mbps data rate showing decrease in channel access loss for higher retry limit and lower packet error rate.

of 2 and packets are retransmitted for a maximum of 5 times. As the packet error rate increases, a higher number of packet transmissions suffer from channel errors and a higher number of retransmissions is needed for each packet. During the retransmissions, a packet stays in the queue and queue occupancy increases. With a larger queue size, more packets can be buffered which reduces queuing loss (Fig. 5.11); but as the packets tend to linger for a longer period, queuing delay increases.

The resulting voice quality rating is shown in Fig. 5.13 for different queue sizes and packet error rates. We use the G.729 codec with $n_a = 2$ in a 10-node PCF based WLAN with 1 Mbps data rate and $m = 5$. With relatively smaller sizes of the queue ($s_q = 1 \sim 15$), the fast decrement of queuing loss outweighs the voice quality impairment caused by the moderate increase in the queuing delay and, therefore, the voice quality rating R increases. However, for relatively larger queue sizes ($s_q \geqslant 16$), the queuing loss plateaus but the higher queuing delay drives R at increasingly lower values, i.e., the voice quality decreases.

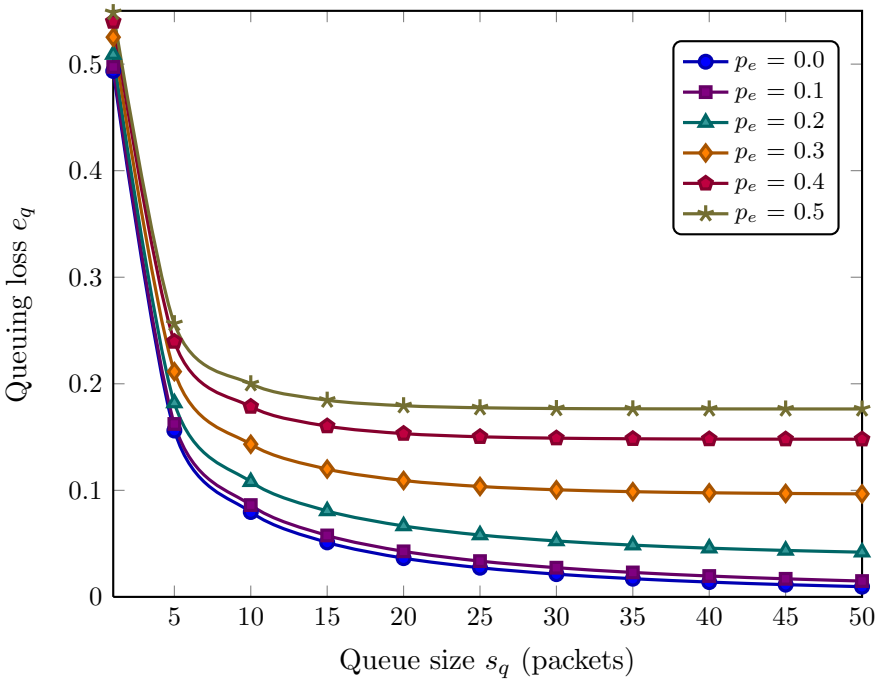


Figure 5.11: The queuing loss for different queue sizes and packet error rates for the G.729 codec with $n_a = 2$ in a 20-node PCF based WLAN with 1 Mbps data rate showing decrease in queuing loss for a larger queue size and lower packet error rate.

5.10.5 Voice Call Capacity

We investigate the voice capacity in terms of the number of supported calls in a PCF based WLAN. We also verify the estimated results from our analytical model using a custom simulator. Although a patch to support PCF in NS-2 was reported by Anders Lindgren [156], it is no longer supported (the project discontinued) and can not be used with the latest version of NS-2 (the current version being 2.34 at the time of writing this thesis). While used in conjunction with the targeted version NS-2.1b8a, the PCF implementation in [156] was found to incorrectly poll and wrongly respond to association requests. PCF implementations were also reported for the GloMoSim simulator in [157–159], however, when the author of this thesis contacted the authors of [157–159], some [158, 159] did not reply and the other [157] did not have the code any more. Currently, PCF is supported by OPNET simulator which is only sold commercially and it is identified by exhaustive search that no current open source network simulator supports the PCF based medium access mechanism. In this light, the author of this thesis designed and developed a custom network simulator which is much similar to NS-2 in architecture and is under the process to be published as a stand-alone open source simulator. A short description of this custom simulator *atisim* is given in Appendix A.2.

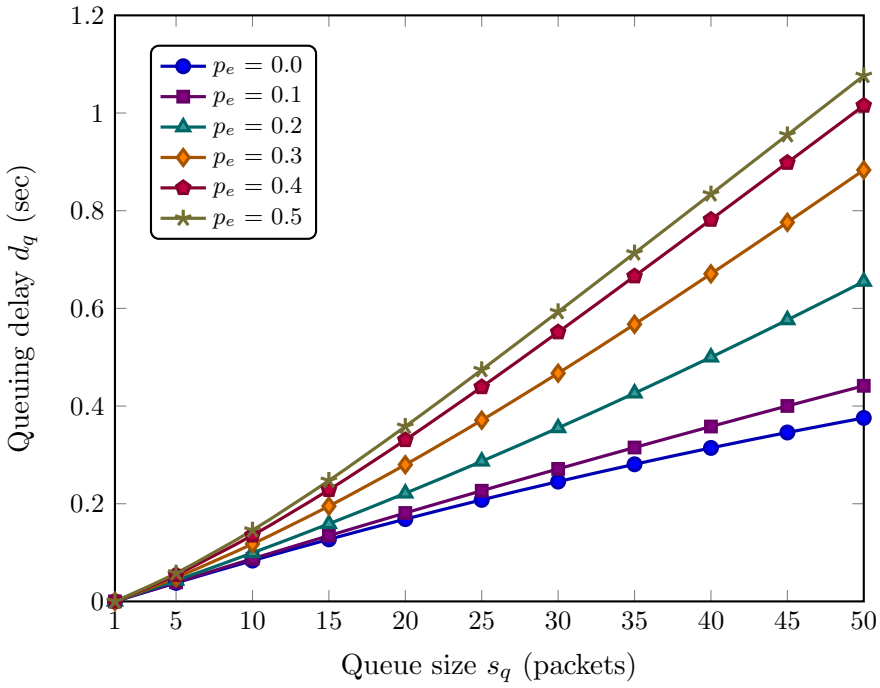


Figure 5.12: Queuing delay for different queue sizes and packet error rates for the G.729 codec with $n_a = 2$ in a 10-node WLAN with 1 Mbps data rate and $m = 5$ where the queuing delay increases for a larger queue size and a higher packet error rate.

In the simulations using *atisim*, we used a single-channel infrastructure WLAN where the point coordinator (PC) resides in the access point. The AP controls access to the channel using the PCF based medium access mechanism. Since PCF is applicable to single hop network only, all of the nodes reside in a single collision domain. To consider the impact of imperfect channel, the "channel" class of our simulator (please see Appendix A.2) introduces channel errors uniformly over all transmissions. Each network configuration was simulated for 500 seconds and the results were averaged over 20 such simulations. The randomization seed of each simulation run were different. All simulation results plotted in this section yielded $\rho < 10^{-8}$ at 99% confidence interval.

Figure 5.14 and 5.15 illustrate the VoIP call capacity in a WLAN with data rates of 1 Mbps and 11 Mbps, respectively. We use the most popular codecs G.729, G.711 and G.723.1 and call capacity is shown as a function of the aggregation level. The simulation results are also presented in the figures along with the analytical results. The simulation results match the analytical data closely which validates our proposed Markov model for the PCF based medium access. As aggregation level n_a increases, packet generation rate $\frac{1}{n_a d_f}$ decreases offering a lower traffic load. As a result, the delay and loss in the medium access and queue decrease and call capacity increases. At the

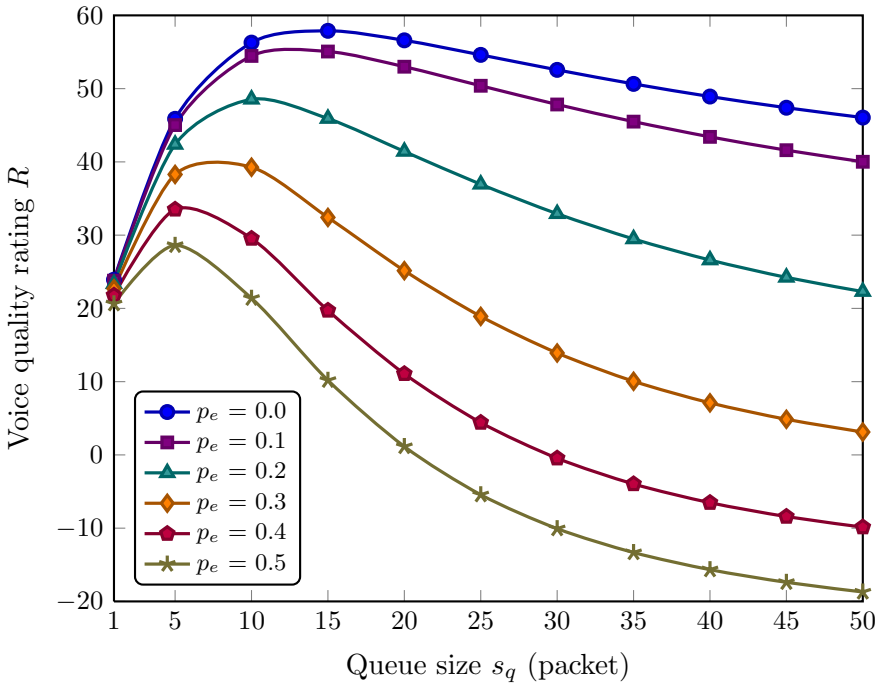


Figure 5.13: The voice quality rating R according to the ITU-T E-model for different queue sizes and packet error rates for the G.729 codec with $n_a = 2$ in a 10-node PCF based WLAN with 1 Mbps data rate and $m = 5$ illustrating the voice quality degradation due to a higher packet error rate.

same time, a greater number of voice frames are merged in a single packet which requires a longer transmission time. Additionally, the increase in the packetization interval $n_a d_f$ results in a longer end-to-end delay. Due to these two opposing effects, initially call capacity increases with increasing aggregation level. But after a certain threshold of n_a, call capacity starts to decrease for all codecs. The call capacity at this point is the maximum achievable capacity by the WLAN. G.729 and G.723.1 compress the voice frames more than G.711 and support a maximum of 31 and 32 calls, respectively, in the 1 Mbps scenario. G.711 compresses voice the least and consumes more bandwidth. As a result, only a maximum of 7 calls can be supported. The 11 Mbps scenario (Fig. 5.15) also shows a similar relationship between call capacity and aggregation level. However, with the higher data rate (11 Mbps), the transmission time becomes shorter. As a result, the super-frame length becomes shorter, and a higher number of calls can be supported. A maximum of 151, 63 and 144 calls are supported using the G.729, G.711 and G.723.1 codecs, respectively.

The effect of packet error rate on the voice call capacity is presented in Fig. 5.16 along with relevant simulation results. We use the G.729 codec and a data rate of 1 Mbps. We also verified similar results for 11 Mbps in our simulations. As packet error

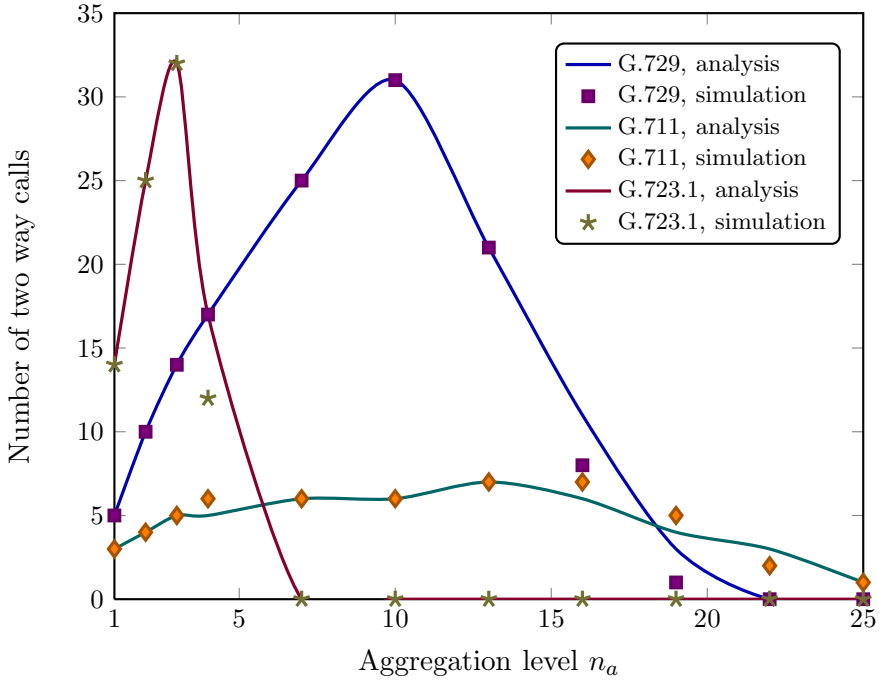


Figure 5.14: The VoIP call capacity for different codecs using different aggregation levels in a PCF based WLAN with 1 Mbps data rate.

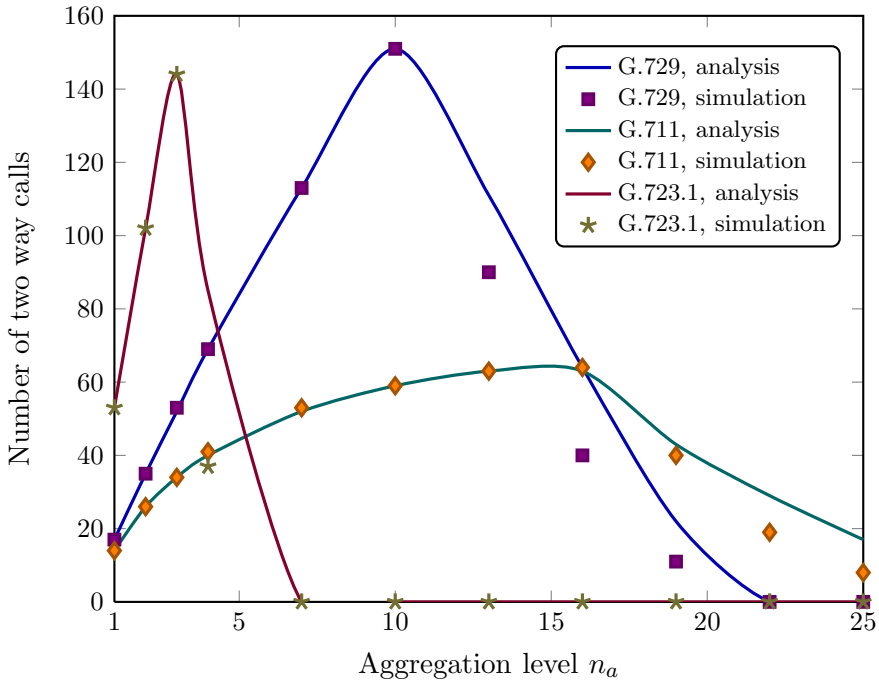


Figure 5.15: The VoIP call capacity for different codecs using different aggregation levels in a PCF based WLAN with 11 Mbps data rate.

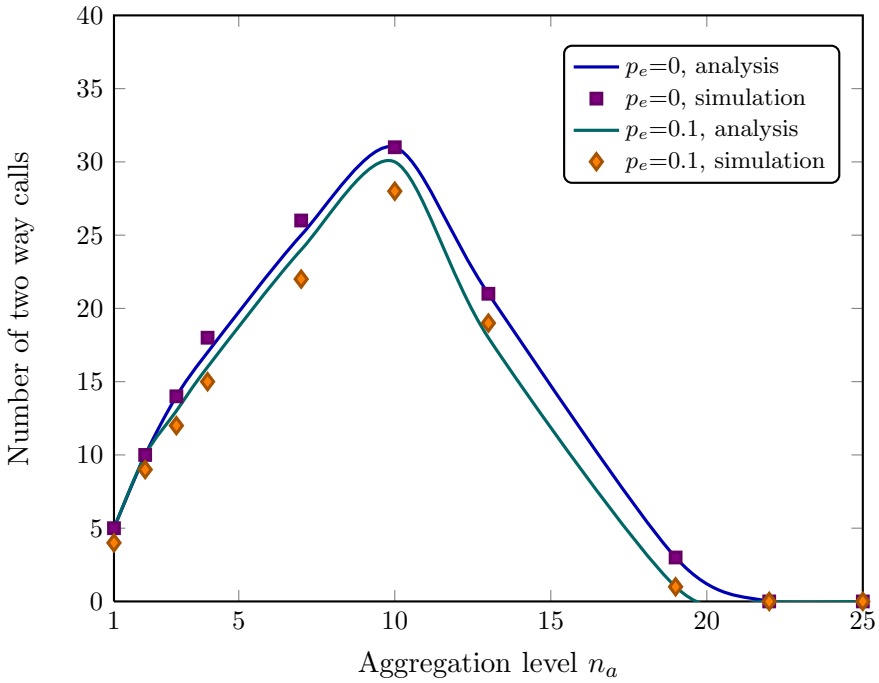


Figure 5.16: The call capacity for different packet error rates using the G.729 codec at different aggregation levels and 1 Mbps data rate.

rate increases, transmission errors are introduced in a higher number of transmissions and the call capacity decreases. When packet error rate is p_e is 0 and 0.1, the call capacity of the 1 Mbps WLAN are found to be 31 and 30, respectively.

5.10.5.1 Impact of Data Transmission Rate r_d

The effect of data transmission rates is shown in Fig. 5.17 for 1, 2 and 5.5 Mbps at an aggregation level of $n_a = 10$. A quick rise in the call capacity occurs when the data transmission rate is initially increased from 1 Mbps. Note that this data rate is used to transmit the data packets only. MAC control packets are always transmitted at the basic rate of 1 Mbps and preamble and PLCP headers are transmitted at the control data rate of 1 Mbps [160]. This is why the increase in the call capacity is not proportional to the increase in the data transmission rate. For the G.729 codec, the maximum call capacity are 31, 55 and 109 at 1, 2 and 5.5 Mbps, respectively. The non-proportional increase in the call capacity with the data transmission rate can be better presented in a log scale (for data transmission rates) as illustrated in Fig. 5.18. We present the call capacity for all data rates and both G.729 and G.711 codecs with an aggregation level of 10. (The capacity of the G.723.1 are not shown as it does not support any call with $n_a = 10$). With 11 and 54 Mbps data rates, the call capacity

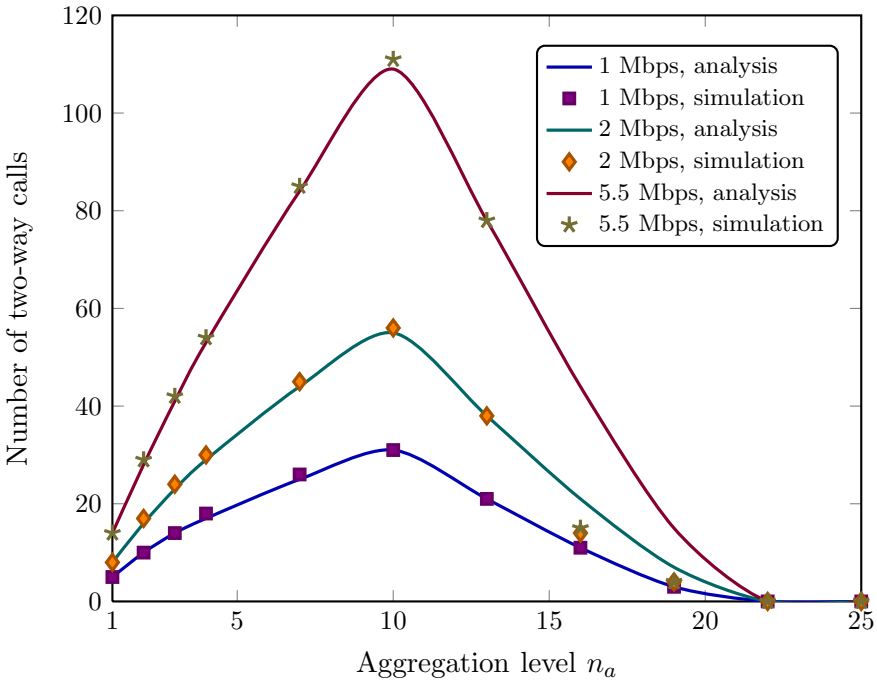


Figure 5.17: The impact of aggregation levels and data rates on the voice call capacity using the G.729 codec.

for G.729 are found to be 151 and 218, respectively. Both Fig. 5.17 and 5.18 show reasonable accuracy between simulation and analytical results.

5.10.5.2 Effect of Call Quality Requirement— Impairment Budget

So far we have analyzed the capacity for medium quality voice calls only. High quality voice requires an stringent impairment budget and limits the capacity to a lower degree. For comparison purposes, Fig. 5.19 presents results for both high and medium quality calls in a PCF based WLAN with 1 Mbps data rate. Although G.711 supports the lowest number of medium quality calls compared to the other two codecs, it supports a relatively moderate number of high quality calls. Both G.729 and G.723.1 have a high initial equipment impairment factor at zero loss (I_e) of 11 and 15, respectively, compared to the 0 initial equipment impairment factor of G.711 (please see Table 3.1). As a result, G.729 and G.723.1 support no high quality call at 1 Mbps data rate although these codecs consume less bandwidth than G.711.

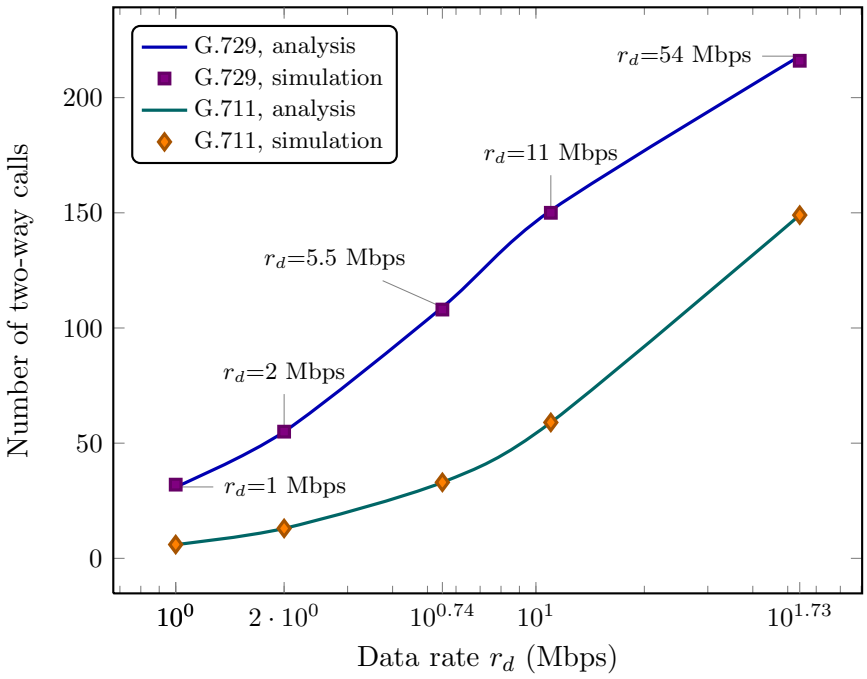


Figure 5.18: The call capacity of the G.729 and G.711 codecs with different data rates r_d at an aggregation level of 10 in a PCF based WLAN (shown on a log-X axis for the data rate).

5.10.5.3 Comparison of Voice Call Capacity of the DCF and PCF based Medium Access Mechanisms

Figure 5.20 presents a comparison of VoIP call capacity with the DCF and PCF based medium access mechanisms with 11 Mbps data rate. All codecs support higher number of calls with the PCF than the DCF based medium access mechanism which is primarily due to the lower delay and loss in medium access and queue. G.729, G.711 and G.723.1 support only 45, 34 and 40 medium quality calls using the DCF mechanism while they support 151, 63 and 144 calls using the PCF mechanism. This performance gain with the PCF based medium access is intriguing enough to promote the use of PCF in future mobile devices. Similar improvement is also observed for high quality calls using the G.711 codec.

5.11 Key Contributions

In this chapter, we modeled the VoIP call capacity of a PCF based WLAN. The following are the key contributions from this work.

- The existing voice capacity analyses for the PCF based WLANs either used an ideal channel or considered the number of retransmissions only and disregarded

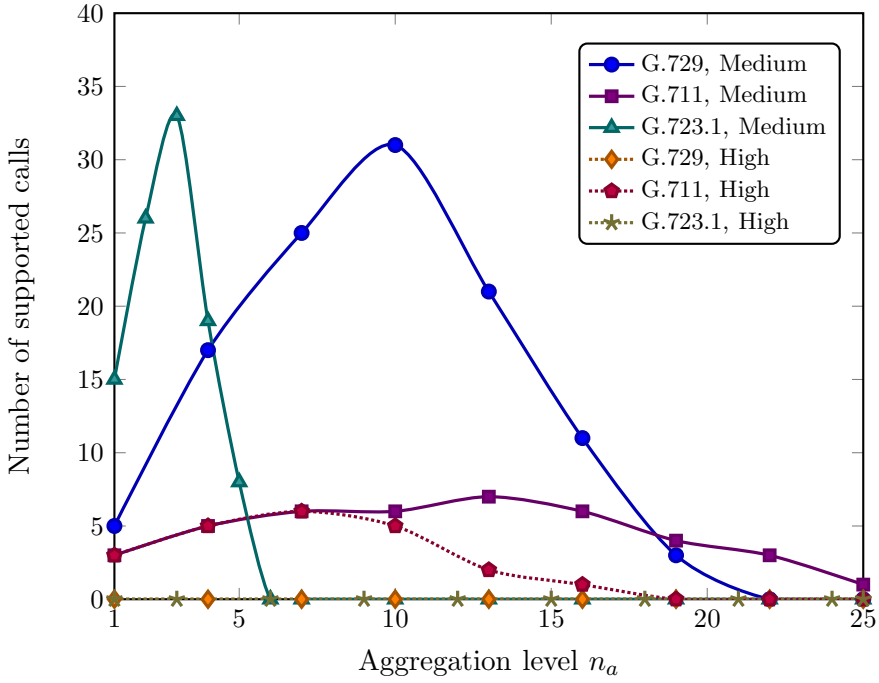


Figure 5.19: The call capacity variation due to voice quality requirement for high and medium quality calls using the G.729, G.711 and G.723.1 codecs at different aggregation levels in a PCF based WLAN with 1 Mbps data rate.

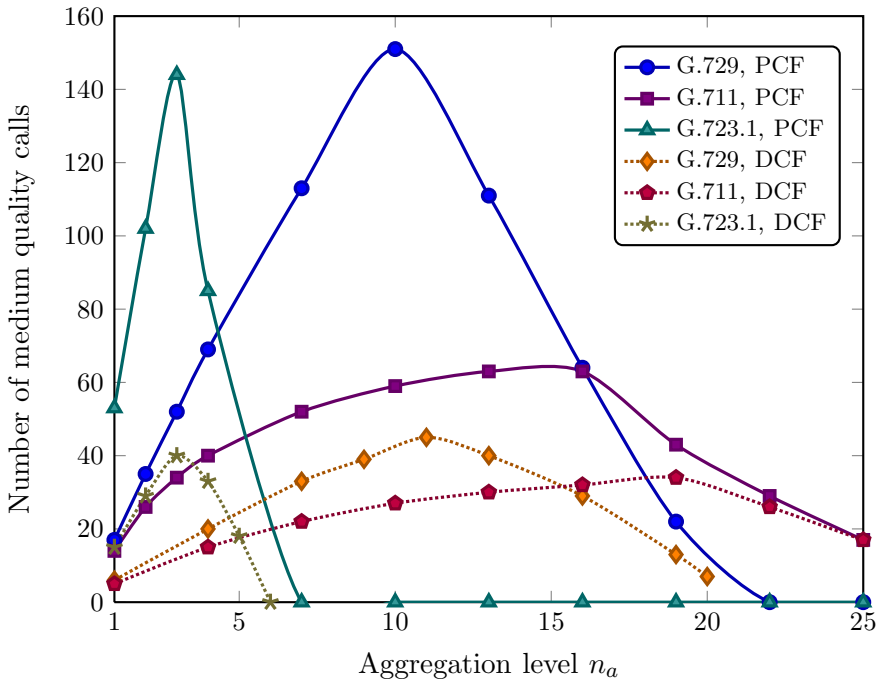


Figure 5.20: PCF versus DCF call capacity comparison for the G.729, G.711 and G.723.1 codecs at different aggregation levels with 11 Mbps data rate.

the difference in the channel time wastage due to channel errors in uplink and downlink transmissions. While the former assumption (the use of an ideal channel only) gives an over-estimation of call capacity, the latter (ignoring the difference in channel time usage for uplink and downlink transmissions) results in an under-estimation. Due to these reasons, the existing works provided inaccurate estimations of the call capacity in real networks. We considered the effect of imperfect channels and the probabilistic delay elements of a PCF transaction in our proposed Markov model (illustrated in Fig. 5.2) to overcome these limitations. The proposed Markov model allows us to determine the delay and loss in the PCF based medium access mechanism with reasonable accuracy. But most importantly, the call quality requirements and the delay and loss in the queue were never considered (in the existing works) which we incorporated in our call capacity estimation model.

- We determined the most useful performance measures for voice traffic and used these measures to formulate a call capacity model as an optimization problem in (5.23). The proposed model utilizes the voice quality impairment budget introduced in Chapter 3. Congestion at the access point can become a bottleneck which degrades the performance of the WLAN. We identified the necessary condition to avoid such a scenario and presented it as a part of the optimization problem so that the formation of such bottlenecks can be avoided in the WLANs.

- We additionally modeled variation from the ideal two way traffic condition and included uplink only and downlink only traffic. Additional considerations to model rigid or variable length CFP repetition intervals were incorporated. We also considered the length of a contention period using a few standard strategies.

- We investigated the impact of channel errors, data transmission rate, retry limit, queue size and impairment budget for high and medium quality calls. Extensive simulations validated our proposed call capacity model. We also compared the voice capacity for high and medium quality calls and the call capacity of PCF and DCF based WLANs.

5.12 Summary

In this chapter, we investigated the call capacity of a PCF based WLAN. We considered an infrastructure WLAN where a PC located at the AP coordinates access to the channel. To model a real network environment closely, we considered an imperfect channel and analyzed its impact in conjunction with the queue size and the

retry limit. The existing studies of PCF with consideration for imperfect channels presented inadequate analyses in determining the delay and loss in the channel access, and ignored the voice quality requirements and queue performance. We presented a Markov model of the PCF based medium access mechanism and used the voice quality impairment budget introduced in Chapter 3 to determine the call capacity of a PCF based WLAN. Using the proposed Markov model, the impacts of different network aspects on the voice traffic performance and voice call capacity of a PCF based WLAN were determined. We also verified some of these results using extensive simulations. Simulation results validated our numerical analysis as reasonably accurate. Our proposed capacity model can be of great assistance in designing voice networks and call admission control systems in PCF based WLANs. The performance benefit from the use of the PCF based medium access mechanism intrigues us to promote its use. However, a major limitation of the PCF based medium access mechanism is that a PC needs to be available to coordinate the channel access which limits the use of PCF strictly to single hop WLANs only. In the following chapter, we propose a mechanism to overcome this limitation.

Chapter 6

Increasing Voice Capacity of Multihop WLANs using vAPs

6.1 Introduction

In the last chapter, we presented a comparison between the voice capacity of the PCF and DCF based medium access mechanisms. The comparison demonstrated that a higher call capacity can be attained when PCF is used, rather than DCF, by utilizing its time synchronized medium access mechanism which ensures that collisions do not occur and eliminates the need of any backoff altogether. Additionally, in a single hop WLAN of n client nodes, the AP needs to transmit n times more packets than any of the client nodes as mentioned in Section 2.6.4 and in [124]. In the DCF mechanism, the AP has to compete for channel access with the same priority as the client nodes and an unbalanced traffic condition occurs. As a result, a bottleneck is formed at the AP, especially when the traffic load is high. This problem does not arise in PCF since the AP can send one downlink packet to each client and each client can send one uplink packet to the AP in every contention free period.

PCF requires an AP (with a co-located PC) to coordinate the channel access and poll the client nodes. Due to the absence of an AP, PCF is not operable over multiple hops and, thereby, fails to provide a large coverage and spatial reuse like DCF. Multihop WLANs offer wireless coverage for a larger area at a lower cost than single hop WLANs since only a few access points with broadband connections can provide the Internet connectivity to a large number of client nodes. This is why, the DCF based medium access mechanism is a popular choice, despite the high delay and loss incurred by it, in public WLANs where cost is a dominating and decisive factor.

Compared to wired networks, voice capacity is lower in the IEEE 802.11 WLANs and, therefore, techniques need to be devised that can increase voice capacity substan-

tially. In this respect, we plan to design a new medium access control mechanism by combining the best features of DCF and PCF. Our proposed medium access mechanism uses a time synchronized channel access mechanism like PCF so that a low delay and loss in the medium access can be attained but, at the same time, it can also be used in multihop WLANs like DCF so that a wide coverage can be provided. The IEEE 802.11 standards do not offer such a mechanism and despite the call capacity enhancement efforts in the current literature as reviewed in Section 2.6, the time synchronized medium access is still not operable over multiple hops. In the following, we explain the proposed scheme which also considers real world factors including an imperfect channel, node mobility and traffic load variation in multihop WLANs. At the same time, we make the proposed scheme open to all nodes in a fair manner to avoid channel hogging.

6.2 Proposed Medium Access Mechanism— The Basic Concept

Our proposed medium access mechanism for a multihop WLAN utilizes a PCF-like time synchronized medium access to avoid the excessive delay and loss in the channel access associated with DCF. But such a time synchronized medium access mechanism needs to be coordinated by a Point Coordinator (PC) which is located at the AP in PCF. DCF is operable over multiple hops due to its ad hoc nature of operation that enables the client nodes to carry out the channel access granting mechanism themselves. But to coordinate the time synchronization in our proposed scheme, one node in each collision domain needs to control access to the channel in lieu of a regular PC. Therefore, we propose that in the absence of an access point (with a co-located PC), a selected client node will carry out the task of a PC as required by the PCF mechanism. A client node can virtually become an access point for a short duration, and such a node is henceforth termed as a vAP (virtual Access Point) in this chapter. Future mobile devices are expected to be equipped with multiple Network Interfaces (NI) [21] which can significantly improve the performance of a multihop WLAN using our proposed vAP based medium access mechanism. Therefore, we use multi-NI nodes as vAPs for the general discussions in this chapter. A specific scenario for single-NI nodes operating as vAPs is discussed in Section 6.6. A typical multihop WLAN where selected client nodes are coordinating the channel access as virtual access points is shown in Fig. 6.1. For tractability of analysis, we split the multihop WLAN into a set of concentric annulus regions as in Chapter 4. At each hop, some nodes are acting as the vAPs and forwarding packets to and receiving from the client nodes in the outer

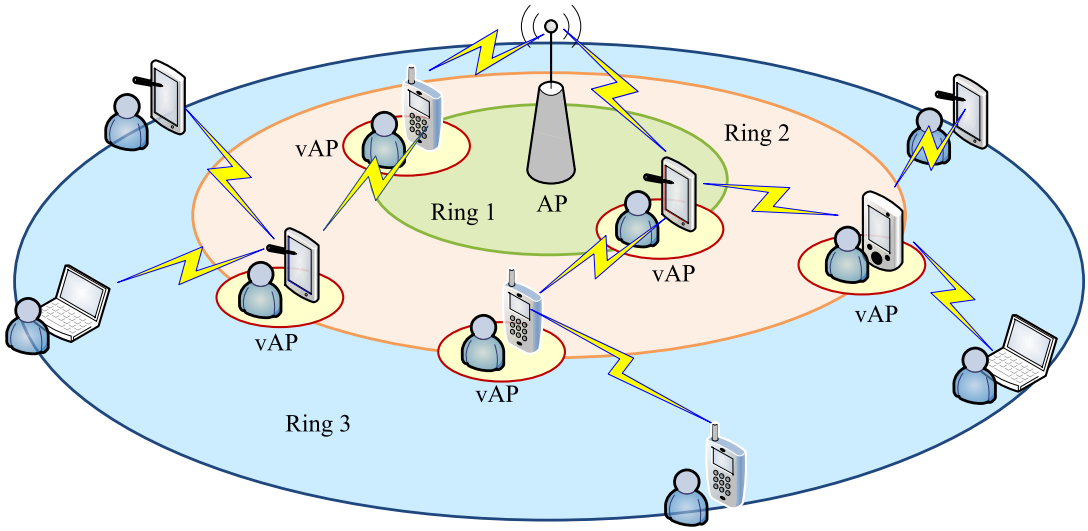


Figure 6.1: A multihop WLAN with virtual access points coordinating channel access at each hop. The virtual access points are marked with yellow circles under them.

rings to and from the access point located at the center of the WLAN. In Fig. 6.1, the vAPs are marked with yellow circles under them.

Our proposed medium access mechanism is built using the existing protocols of the DCF and PCF mechanisms so that a non-vAP compliant node (a node that is not aware of our proposed scheme) can not interfere, and can operate under a vAP's coordination without any problem. In addition to the medium access mechanism, a detailed vAP selection mechanism is devised so that exactly one node becomes a vAP in a collision domain for a given contention period to avoid collisions. Additionally, no change in the PHY layer is necessary and the existing DSSS, FHSS or IR based PHY layer, as defined in the IEEE 802.11 standards, can be used. The modifications proposed in addition to the IEEE 802.11 standards to define the medium access mechanism and the vAP selection procedure are purely in the lengths of waiting time, handshake mechanism and packet formats. We outline the medium access mechanism and the vAP selection process in the following two sections. The vAP selection process ensures fairness to all nodes using arbitration and avoids channel hogging. A real world mobility model is used to further guide the selection process in order to avoid performance degradation due to node mobility.

6.3 Virtual Access Point (vAP) based Medium Access Mechanism

Our proposed medium access mechanism closely follows the 802.11 PCF based medium access mechanism. Once a client node is designated as a virtual access point, it coordinates the channel access for exactly one contention free period. The vAP selection mechanism, as explained in the following section, is randomized to ensure fairness to all nodes. Here, by fair, we mean that every willing node is equally likely to become a vAP if it has the same resources as the other competing nodes. The vAP selection process needs to be executed after every contention free period because client nodes, as opposed to regular APs, are expected to be mobile as they roam about. If a client node that is currently a vAP (and is coordinating the channel access) moves away, connections to some of the nodes in its old collision domain can be lost. If this vAP had some packets in its queue to be sent to those nodes, the transmission of those packets will result in no acknowledgment and eventually, those packets will be dropped. Therefore, it is essential to hand over the packets in the queue to the next hop nodes as soon as possible. If the vAP, however, remains in its current location any longer, it can bid in the following super-frames and also coordinate channel access, if wins. But every time the CFP expires, the authority of the last vAP as the PC ceases.

Therefore, we repeat the selection process before every CFP giving the most eligible node a chance to become the vAP. A vAP can be reselected if it meets the selection guideline, i.e., a client node may continue to be repeatedly selected as the vAP as long as it remains the most eligible among all the competing nodes in the selection process. Once a node becomes the vAP for a collision domain, it transmits a Beacon frame like the standard PCF based medium access mechanism and starts a CFP. However, unlike PCF and more like HCCA, the CFP can start at any time since there is no AP who can periodically transmit the Beacon frames announcing the next expected Beacon transmission time. The vAP waits for a SIFS after transmitting the Beacon frame and polls the first node in its polling list (explained in the following section) with a downlink data frame as done in the standard PCF mechanism. The client node waits for a SIFS and transmits an uplink data frame or acknowledgment frame to the vAP. The polling frame uses the standard PCF header so that the client node can reply correctly even if it is not aware of the vAP mechanism. Once every node in the collision domain of the vAP is polled in this way, the vAP transmits the CF-End frame announcing the end of the CFP.

The vAP then forwards all its collected uplink packets to the next hop destination node which can be another vAP that is nearer to the access point. This second node can act like a client in the first vAP's collision domain and channel, or it can be a

vAP operating in a second channel where the first vAP is a client. In this manner, the time synchronized medium access can be used at every hop over a multihop route from a client node to the AP as shown in Fig. 6.2. In Fig. 6.2, nodes A1 and B2 are vAPs in two adjacent collision domains and need to exchange packets between each other. However, adjacent vAPs can not operate in the same channel since it will result in frequent collisions. Such a situation will never take place since once a vAP starts a CFP by transmitting a Beacon frame, essentially all other nodes will start using virtual carrier sensing (explained in Section 2.3.3), and will not access the channel or become another vAP.

In the scenario presented in Fig. 6.2, nodes B2, B3 and B4 can access both channels A and B using two NIs each, and are shown by having shadow nodes in both channels. Nodes A2, A3 and A4 are such shadow nodes indicating that they are the NIs of nodes B2, B3 and B4, respectively, and are connected to channel A; while the NIs B2, B3 and B4 are connected to channel B. Being solely client nodes, B3 and B4 need to operate in only one channel (either channel A or B), preferably the one where the vAP is nearer to the AP (compared to the vAP in the other channel). But B2 needs to communicate in both channels while it is a vAP in channel B (using the interface B2) and a client node in channel A (using the interface A2). Although not shown in this figure, node A1 can also be a client in another channel in the same manner, and operate under the coordination of another vAP or a regular AP (if A1 is located in the first ring, i.e., ring 1) of the WLAN. The vAP B2 hands over all the packets collected from nodes B1, B3, B4 and B5 in channel B to the vAP A1 in channel A using the A2 interface.

6.4 The vAP Selection Mechanism

The vAP selection process is arbitrated using a bidding mechanism to ensure fairness to every node. The bidding mechanism uses a slightly modified DCF based handshake and negative voting or *veto* system. These modifications are in the lengths of the waiting time, handshake mechanism and packet format, as mentioned before, and ensure that non-vAP compliant nodes can operate without any problem. The winner of the bidding becomes the point coordinator for the following super-frame.

A node that is willing to bid in order to become vAP first observes the channel activity for a reconnoiter period of length t_{recce}. During this period, the node tries to detect the presence of a genuine access point from the reception of Beacon, CF-End or Polling frames. If an AP is detected during this period, the node does not bid and uses the standard PCF mechanism to send its own data to and from the AP since the AP can directly communicate with all the client nodes in the associated collision domain. Therefore, the proposed mechanism is not applicable to single hop

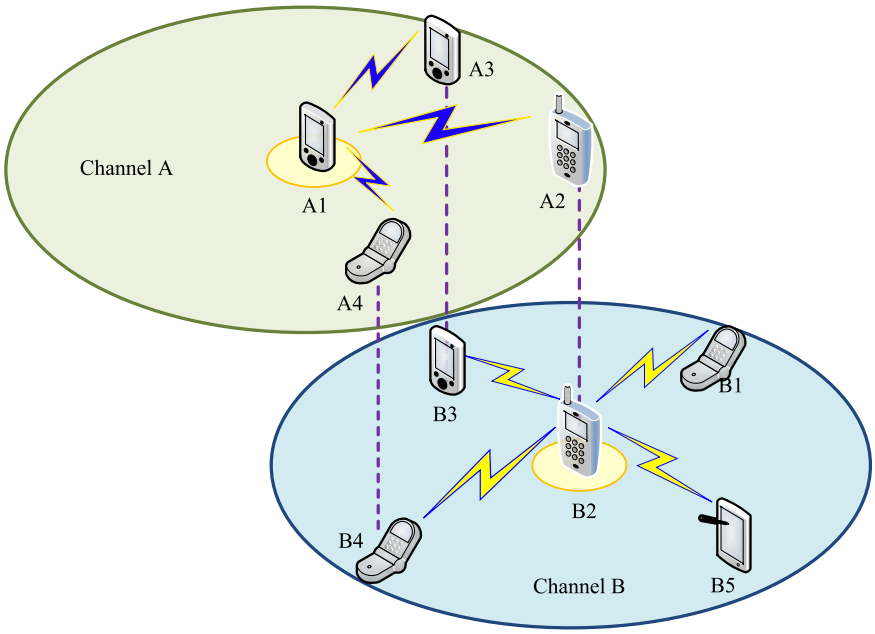


Figure 6.2: Packet handover from one vAP to another vAP using the proposed scheme. Nodes A1 and B2 (placed on yellow circles) are vAPs in channels A and B, respectively. A2 and B2 are two network interfaces of the same node and can operate in either of the channels. A similar analogy holds for (A3, B3) and (A4, B4), respectively. The vAP B2 hands over all the packets collected from nodes B1, B3, B4 and B5 in channel B to the vAP A1 in channel A using the A2 interface.

networks as well as the first hop (the nearest from the AP) of a multihop WLAN in the particular channel where the AP is operating, although other channels where an AP is not detected can be utilized by a vAP. If the presence of any AP is not detected from the ongoing transmissions, the node still continues to observe each packet (by using the network card in promiscuous mode) during t_{recce} to identify all other nodes in its collision domain and add them to a list which is later used as the polling list for the time synchronized channel access.

The length of t_{recce} has to be at least as long as t_m, which is the time to transmit and acknowledge a maximum size packet. After an initial channel time of t_m (idle or busy), the node additionally waits for the channel to be idle for a DIFS period of length t_{difs}. The length of the reconnoiter period can be longer than its minimum length of $(t_m + t_{difs})$ if channel activities delay the vAP's later transmissions, for instance, by the use of the RTS/CTS based channel reservation. A regular AP waits for $(t_m + t_{pifs})$, where t_{pifs} is the length of a PIFS, before transmitting the Beacon frame and can retain control of the channel without any competition since $t_{pifs} < t_{difs}$. After waiting for t_{recce}, the node determines a bidding value ζ randomly, and additionally waits for ζW number of idle slots (W is the initial contention window size) before transmitting a

Bidding frame. The Bidding frame is a simple data frame with zero bytes of payload and can be overheard by the other bidding nodes (who also want to be the vAP). The Bidding frame bears a new frame subtype (the IEEE 802.11 standards define a 2-bit type field and a 4-bit sub-type field in the MAC layer headers) called "Bidding" so that such packets can be easily identified by the access points and other vAP-enabled client nodes. A vAP-enabled client node is a regular client node which supports our proposed vAP based medium access mechanism. The Bidding frame is addressed to a randomly chosen station from the polling list and the node acknowledges with an ACK frame after a SIFS. The acknowledgment is performed at the MAC layer before processing the payload and any node performs it regardless of the currently employed medium access mechanism which can be any of DCF, PCF or vAP. The higher protocol layers of the acknowledging node will eventually discard the Bidding frame since it contains no data of interest to the application layer like the MAC control frames.

The successful reception of the acknowledgment to the bidding frame makes the bidder the new virtual access point for the following CFP. On the other hand, if an ACK frame is not received after a SIFS following the Bidding frame transmission, the bidder node has to start the bidding process again. This can happen if the combination of the start time of the bidding process and the bidding value ζ of a node result in a collision of the bidding frame transmission with some other contemporary frame transmission (which can be another bidding frame by another bidder or a regular data frame). However, it is not necessary to repeat the reconnoiter process since if an AP is not detected already, there is less chance that one would appear later and if an AP becomes available later anyway, it can secure control of the channel immediately, as explained in the following discussion.

When there are more than one node willing to bid, the bidding value ζ prioritizes the bidders since the node with the smallest ζ wins. We explain this parameter in details in the following section. The Bidding frame performs a secondary task by allowing a genuine AP to regain control of the channel. If a regular AP becomes available later, for instance after a reboot, it can acquire control of the channel in the following super-frame since the vAP waits longer than a regular AP. Additionally, the AP can immediately transmit a jamming frame after a SIFS period following the Bidding frame. The jamming frame contains random noise signals and its transmission time is equivalent to that of an ACK frame. The sole purpose of the jamming frame is to block the ACK frame's reception by the bidding node. This mechanism works like the jamming signal of the Ethernet protocol where a node transmits a jamming signal whenever it detects a collision so that every other node in the collision domain becomes aware of it. Only a genuine AP, and no other node, transmits a jamming frame which collides with the ACK frame ensuring that the bidding fails and the AP

can regain control of the channel immediately. Thus, the jamming frame acts as a *veto* to the bidding process. Although any other vAP-enabled client node can not transmit the jamming frame to veto an ongoing bidding process, a vAP-enabled node can transmit the jamming frame if it has detected that another vAP/AP (with a different MAC address) has already transmitted a Beacon frame in the same channel and a CFP is underway which has not yet finished (as determined by using virtual carrier sensing or the reception of a CF-End frame). In this way, a vAP-enabled client node can save most of the transmissions of a CFP by jamming only one packet's (the consequently transmitted packet in the ongoing CFP) transmission when two vAPs are detected. However, such a situation can occur only when two would-be vAPs are located in a way that gives rise to the hidden node problem, and two such vAPs try to operate in the same channel. Thus, the reception of an acknowledgment to the bidding frame ensures that there is no node between two distant vAPs forming a hidden node problem that can cause repetitious collisions. Once a node wins in a bidding, it waits for a SIFS period and transmits a Beacon frame with the broadcast address initiating a vAP controlled CFP as mentioned in the previous section.

6.5 Priority in Bidding— Determination of ζ

When there are multiple client nodes bidding at the same time, the node with the lowest value of the bidding priority parameter ζ wins and, in case of a tie (having the same value of ζ), the first one to transmit the bidding frame successfully (followed by a successful reception of an ACK frame) wins. However, the network will perform better if the associated vAP is more resourceful in terms of its memory, number of NIs, support for higher data rate and multi-rate MAC, etc. The bidding priority ζ is determined using a node's resources and capabilities to carry out the task of the vAP based medium access. This parameter performs an important role in boosting the network performance since it allows a node with higher resources to win in the bidding and, thereby, improves the network performance. We propose the following guidelines that can be used in determining ζ and mobile device manufacturers (device vendors) can implement these guidelines in manufacturing vAP compliant devices.

- **Compatibility to the IEEE 802.11 a/b/g standards:** The client nodes can be supporting any of the IEEE 802.11 a/b/g standards. Although the 802.11g standard is backward compatible to both 802.11a and 802.11b standards, 802.11a and 802.11b compliant nodes are not compatible to each other and need to communicate at the lowest data rate. Compatibility to all of these three standards means that a vAP can exchange data packets with the client nodes at the highest

data rate which will improve the network performance.

- **Multiple channels:** As mentioned before, two nodes within the transmission range of each other can not be vAPs in the same channel as it will incur repetitious collisions. The vAPs can perform best if adjacent vAPs operate in different channels. Although the vAP based medium access can be used in a single channel WLAN (details explained in Section 6.6), the ability to operate in multiple channels boosts the network performance significantly.

- **Multiple network interfaces:** Multiple channels can be properly utilized only with multiple interfaces. Most mobile devices are likely to bear multiple interfaces in the near future [21] which can be utilized to improve the WLAN performance and avoid connection failures due to network outages.

- **Multi-rate MAC:** The use of the multi-rate MAC mechanism can increase the network performance as the data transmission rate can be altered depending on channel conditions. In this mechanism, the highest data rate is adaptively chosen which does not result in a significant packet error rate. As a result, the voice performance improves significantly.

- **History of success in bidding:** Recent bidding successes are strong indications that the respective node is one of the most resourceful among its competitors (i.e, other bidders). Since this parameter is a function of time, an aging mechanism should also be used so that more recent bidding successes get greater weights than earlier successes.

- **Mobility of the user:** A less mobile node can serve the collision domain better since the probability of a connection getting lost becomes lower. Nodes that are highly mobile can cause packet loss if they become vAP. We discuss the effect of mobility on our proposed mechanism in detail in Section 6.8.

- **Queue size:** A vAP needs to store and forward a large number of packets and, therefore, it should have a sufficiently large buffer in its queue. A node with a higher queue size can be given higher priority in the bidding mechanism.

- **Battery power:** A mobile device with low battery power should not be entrusted with a responsibility that requires a high number of transmissions and such a node is, therefore, less preferred as a prospective vAP.

We assign weights $w_1, w_2, \ldots, w_r$ and the corresponding values $v_1, v_2, \ldots, v_r$ to an r number of such resources where $0 < w_i < 1$ and $0 < v_i < 1$ for $1 \leqslant i \leqslant r$. Both the weights and the values are assigned in a way that allows the cumulative resources

of a vAP to bring a greater positive impact to the network performance. Since the use of a lower ζ allows a node to have a higher probability of winning the bidding, a resource will be assigned a lower weight if it can bring a greater improvement in network performance. Similarly, if a resource can have a number of levels, e.g., a node can have different number of network interfaces, a node with a higher amount of such a resource (e.g., a higher number of NIs) will use a lower value for it. Now the resourcefulness index ζ is given by

$$\zeta = \sum_{i=1}^{r} w_i v_i \quad \text{for } 0 < w_i < 1,\ 0 < v_i < 1. \tag{6.1}$$

Although the weights should not vary over time, some of the resources (e.g., bidding success history) can have varying values. Therefore, ζ might not be constant over time for a given node. Before transmitting the Bidding frame, a bidder will backoff for a random number of idle slots chosen uniform randomly between 0 and ζW. The expected length of this additional backoff is $\zeta t_{idle}\frac{W}{2}$ where t_{idle} is the length of an idle slot. However, for the rest of this chapter, we consider a homogeneous network with a uniform priority for all nodes. For such networks and in the absence of a comparative study on network resources, we assign $w_i = \frac{1}{r}$ and $v_i = 1$ for $i = 1, 2, \ldots, r$ which gives $\zeta = 1$, and a value of $t_{idle}\frac{W}{2}$ for the above mentioned backoff delay is used. An exhaustive analysis of these resources is outside the scope of this thesis and we only provide a general guideline here. A comparative study of all possible resources and the determination of w_i and v_i to improve network performance are left to future works as explained in Section 7.2.

6.6 End-to-end Communication in a vAP enabled WLAN

While forwarding all the accumulated packets, the vAP can send its own packets first followed by the packets collected from other client nodes. Considering the delay and loss in the medium access and queue as estimated later in Section 6.9, this priority in handing over the packets can improve the voice performance of the vAP considerably which is the primary incentive for a client node to become a vAP. However, a client node serves as a vAP on a voluntary basis since not every node can be expected to be well-equipped to coordinate the channel access. Therefore, a vAP might not be available in every collision domain, i.e., at each hop, of a multihop WLAN. But in dense networks and WLANs with users who tend to linger in their places for long time, e.g., in coffee shops or stadiums, some nodes that are willing to be vAPs can be expected in every collision domain. When the next hop destination node is another vAP itself coordinating access in a different channel, the data packets are transmitted

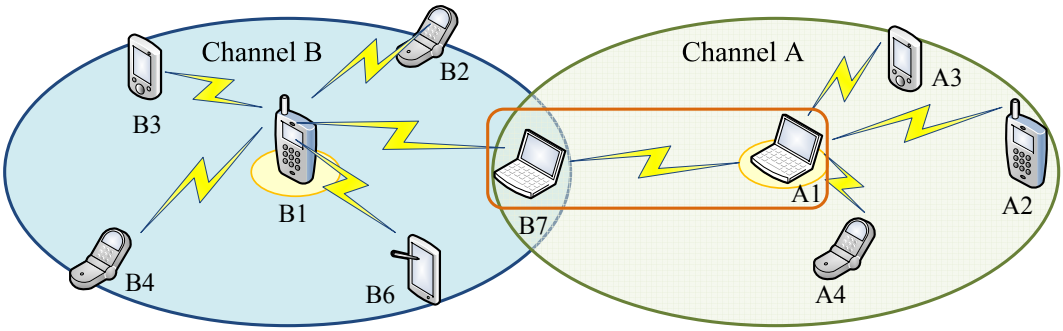


Figure 6.3: Packet transfer from a vAP (A1) to a non-adjacent vAP (B1) through a non-vAP client node (B7). Node B7 can communicate with A1 and B1 using the vAP based medium access mechanism in channels A and B, respectively.

using the vAP based medium access mechanism. On the other hand, when there is no node willing or capable to become a vAP in the WLAN, using DCF at each hop is the only available option.

But especial considerations are required when there are only a few nodes that are willing to be vAPs and in the end-to-end path from a client node to the AP, the voice packets need to be forwarded by a node that is not a vAP. We illustrate one such scenario in Fig. 6.3 where an intermediate node is either not willing or not capable of being a vAP. For instance, the intermediate node B7 in the figure is capable of operating in (and connected to) only two channels (channels A and B) and both of these channels already have a vAP each. In this case, the intermediate node (B7) collects all packets from one vAP (A1 in channel A) and hands them over to the other vAP (B1 in channel B).

However, if the intermediate node is not directly connected to another vAP and can not become a vAP for other reasons, e.g., low battery, it will use DCF to handover the packets to the next hop destination. We illustrate this scenario in Fig. 6.4 where a vAP A1 hands over the packets using the vAP based medium access to an intermediate node A2 who is not directly connected to another vAP. Therefore, node A2 hands over the packets to the next hop destination B2 using DCF in a second channel. Node B2 can, in turn, deliver the voice packets to a vAP/AP B1 (which it is connected to) using the vAP based medium access mechanism.

An especial situation can still occur when only one channel is available or the nodes have only one NI each so that all the nodes must operate in a single channel. For instance, if channel B is not available in Fig. 6.4, all nodes are required to operate in channel A. In this case, node A2 does not get much time to forward all the packets to B2 using DCF since the vAP (either A1 or B1, since two vAPs will not be selected to form a hidden node problem as explained in Section 6.4) will be using most of the

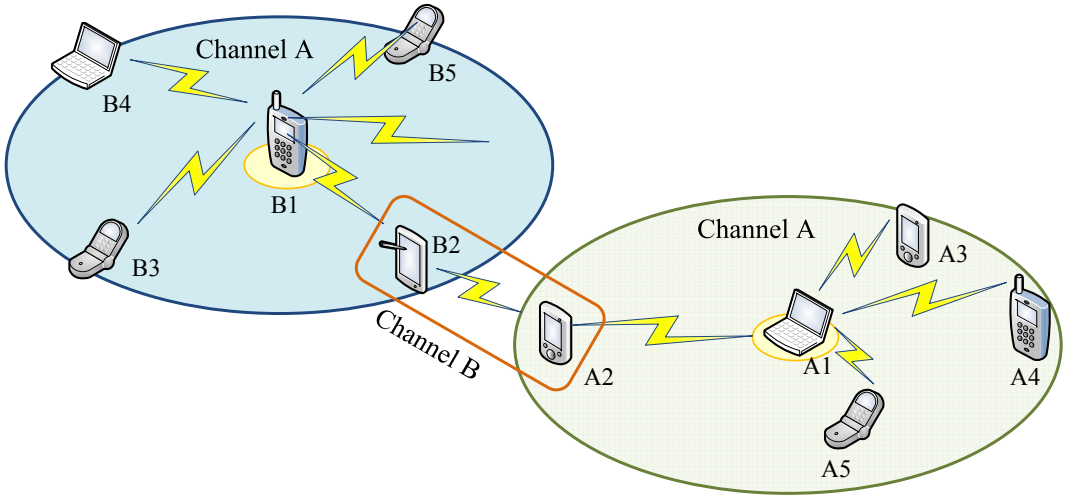


Figure 6.4: Packet transfer from a vAP (A1) to another non-adjacent vAP (B1) through two non-vAP client nodes (A2 and B2). Nodes A2 and B2 use the RTS/CTS technique to reserve the channel and a sequence of basic handshake to transfer all the accumulated packets.

channel time conducting the CFPs. In such a case, the intermediate node A2 still uses the DCF mechanism but utilizes the virtual carrier sensing technique to reserve the channel and hand over all packets to the next hop destination B2. In this mechanism, node A2 transmits a RTS frame to B2 but the RTS carries a long value as the duration of the handshake that should be sufficient to transfer all the packets in its queue. After receiving the RTS frame, all nodes within the transmission range, i.e., A1, A3, A4, A5 and B2 refrains from transmitting any frame of their own although B2 transmits a CTS giving a go ahead signal to A2. The CTS frame also carries the transaction length so that all nodes within the transmission range of B2, i.e., B1, B3, B4 and B5, avoid channel access within this time. After a SIFS following the CTS frame, A2 transmits the first packet which B2 acknowledges after a SIFS. A2 waits for another SIFS and transmits its second packet which B2 acknowledges after another SIFS. In this way, all the packets that are queued at A2 are handed over to the next hop destination B2 using a sequence of the basic DCF handshake mechanism. Even if another client node now moves into the scenario, it can not interfere with these sequential transmissions since A2 is waiting for only one SIFS in between the sequential transmissions of the packets (we recall from Section 2.3.2.1 that a node must wait for a DIFS period before accessing the channel while using DCF and the length of a DIFS is longer than that of a SIFS). All other nodes also remain silent because their NAV is non-zero for the whole time and also because the channel is never free for more than a SIFS.

6.7 Fairness Measure at Each Hop

As the intermediate vAPs forward packets for all client nodes and vAPs located in the outer rings, the effective packet arrival rate increases for vAPs in the rings nearer to the AP. In such a ring, polling one client node in the collision domain once in every CFP becomes unfair to the nodes in the outer rings because the end-to-end delay becomes very high for such nodes. This is a common problem also with the DCF based medium access mechanism. Therefore, we use an additional measure to make the proposed mechanism fair to all voice streams at each hop. We propose that the client nodes in each collision domain will report the number of voice packets in its queue using an unused header field of the 802.11 MAC header. For instance, in the 802.11 standards, the SERVICE field of the MAC header is left unused which can be used for this purpose (possibly after being renamed). When the vAP/AP polls a client node, the client node puts the number of voice packets left in its queue in the designated field of the uplink data packet. The vAP/AP will poll the same node until the queue becomes empty. We additionally propose to include this technique in the original PCF mechanism which is to be carried out by a regular AP. This mechanism is very simple to implement but will ensure a fair share of the channel resource to be allocated to each voice stream. It also ensures that a vAP, which is likely to store and forward a higher number of packets than other regular client nodes, is granted longer access to the channel to empty its queue quickly. The regular client nodes which are not vAPs, are likely to send their own packets only. Therefore, the regular clients will be polled only once while the nodes which are vAPs in other channels or forwarding a higher number of packets will be polled a higher number of times.

6.8 Effect of Node Mobility on the vAP based Medium Access

Using the client nodes as virtual access points has one down side, that is, the client nodes are expected to be mobile and selecting a frequently moving node as a vAP will jeopardize its purpose. The vAP selection mechanism, requiring the reconnoiter and bidding processes, incurs a delay overhead. Although the use of time synchronized medium access reduces the channel access delay, the delay overhead in the vAP selection mechanism has to be traded off against the benefits. In this regard, the maximum tolerable level of mobility for a user should be carefully derived to prevent a frequently moving node from becoming a vAP.

To analyze the impact of mobility on our proposed scheme, we use a statistical study [4] on real human walking pattern. Rhee *et al.* [4] investigated the mobility

pattern of human by analyzing mobility track logs of 44 human participants carrying GPS receivers from September 2006 to January 2007. The samples were taken from North Carolina State University (NCSU), Korea Advanced Institute of Science and Technology (KAIST), New York city (NYC) and Disney world (Disney). Despite the low number of participants, the total duration of all walking together is over 1000 hours which makes the findings statistically significant. Human movement is a sequence of pause and walk states and a mobile device attached to a human being is like a static WLAN node during a pause state. It was identified in [4] that the truncated Pareto distribution exhibits the best fit for the pause time distribution of all the collected data than other most likely distributions. Following the observations and findings in [4], we use the same truncated Pareto distributions to determine the pause time of a mobile device that a human is carrying.

A truncated Pareto distribution is determined by three parameters: α is the shape parameter while L and H denote the minimum and maximum values, respectively. The distribution is bounded or truncated by L and H, and the probability density function of a random variable x is given by

$$P(L \leqslant x \leqslant H) = \frac{\alpha L^{\alpha} x^{-\alpha-1}}{1 - \left(\frac{L}{H}\right)^{\alpha}}. \tag{6.2}$$

When a node moves from its current location, connectivity might be hampered with some of the nodes in its old collision domain and in order to coordinate time synchronized channel access, the vAP should not move before it completes the ongoing CFP. Therefore, the pause time length t_{pause} of a vAP should be greater than the super frame length t_{sf}. Using the cumulative distribution function of a truncated Pareto distribution, we represent this condition as

$$P\left(t_{pause} > t_{sf}\right) = \frac{L^{\alpha} {t_{sf}}^{-\alpha} - H^{-\alpha}}{1 - \left(\frac{L}{H}\right)^{\alpha}}, \tag{6.3}$$

where $0 \leqslant L \leqslant t_{sf} \leqslant H$ and $0 < \alpha$. The associated values for this pause time distribution for the real world scenarios from [4] are presented in Table 6.1. The resulting requirement on the super-frame length is shown in Fig. 6.5.

We further use this real world data to determine the number of nodes that can be polled during a "pause" time of the user. We use the G.729 codec with an aggregation level of 2 in an 11 Mbps 802.11b WLAN and determine the number of nodes that can be polled for two way voice connections. During each pause of the user, the user's mobile device bids and then transmits a Beacon and coordinates time synchronized channel access as outlined in Sections 6.3 and 6.4. The number of nodes that can be polled for two way voice calls, i.e., where each call requires one uplink and one downlink frame to be transmitted in each CFP, is illustrated in Fig. 6.6 using the

Table 6.1: Parameters for the truncated Pareto distributions which represent the pause time distribution of real human mobility [4].

Scenario	α (shape)	L (lower limit)	H (upper limit)
NCSU	0.99	0.1	430
KAIST	0.45	0.1	270
NYC	0.49	0.1	670
Disney	0.80	0.1	150

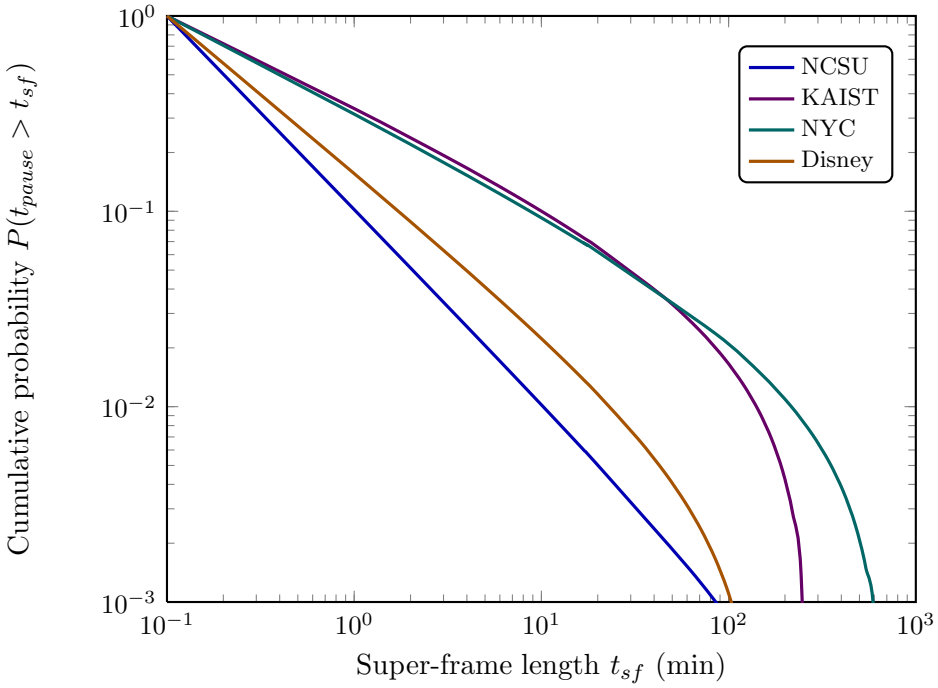


Figure 6.5: Pause time distribution for human mobility [4] illustrated on a log-log axis showing the probability of a pause time length to be longer than a given super-frame length.

real world pause time distributions presented in [4]. The results suggest that a high number of nodes can be served despite the mobility of the wireless nodes. However, we propose that a node should not bid whenever

$$P\left(t_{pause} > t_{sf}\right) < 0.9, \tag{6.4}$$

so that a vAP can ensure that all the packets are properly forwarded and its mobility does not degrade the performance of the network. Figure 6.6 dictates that in all of the above mentioned real world scenarios (Table 6.1), about 1.128×10^4 nodes can be polled during a pause time with probability 1.0.

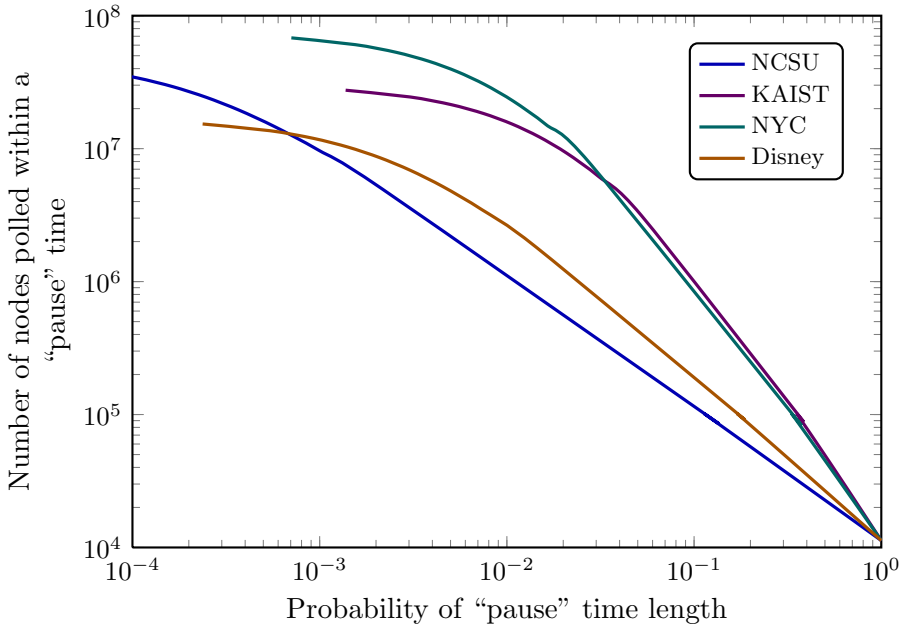


Figure 6.6: Number of client nodes that can be polled within a "pause" time with a given probability determined using real human mobility data [4] illustrated on a log-log axis. The G.729 codec with an aggregation level of 2 is used and one two-way voice call is initiated by each client node in this 11 Mbps 802.11b WLAN.

6.9 Delay and Loss in Medium Access and Queue at Each Hop with vAPs

For easier analysis, we split the WLAN into ϕ concentric annulus regions or rings as in Chapter 4. Each ring is ω meters wide which can be determined considering coverage and interference. Most of the analyses for multihop WLAN that are explained in Chapter 4 also hold for a WLAN with vAPs. Therefore, we only analyze network conditions where additional considerations are required.

The WLAN has a total of n_s nodes and each node initiates n_v number of two-way voice calls. The total area Λ_i and node density Ω_i of the i-th ring are given by $\Lambda_i = \pi\omega^2(2i-1)$ and $\Omega_i = \frac{n_s}{\pi\omega^2\phi^2}$ (for uniform node distribution), respectively. Although this node density is applicable to uniform node distribution only, for other node distributions like exponential or normal, a similar analysis holds as demonstrated in Chapter 4. The vAPs in the ring i need to forward all packets generated by the $\sum_{j=i}^{\phi}\Lambda_j\Omega_j$ nodes that are located in all outer rings $j = i, i+1, \ldots, \phi$.

With a transmission range of r_t, there are $\pi r_t^2\Omega_i$ nodes in a collision domain in the i-th ring. Each of the nodes has an effective packet arrival rate of $\frac{2n_v}{n_a d_f \Lambda_i \Omega_i}\sum_{j=i}^{\phi}\Lambda_j\Omega_j$

packet/sec. However, when n_c channels are utilized simultaneously, a vAP operating in the i-th ring needs to execute η_i transactions where η_i is defined as

$$\eta_i = \frac{\pi r_t^2 n_v}{n_a d_f n_c \Lambda_i} \sum_{j=i}^{\phi} \Lambda_j \Omega_j. \tag{6.5}$$

The length of t_{recce} is given by

$$t_{recce} = t_m + \ t_{difs}. \tag{6.6}$$

The bidding takes t_{bid} time to complete which is the sum of the random waiting time and the time taken by transmissions of the Bidding and acknowledgment frames

$$t_{bid} = \zeta t_{idle} \frac{W}{2} + t_{dummy} + t_{sifs} + t_a, \tag{6.7}$$

where t_{sifs} and t_{idle} are the lengths of a SIFS and an idle slot, respectively. Similarly, the transmission times of a dummy frame and an acknowledgment frame are t_{dummy} and t_a, respectively.

Using the Markov model for the PCF mechanism introduced in Chapter 5, we determine the length t_t of a transaction and derive the length $t_{sf}(i)$ of a super-frame in ring i as

$$t_{sf}(i) = t_{recce} + t_{bid} + t_{sifs} + t_b + \eta_i t_t + t_{sifs} + t_{ce}, \tag{6.8}$$

where t_b and t_{ce} are the transmission times of a beacon frame and a CF-End frame, respectively, and t_{recce} can be eliminated once it is determined that there is no AP in the associated collision domain, i.e., in the later super-frames as long as the node does not move. Using the analysis for channel access delay with the PCF based medium access mechanism presented in Section 5.4, the delay $d_c(i)$ and loss $e_c(i)$ in medium access in ring i are defined as

$$d_c(i) = \begin{cases} \frac{t_{sf}(i)}{2} & \text{for uplink packets,} \\ \frac{t_{sf}(i)}{\eta_i} & \text{for downlink packets,} \end{cases} \tag{6.9a}$$

$$e_c(i) = \mathbb{P}_r^{m+1}. \tag{6.9b}$$

The transitional probability $\mathbb{P}_r$ was defined in (5.3) in Chapter 5.

The vAPs are regular, and preferably, multi-interface nodes, and the delay and loss in the queue can be modeled using the multi-server queuing system explained in Section 4.4. For the single-NI vAP scenario presented in Section 6.6, the single server queuing system explained in Section 3.7 can be used. Using an appropriate queuing model, we can now derive the delay $d_q(i)$ and loss $e_q(i)$ in the queue of a vAP node in the ring i.

6.10 End-to-end Delay and Loss over Multiple Hops

As explained in Section 4.5, the end-to-end delay and loss for the packets generated in the ring ϕ is greater than those for packets generated in any inner ring. Therefore, testing the total voice quality impairment due to the end-to-end delay and loss for a packet generated in the ring ϕ (outer most ring) suffices the same for the whole network. The cumulative delay Δ_i and loss Γ_i faced by a packet generated in the ring ϕ as the packet hops through the rings $\phi, \phi-1, \ldots, i$ (data is forwarded by single hop, i.e., one ring at a time) can be determined by using the additive and multiplicative nature of delay and loss, respectively, in the medium access and queue. We redefine the end-to-end delay and loss (previously defined in (4.12) and (4.13), respectively) as experienced by a packet that has been generated by a node in the ring ϕ while the packet traverses the ring i as

$$\Delta_i = \sum_{j=i}^{\phi} \big(d_c(j) + d_q(j) \big) \quad \text{and} \tag{6.10a}$$

$$\Gamma_i = \begin{cases} e_q(\phi) + (1 - e_q(\phi))\, e_c(\phi) & \text{for } i = \phi, \\ \Gamma_{i+1} + (1 - \Gamma_{i+1})\, e_q(i) + e_c(i) & \\ \quad - (\Gamma_{i+1} + (1 - \Gamma_{i+1})\, e_q(i))\, e_c(i) & \text{otherwise.} \end{cases} \tag{6.10b}$$

The end-to-end delay Δ_i and loss Γ_i for a packet generated in the ring ϕ attain their maximum values in the first ring, i.e., when $i = 1$. Using (3.17) and (3.20), we determine the maximum end-to-end delay d'_e and loss e'_e experienced by voice packets in the WLAN as

$$d'_e = d_l + n_a d_f + \Delta_1 + d_j, \tag{6.11a}$$

$$e'_e = \Gamma_1 + (1 - \Gamma_1) e_j. \tag{6.11b}$$

6.11 Capacity Model Employing Virtual Access Points

Characterization of the formation of critical zones is slightly different for the vAP enabled WLANs. Although the expected channel access delay for the uplink packets is $\frac{t_{sf}(1)}{2}$ in the first hop, a total of $2\eta_1$ packets are served in each super-frame. Therefore, the service rate of the collision domain is much higher (compared to that in a DCF based WLAN) and we redefine the requirement in (4.14) to avoid a bottleneck due to the formation of a critical zone as

$$n_c n_a d_f \eta_1 \geqslant n_s n_v t_{sf}(1). \tag{6.12}$$

Since the number of VoIP calls is given by $n_s n_v$, we present the voice capacity of the multichannel, multihop WLAN with vAPs using the following optimization problem.

$$\begin{aligned} Max \quad & n_s n_v \\ s.t. \quad & I_d\left(d'_e\right) + I_{e\_eff}\left(e'_e\right) \leqslant I_b, \text{ and} \\ & n_c n_a d_f \eta_1 \geqslant n_s n_v t_{sf}(1). \end{aligned} \tag{6.13}$$

The above formulation maximizes the total number of calls while maintaining the voice quality impairments within their total budget (3.16) so that voice quality does not degrade, and ensures that the traffic load in any collision domain remains within an acceptable range.

6.12 Results and Analyses

Our proposed mechanism only adds an ad hoc channel acquisition procedure to allow a client node in an outer ring to coordinate the time synchronized channel access without any supervision of an AP in a multihop WLAN. Therefore, most analyses for multihop WLANs hold also in a vAP enabled WLAN if we consider the modified delay and loss in the channel access as per (6.9a) and (6.9b). The delay and loss in the channel access can be easily determined using the Markov model for the PCF based medium access introduced in Chapter 5. The use of vAPs can improve voice performance significantly in a multihop, multi-channel WLAN by utilizing the time synchronized medium access over multiple hops a multihop WLAN. We present some numerical results in this section that signifies the performance improvement. Although, only voice traffic is considered, our proposed mechanism offers similar improvement for other traffic categories as well.

6.12.1 Performance Improvement with vAP based Medium Access

The use of vAPs reduces the delay and loss at each hop and, as a result, the end-to-end delay and loss for voice packets are reduced leading to a better voice quality. The increase in voice quality ultimately leads to a higher voice capacity. In the previous chapter, we showed that the simulation results closely follow the theoretical results derived using our Markov model for the PCF based medium access mechanism. We used the same Markov model and a PCF-like mechanism for the vAP based medium access in this chapter (the difference being the modified length of the contention period due to the reconnoiter and bidding procedures). Therefore, similar match between numerical results from theoretical analyses and simulations can be expected. Using the analysis presented in this chapter, we discuss some numerical results in the following sections to highlight the performance improvement using vAP based medium access.

6.12.1.1 Channel Access Delay d_c

Channel access delay for the DCF, PCF and vAP based medium access mechanisms are shown in Fig. 6.7. Packet error rates of 0.0 and 0.2 are considered for all three access mechanisms although a similar trend of results was identified with other packet error rates as well. To emphasize the performance improvement, we consider a wide range of network sizes $n_s = 1 \sim 50$ and generate voice calls using the G.729 codec at an aggregation level of $n_a = 1$. As the number of stations in a collision domain increases, the channel access delay for the uplink traffic increases almost linearly with the vAP and PCF mechanisms. On the other hand, the downlink delay is much lower than the uplink delay with the vAP and PCF mechanisms. This downlink delay in the vAP mechanism is also lower than the channel access delay with the DCF mechanism (the channel access delay with DCF is the same for uplink and downlink traffic). As the number of nodes increases, the number of transactions in a super-frame increases and the delay overhead per packet due to the waiting time in bidding and reconnoiter periods reduces. As a result, the downlink channel access delay converges to $2(t_d+t_{sifs})$ for a high η_i. The medium access delay for the uplink packets in the vAP mechanism is slightly higher than the PCF mechanism (uplink traffic) due to the overhead of bidding and reconnoiter mechanisms. However, with a higher packet error rate, a packet is retransmitted a higher number of times resulting in an increased channel access delay. The channel access delay as seen by a HoQ packet in the DCF, PCF (uplink) vAP (uplink) and vAP (downlink) scenarios are 4.629, 13.084, 13.648 and 6.984 ms, respectively, for $n_s = 4$ and $p_e = 0.0$. With $n_s = 20$, these delay elements become 15.774, 26.3, 26.864 and 2.846 ms, respectively. The channel access delay with DCF was determined using the Markov model introduced in Section 3.8.

6.12.1.2 Channel Access Loss e_c

Channel access loss with the vAP based medium access mechanism is the same as that with PCF. We illustrate the channel access loss with retry limits of 3, 5 and 7 using both the DCF and vAP based medium access in Fig. 6.8 for different packet error rates. With DCF, packet errors or collisions without capture or both result in transmission failures which increase packet loss as shown in Fig. 6.8. On the other hand, a transaction in the PCF based medium access can fail if either the uplink or the downlink transmission fails and, therefore, the packet loss becomes high, especially with a high packet error rate. For instance, for packet error rates higher than 0.4 with $m = 5$ (similarly for p_e greater than 0.35 with $m = 7$), the vAP based medium access results in a higher packet loss than DCF. Although not shown in Fig 6.8, a similar trend is observed for lower retry limits at much higher packet error rates. But this

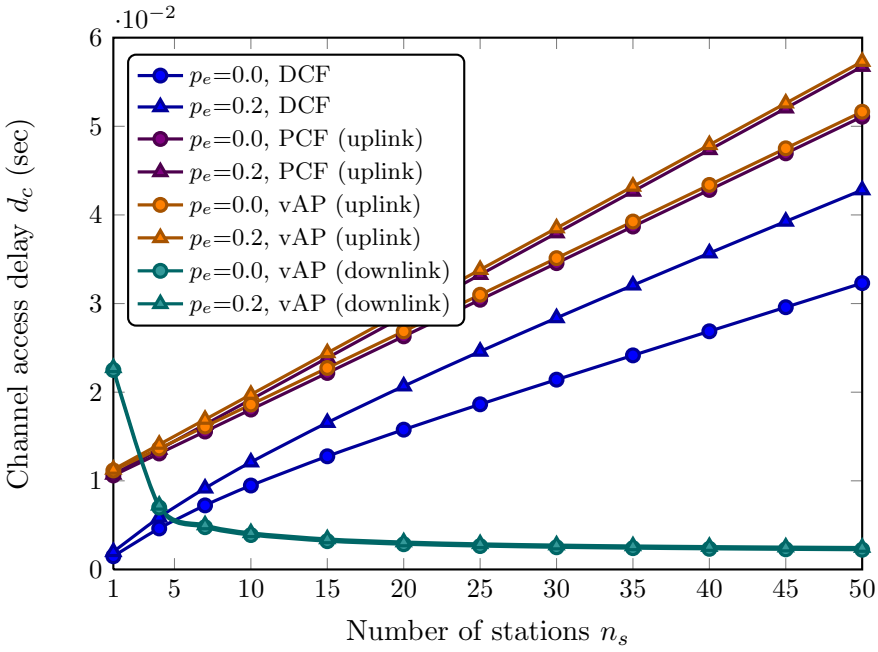


Figure 6.7: The channel access delay with the DCF, PCF and vAP based medium access for different network sizes and packet error rates using the G.729 codec at an aggregation level of 1 in an 802.11b WLAN with 1 Mbps data rate.

phenomenon seldom affects the voice capacity since in the operable condition with real channels, the acceptable packet error rate is below this level. For instance, experiments in real wireless networks conducted in [161] reported packet error rates of less than 0.3 for office or lab environments. Additionally, the use of the multi-rate medium access mechanism can be utilized to keep the packet error rate to a minimum. A packet error rate of $0.05 \sim 0.25$ is shown to have achieved in [154] using Adaptive Multi Rate Retry (AMRR) feature of the multiband *Atheros* driver *madwifi*[1] for different traffic load. For a packet error rate less than or equal to 0.3, the vAP based medium access results in lower channel access loss than DCF for all retry limits. At $p_e = 0.3$, the channel access loss with the vAP (and PCF) based medium access are 0.039, 0.02 and 0.01 for a retry limit of 3, 5 and 7, respectively. For the same packet error rate with the DCF based medium access, the channel access loss becomes 0.13, 0.047 and 0.016, respectively. From the expression for channel access loss in (6.9b) for the vAP based medium access, we note that channel access loss is independent of the number of competing nodes which has a severe degrading effect in DCF. As collisions can not occur with the PCF based medium access mechanism, transmissions can fail only due to channel errors. With the DCF based medium access, however, transmissions can

[1]http://www.madwifi.org

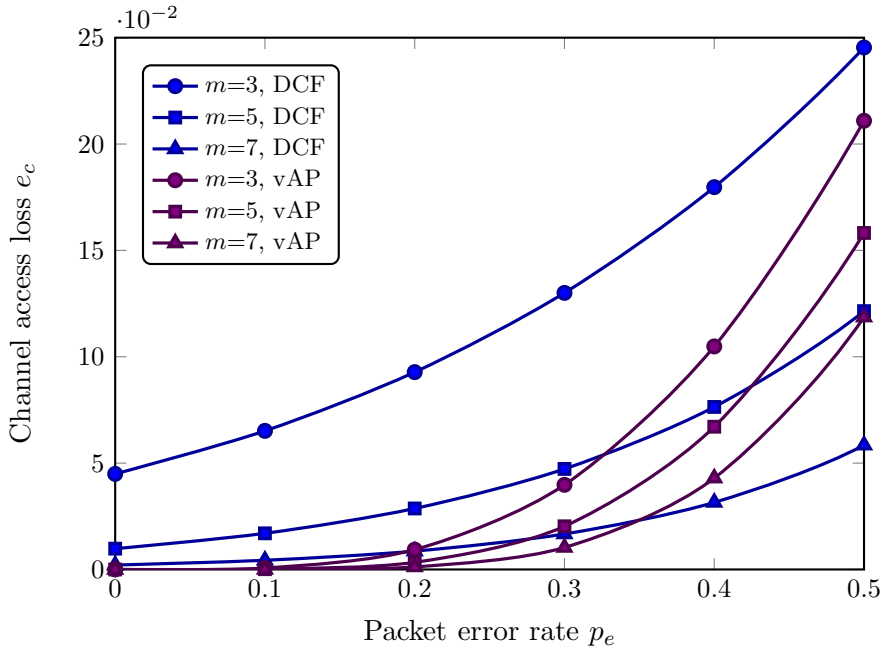


Figure 6.8: The channel access loss with the DCF and vAP based medium access mechanisms for different packet error rates and retry limits using the G.729 codec at an aggregation level of 1 in a 50-node 802.11b WLAN with 1 Mbps data rate showing higher channel access loss for higher packet error rates and lower retry limits. Also shows the channel access loss with the vAP based medium access mechanism is lower than that with DCF for $p_e \leqslant 0.35$.

fail due to either of channel errors and collisions (i.e., the collisions without captures). The probability of collision (3.33) in the DCF mechanism increases substantially with the number of competing nodes which incur a high packet loss in the channel access.

6.12.1.3 Queuing Delay d_q

The lower channel access delay and loss offer a higher service rate for the packets in the queue. Therefore, the queuing delay is much shorter in the vAP based medium access than DCF and much like in PCF. As shown in Fig. 6.9, queuing delay with the vAP based medium access is only slightly higher compared to PCF due to the slightly higher channel access delay. This delay for the vAP and PCF based medium access mechanisms are found to increase at a higher rate when the network size is between 10 and 15. For instance, the queuing delay with the DCF, PCF and vAP based medium access are 453.428, 80.409 and 111.136 ms, respectively, for $n_s = 10$ with $p_e = 0.0$ which become 617.35, 444.23 and 474.77 ms, respectively, for $n_s = 15$. For $n_s > 15$, queuing delay increases at a much slower rate.

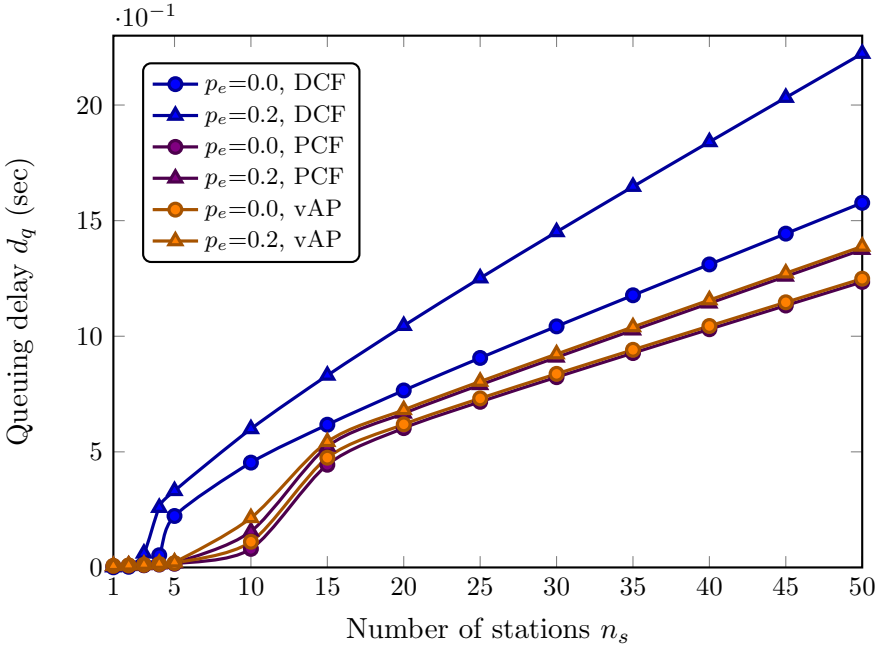


Figure 6.9: The queuing delay with the DCF, PCF and vAP based medium access mechanisms for different network sizes and packet error rates using the G.729 codec at an aggregation level of 1 in an 802.11b WLAN with 1 Mbps data rate.

6.12.1.4 Queuing Loss e_q

Due to the lower delay in channel access for the vAP based medium access, the queue remains less prone to becoming full and the queuing loss decreases. But as the network size increases, frequent collisions in the DCF based medium access give rise to higher number of retransmissions. As a result, the packets keep occupying the queue and queuing loss increases as illustrated in Fig. 6.10. The rate of increase in the queuing loss with the DCF mechanism increases sharply when the network size is increased from 4 to 15 nodes. However, in the PCF and vAP based medium access, queuing loss starts to increase significantly for $n_s > 10$ but still maintain a much lower loss compared to DCF. The queuing loss with the DCF, PCF and vAP mechanisms for the 5-node WLAN with $p_e = 0.0$ are found to be 0.099, 0.0 and 0.0 which become 0.472, 0.001 and 0.002 for $n_s = 10$ and 0.608, 0.098 and 0.12 for $n_s = 15$, respectively. In the 50-node WLAN, the queuing loss for these medium access mechanisms become 0.845, 0.608 and 0.612, respectively. Although the collision domain size dictates the queuing loss largely, a higher packet error rate also increases the queuing loss significantly, especially with DCF. With a packet error rate of 0.2 in the 50-node WLAN, the channel access loss become 0.89, 0.64 and 0.65, respectively.

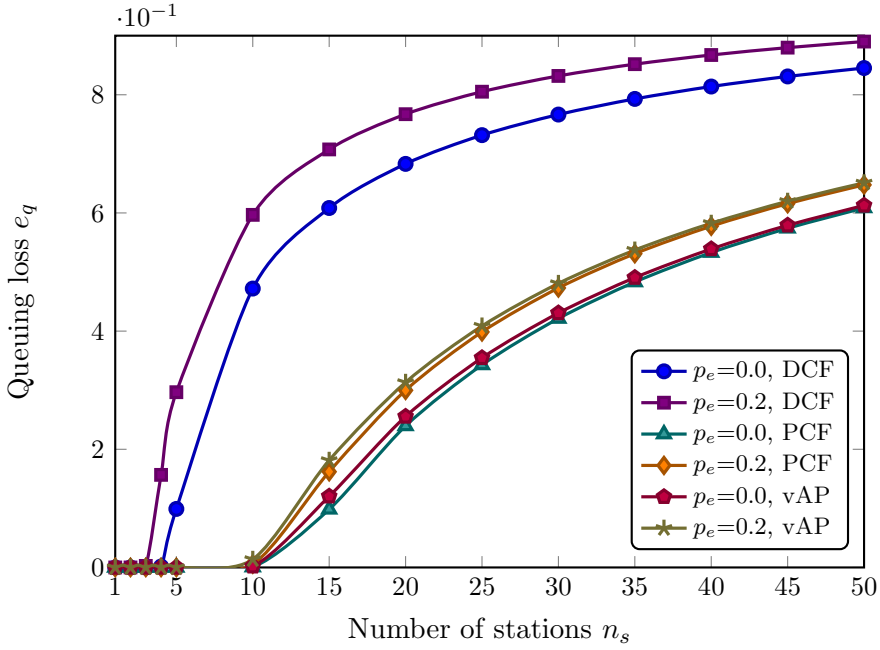


Figure 6.10: The queuing loss with the DCF, PCF and vAP based medium access mechanisms for different network sizes and packet error rates using the G.729 codec at an aggregation level of 1 in an 802.11b WLAN with 1 Mbps data rate.

6.12.2 Maximum End-to-end Delay Δ_1 and Loss Γ_1

The primary benefit by using the vAP based medium access is attained from the lower end-to-end delay and loss. End-to-end delay Δ_1 is illustrated in Fig. 6.11 for the G.729, G.711 and G.723.1 codecs. We use $n_a = 10$ for G.729 and G.711 and $n_a = 3$ for G.723.1 (since G.729 and G.711 attain their maximum call capacity at $n_a = 10 \sim 15$ while G.723.1 attains its maximum call capacity at $n_a = 3$). A data rate of 54 Mbps was used to transmit the voice packets. To demonstrate the effect of the maximum number of hops, we measure performance in 10 WLANs, each with its nodes spread over different number of hops, starting from 1 to 10. We also consider the availability of multiple channels for the time synchronized medium access and present results for $n_c = 1$ and $n_c = 3$. When the nodes are spread over a WLAN with a higher number of rings (or hops) ϕ, the total area $\pi\omega^2\phi^2$ of the WLAN increases and the WLAN can offer a larger coverage with one access point. However, since the voice packets face cumulative delay and loss in the channel access and queue at each hop, extra overhead in terms of delay and loss occur in the end-to-end path. For instance, the maximum end-to-end delay for $n_c = 1$ and using the G.729 codec are 25.37, 44.018 and 56.517 ms when the maximum number of hops ϕ are 1, 5 and 10, respectively. Using the ITU-T E-model, we find that the delay impairment caused by these delay are 0.976, 1.455 and

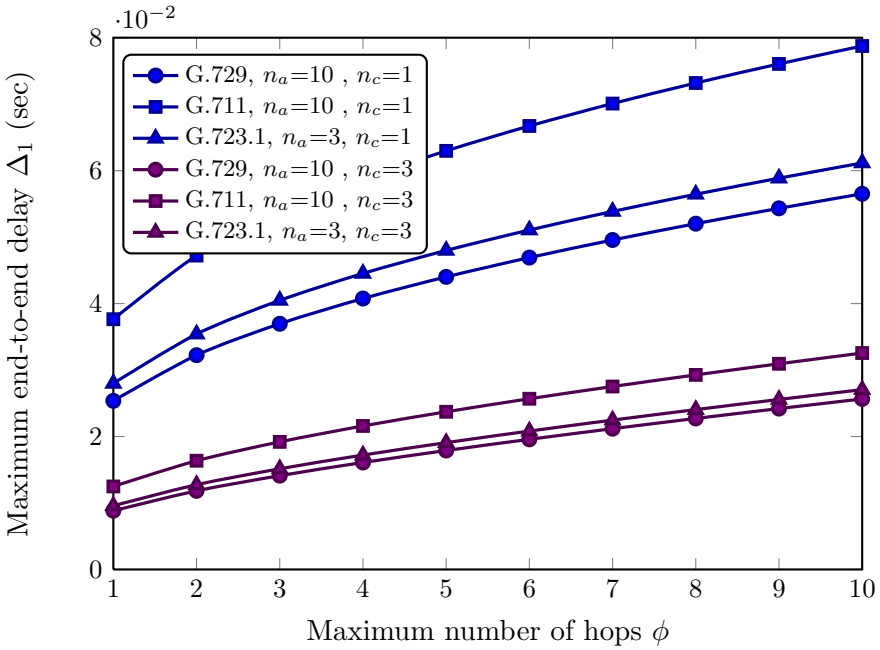


Figure 6.11: The maximum end-to-end delay with virtual access points for different number of hops and the G.729, G.711 and G.723.1 codecs using n_c =1 and 3 in a WLAN with 54 Mbps data transmission rate.

1.751, respectively. These delay using the G.711 codec are 37.651, 62.976 and 78.753 ms which incur delay impairment factors 1.2981, 1.8989 and 2.2465, respectively.

Similarly, the maximum end-to-end packet loss Γ_1 for $\phi = 1, 2, \ldots, 10$ are illustrated in Fig. 6.12 for G.729 codec. Packet loss in the medium access and queue accumulate over multiple hops to give the end-to-end delay and loss. We consider three packet error rates of 0.1, 0.2 and 0.3 and two retry limits of 3 and 5 to demonstrate the end-to-end loss Γ_1 incurred in the end-to-end path. For these results, G.729 codec was used with $n_a = 2$ in a 54 Mbps 802.11b multihop WLAN. Packet error rate has a severe degrading effect on the voice quality as it incurs a high end-to-end loss although it can be reduced to some extent by using a higher retry limit. For the 10-hop WLAN in Fig. 6.12, the maximum end-to-end packet loss are found to be 0.6476, 0.7124 and 0.8242 when retry limit is 3 and the packet error rates are 0.1, 0.2 and 0.3, respectively. Using a retry limit of 5 reduces these end-to-end packet loss to 0.6438, 0.6936 and 0.7847, respectively. According to E-model, the equipment impairment factor for the G.729 codec in these cases are 75.8606, 76.9394 and 78.6272, respectively. Since the total impairment budget (of delay and loss impairments) are 28.3553 and 13.1952 for medium and high quality calls, respectively, the impact of such high end-to-end loss can degrade voice quality severely and should be considered carefully in network design and planning.

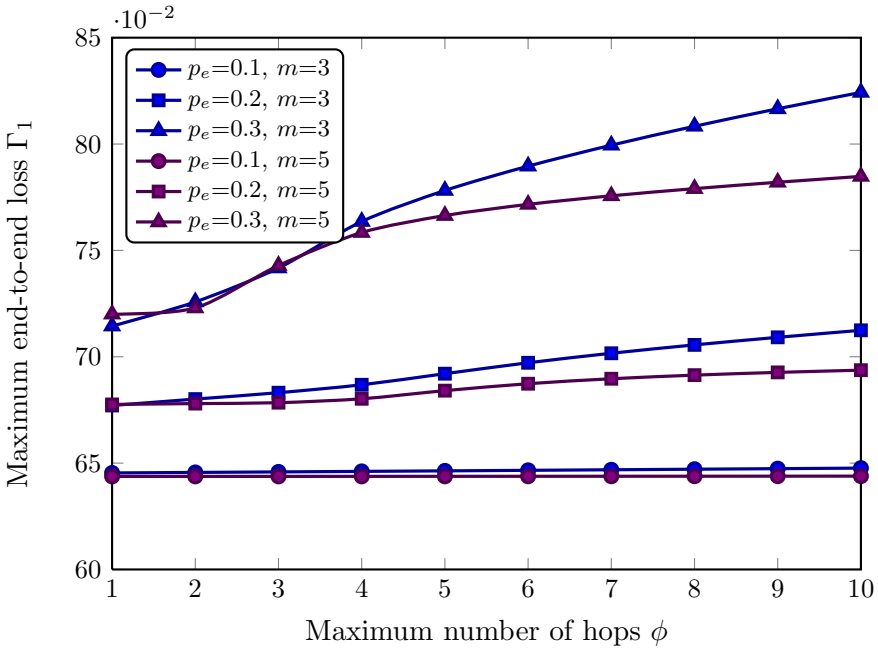


Figure 6.12: The maximum end-to-end loss with virtual access points for different number of hops, packet error rates and retry limits using the G.729 codec at an aggregation level of 2 in a 802.11b WLAN with 54 Mbps data transmission rate.

6.12.3 Voice Capacity of a vAP enabled WLAN

Most analyses for multihop scenarios as presented in Chapter 4 hold for vAP enabled WLANs, i.e., voice capacity is affected by different parameters in a similar manner. Therefore, we re-investigate only a few aspects and focus on the performance comparison of the DCF and vAP based medium access in the multihop WLANs. Figure 6.13 illustrates the voice capacity for the G.729 codec at different aggregation levels and a data rate of 11 Mbps. We considered three WLANs with 1, 2 and 3 number of hops where nodes are distributed uniformly. We note that vAPs are not used when the WLAN consists of only one hop; therefore, the $\phi = 1$ scenario indicates performance of a single collision domain as in PCF and is presented for comparison-only purposes. Voice capacity increases with the number of hops in these scenarios which is primarily due to smaller collision domains at each hop. When client nodes are spread over a WLAN with a higher number of hops ϕ with the same n_s and ω, the nodes are located more dispersedly and the collision domain size at each hop becomes smaller. Since fewer nodes are polled at each hop by a vAP, the delay in the channel access and queue also become shorter. Voice capacity with the vAP based medium access is found to be much higher in multihop WLANs compared to the DCF based medium access. For instance, the maximum voice capacity are identified as 148, 229 and 200 for the vAP

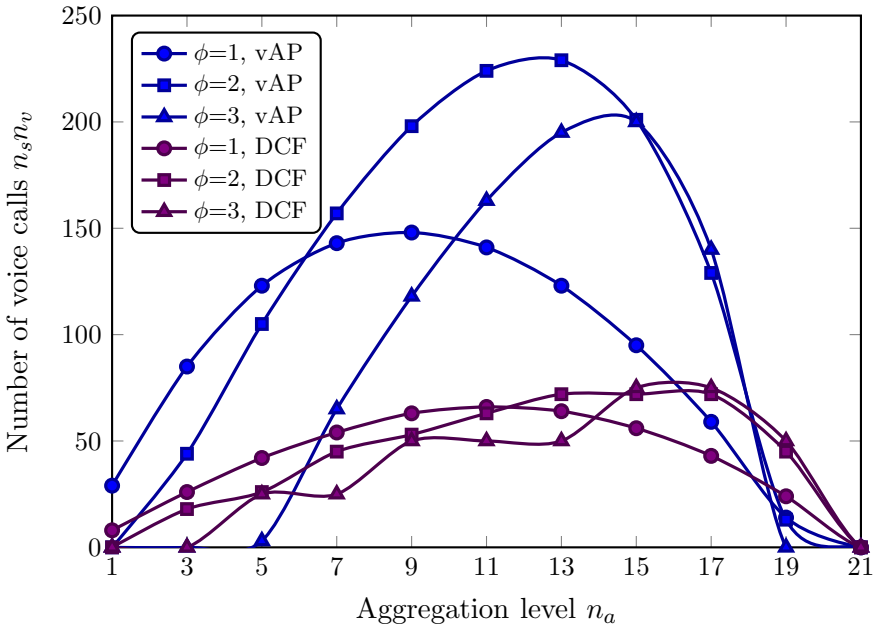


Figure 6.13: The VoIP call capacity for the G.729 codec at different aggregation levels in vAP and DCF based 802.11b WLANs with 11 Mbps data rate and 1, 2 or 3 hops illustrating initial increase in voice capacity with higher number of hops and aggregation levels although further increase in the number of hops and aggregation level decrease call capacity.

based medium access and 66, 72 and 75 for DCF when ϕ is 1, 2 and 3, respectively. As explained before, the primary disadvantage with DCF is the channel time wastage due to collisions and the backoff mechanism while the vAP and PCF based medium access offer lower channel time usage per packet and result in lower delay and loss also in the queue. However, for a considerably high aggregation level, the packet arrival rate becomes very low and nodes tend to remain unsaturated for a considerable amount of time. In such a case, the overhead with the DCF based medium access is reduced since the probability of collisions also remains low. But in the PCF or vAP based medium access, each node still has to be polled because the AP is not aware of the nodes which do not have any packets to send. This is why, the call capacity in DCF outperforms that in the vAP based medium access for such high aggregation levels. In our example scenario, the call capacity with $n_a = 19$ for ϕ=1, 2 and 3 using the vAP mechanism are 14, 13 and 0 which are 24, 45 and 50 with the DCF mechanism, respectively. However, the high delay impairment caused by a too long packetization interval for $n_a \geqslant 21$ prevents both medium access mechanisms from supporting any voice call. For aggregation levels less than 19, the vAP based medium access offers a higher VoIP call capacity than DCF. In fact, we are mostly interested in the region

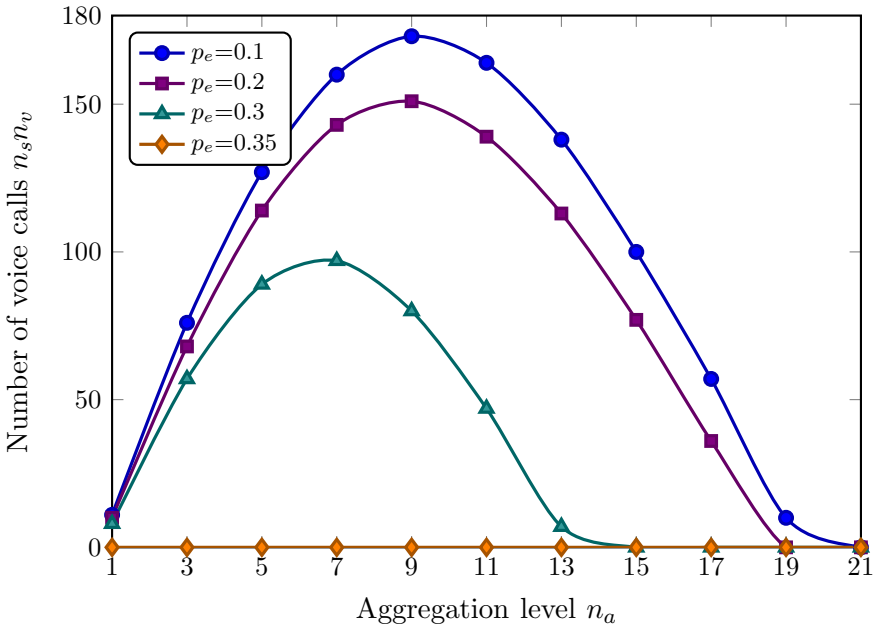


Figure 6.14: The voice capacity for the G.729 codec at different aggregation levels and packet error rates in a vAP based two-hop 802.11b WLAN with 11 Mbps data rate illustrating the decrease in call capacity with increasing packet error rate.

$n_a = 9 \sim 17$ where the G.729 codec attains its maximum call capacity. In this region, the vAP based medium access mechanism outperforms DCF and offers considerably higher call capacity.

We present the voice capacity with imperfect channels in Fig. 6.14 for a vAP based two-hop WLAN where the codec and data rate remain the same as before. With increasing packet error rate, the call capacity decreases sharply since the success of a transaction requires two consecutive successful packet transmissions in each way (uplink and downlink) and results in a higher packet loss. The maximum call capacity are determined to be 173, 151 and 97 when the packet error rates are 0.1, 0.2 and 0.3, respectively. It is identified that for $p_e \geqslant 0.35$, the end-to-end loss becomes very high incurring a high effective equipment impairment factor for G.729, and no voice call can be supported.

The effect of the maximum number of hops in a multihop WLAN is presented in Fig. 6.15 for three aggregation levels using the G.729 codec in an 11 Mbps WLAN. Both the DCF and vAP based medium access mechanisms show a similar extent of coverage, i.e., no calls are supported for $\phi \geqslant 5$, $\phi \geqslant 6$ and $\phi \geqslant 8$ when n_a is 4, 6 and 10, respectively. The benefits from using multihop WLANs are two folds: firstly, a wider area can be covered by a single AP and the operational costs are reduced; secondly, the nodes are spread over a larger area which produces smaller collision

domains at each hop which can increase the voice capacity and improve the network performance. With the decreasing trend of the WLAN device cost, the latter benefit of having smaller collision domains is of higher importance. Figure 6.15 demonstrates that for both medium access mechanisms, the maximum call capacity are attained within $\phi = 2 \sim 4$. For this range of the maximum number of hops, the vAP based medium access offers much higher call capacity than DCF. The vAP based WLAN supports maximum of 107, 151 and 183 calls when the aggregation level n_a are 4, 6 and 10 while the DCF based WLAN supports 35, 49 and 68 calls, respectively. As mentioned before, for small collision domains with low packet arrival rate, the DCF based medium access performs slightly better than vAP due to the lower number of collisions and decreased backoff time. For instance, the vAP based medium access shows a significantly higher call capacity than DCF when n_a is 4 or 6 for $\phi = 1 \sim 8$ but with an aggregation level of 10, the packet arrival rate becomes low since every node generates only 10 packets/sec for each call. Additionally, when the client nodes are spread over a wider WLAN, i.e., with higher number of hops, each collision domain is occupied by only a few nodes which results in lower delay and loss also with the DCF based medium access. In the results demonstrated in Fig. 6.15, only for $\phi = 7$ in the $n_a = 10$ scenario, the call capacity of DCF is found to be higher than that with vAPs. In this case, the vAP and DCF based medium access mechanisms result in call capacity of 35 and 49, respectively.

6.12.4 Formation of Critical Zone

To demonstrate the formation of critical zones, we use a two-hop WLAN with 9 nodes (including the AP) as illustrated in Fig. 6.16. The AP serves the Internet connectivity to all the shown nodes and has two client nodes A1 and A2 in its own transmission range, i.e., within the first hop, which operate in channel A. Each of nodes A1 and A2 is, in turn, connected to a separate 4-node second-hop collision domain which are marked as W1 and W2, respectively. The nodes in collision domains W1 and W2 can operate in channel B without causing any interference to channel A. Additionally, the transmissions in W1 do not interfere with those in W2, and vice versa, since W1 and W2 are located outside the interference range of each other. Each of the nodes in the WLAN initiates a two way voice call using the G.711 codec and an aggregation level of 2 so that each voice call requires two voice streams (one from the node to the AP and another from the AP to the node) where each stream generates VoIP packets at a rate of 50 packet/sec. We use such a low aggregation level only to demonstrate the effect of a critical zone even in such a small WLAN. The data transmission rate and packet error rate are 1 Mbps and 0.05, respectively.

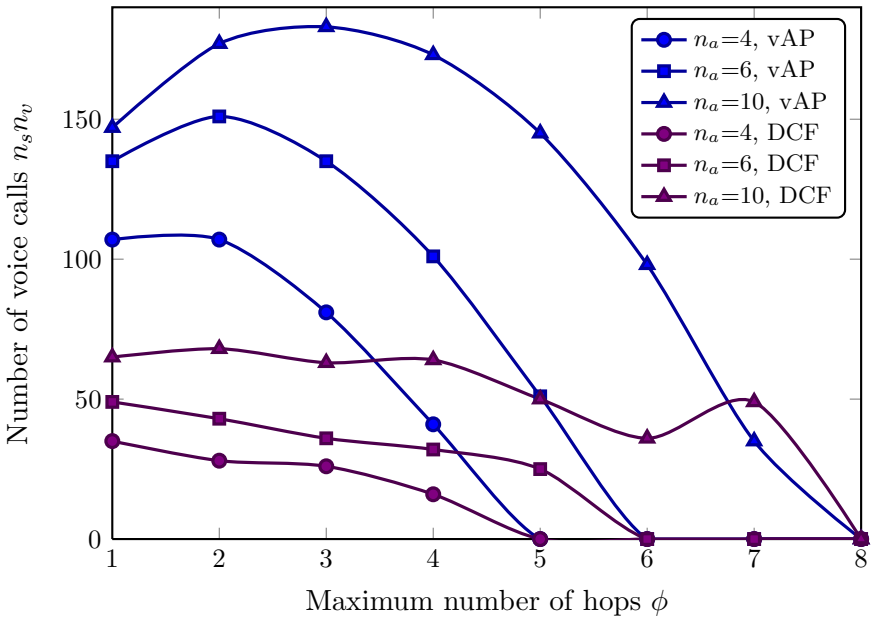


Figure 6.15: The voice capacity for the G.729 codec in a multihop 802.11b WLANs with 11 Mbps data rate using different aggregation levels and number of hops employing the vAP and DCF based medium access illustrating the initial increase in call capacity with the number of hops and later decrease for higher number of hops, also outlining the maximum number of hops to support medium quality calls.

When the traditional DCF based mechanism is used in every collision domain, the channel access delay in the second hop collision domains (W1 and W2) becomes 5.201 ms which incurs a queuing delay of 5.637 ms. Packet loss in channel access and queue are 2.774×10^{-6} and 1.963×10^{-29}, respectively. However, in the first hop, nodes A1 and A2, together with their gateway AP, suffer from increased packet arrival rate forming a bottleneck. Given the packet arrival rate and number of nodes, the delay in the channel access and queue for the first hop become 4.1 ms and 362.024 ms, respectively. Similarly, the loss in the channel access and queue become 9.405×10^{-7} and 0.0854, respectively. As a result, the first hop nodes form a critical zone which limits the performance of the WLAN. The maximum end-to-end delay Δ_1 and loss Γ_1 for this WLAN is 377 ms and 0.0854, respectively, which after including the packetization delay, look ahead delay and dejitter buffer delay, results in an R-score of 6.17. For a medium quality call, we need $R \geqslant 70.07$ as shown in (3.15) and, therefore, the given WLAN can not support the voice quality of the ongoing calls. This will result in the calls being terminated and, at least, severe user dissatisfaction.

However, if the nodes in the critical zone adopt vAP based medium access, the above demonstrated bottleneck can be avoided. For instance, if the nodes A1 and A2 become vAPs in the second hop collision domains W1 and W2, respectively, then all

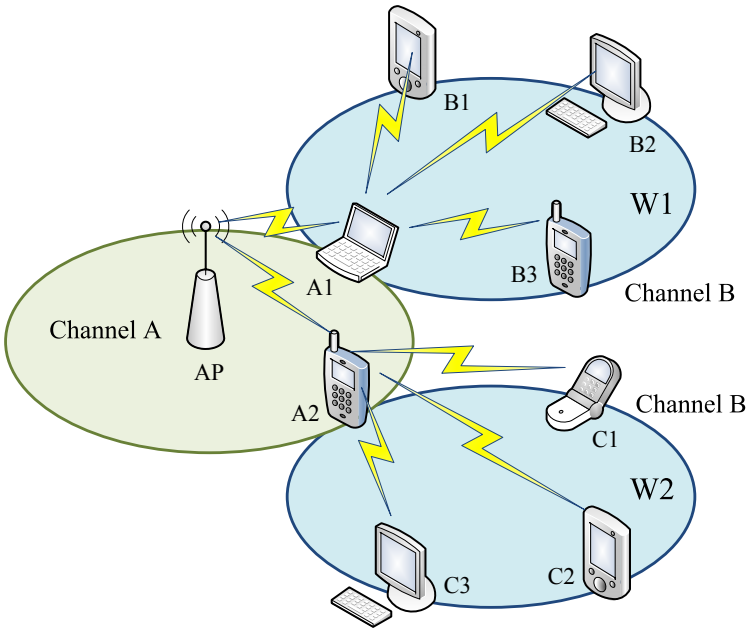


Figure 6.16: Demonstration of performance degradation due to the formation of a critical zone in a two-hop WLAN and avoiding the bottleneck by the use of virtual access points. G.711 with $n_a = 2$ is used to initiate one voice call per node with 5% packet error rate and 1 Mbps data rate. While using DCF forms a bottleneck in the ring 1 (channel A with nodes A1, A2 and AP) and no calls can be supported, using nodes A1 and A2 as virtual access points in collision domains W1 and W2 removes the bottleneck and all calls can be supported.

collision domains can utilize the time synchronized medium access (AP can coordinate channel access in the first hop). In this case, the delay in channel access and queue for the second hop collision domains become 12.84 and 3.03 ms, respectively. Similarly, loss in channel access and queue are determined to be 4.405×10^{-7} and 3.075×10^{-50}, respectively. Considering the cumulative packet arrival rate and collision domain size, the loss in channel access and queue for the first hop are 4.405×10^{-7} and 3.89×10^{-53}, respectively. Note that the channel access loss remains the same at each hop since it is independent of the collision domain size and packet arrival rate. The delay in the channel access and queue in the first hop are 12.01 and 2.576 ms, respectively. In this case, the end-to-end network delay Δ_1 and loss Γ_1 are 30.46 ms and 8.81×10^{-7}, respectively. After including the packetization, look ahead and dejitter buffer delay we find the R-score to be 95.45. Since we need $R \geqslant 80.16$ as shown in (3.15) for a high quality call, each of the nodes can now support a high quality call each. Using the mobility pattern from the statistical data of real human mobility [4] as shown in Fig. 6.6, we additionally find that the vAPs (A1 and A2) can poll all the 3 nodes in

their own collision domains (W1 and W2) with probability 1.0 during a pause time despite the overhead of the reconnoiter and bidding mechanism.

6.13 Key Contributions

The IEEE 802.11 PCF mechanism offers lower delay and loss in both channel access and queue compared to DCF by utilizing time synchronized medium access. On the other hand, due to the ad hoc nature of operation, DCF can be used over multiple hops providing a large coverage with a few access points while the use of PCF is strictly limited to single hop WLANs only. In this chapter, we proposed a new medium access mechanism combining the best features of PCF and DCF. The following are the key contributions from this chapter.

- In this chapter, we proposed a new medium access mechanism to offer the lower delay and loss of PCF while maintaining the wider coverage of DCF in a multihop WLAN. The proposed mechanism uses PCF like time synchronized medium access mechanism over multiple hops (in a multihop WLAN) where an access point is not available. Thus, the proposed mechanism can ensure a lower delay and loss in the channel access over multiple hops compared to DCF which is the only access mechanism applicable to such scenarios. The lower delay in the medium access results in lower delay and loss in the queue and improves the VoIP call capacity significantly.

- We used a randomized bidding mechanism to allow the selection of a client node which can temporarily coordinate the channel access like an AP. Thus, the larger coverage of DCF can still be provided while utilizing the time synchronized medium access of PCF at each hop. We used a reconnoiter mechanism to ensure that a client node does not try to become a vAP when there is a regular AP available. The reconnoiter and bidding mechanisms ensure that a regular AP can always gain control of the channel without any competition. We additionally used a veto system which a regular AP can use, if it becomes available later, to give a veto to an ongoing bidding and gain control of the channel immediately. If an AP is not aware of this veto mechanism, it can still acquire control of the channel in the following super-frame.

- The proposed mechanism is built using the existing protocols and techniques of the standard DCF and PCF mechanisms so that a client node that is unaware of the proposed scheme can operate under the coordination of a vAP without any problem. The downlink frames can be sent to nodes that are not PCF-compatible, however, in order to transmit the uplink packets, the nodes need to

recognize the polling frames that are used in the standard PCF mechanism, i.e., they have to be PCF-compliant although vAP-compliance is not required. To enable the vAP capabilities in a client node, only the MAC layer needs to be extended and any existing 802.11 physical layer can be utilized.

- When there is no vAP at an intermediate hop of a multihop WLAN, a vAP-enabled client node (but not necessarily currently a vAP itself) uses a RTS/CTS type channel reservation technique to quickly send all the packets in its queue. Our proposed mechanism also provides a better share to nodes that are forwarding a large number of voice packets to ensure fairness on a per voice call basis at each hop.

- We used the voice quality impairment budget in (3.16) to formulated a call capacity model as shown in (6.13) for a vAP enabled WLAN. The comparative analyses between the call capacity of vAP and DCF based medium access mechanisms showed that, while maintaining the voice quality, a superior voice capacity can be provided with the additional benefit of large coverage in a vAP enabled multihop WLAN.

6.14 Conclusion

In this chapter, we proposed the concept of virtual access point to extend the application of PCF to multihop, multichannel networks and applied it to increase VoIP call capacity. We utilized the Markov model presented in Section 5.3 to determine the delay and loss with the vAP based medium access mechanism considering the impact of channel errors. These delay and loss were used to determine the end-to-end delay and loss as outlined in Sections 6.9 and 6.10 and a call capacity model was formulated. In the previous chapters, starting from Chapter 3 to 6, we progressively developed call capacity estimation models for the DCF and PCF based medium access mechanisms under different network conditions. The theoretical results were substantiated by extensive simulation results. The capacity modeling task is extended in the current chapter by improving the VoIP call capacity and coverage through the introduction of virtual access point concept. Appropriate protocols and access mechanisms were devised to address different network scenarios. In the following chapter, we conclude this thesis by presenting conclusive remarks and directions on further extensions of the current work.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

The wide acceptance and ubiquitous use of VoIP are primarily due to its low cost of service. In order to offer the convenience of mobility to VoIP users without incurring a high service cost, we need to use the IEEE 802.11 WLANs for the last mile wireless coverage. IEEE 802.11 standards offer a moderate coverage and bandwidth at a low operational cost; and additionally, the wide spread availability of Wi-Fi compatible radio network interfaces in portable devices (mobile phones, PDAs and laptops) makes the WLANs an ideal wireless coverage solution for VoIP based voice calls. But the voice capacity of such networks is low; and in the absence of a quality assurance mechanism the voice quality degrades, especially under a high traffic load. To avoid customer dissatisfaction, voice quality requirements should be carefully considered in network design and planning. In this respect, an analytical call capacity model can greatly assist network designers in determining the call capacity of a target network for a chosen set of network aspects.

Due to the fast growth of the VoIP subscriber base, VoIP over Wi-Fi has attracted considerable interest from the network researchers, and a substantial research effort is invested in designing analytical call capacity models for IEEE 802.11 WLANs. But the existing analytical call capacity models used unrealistic and over-simplified assumptions that make their findings unusable in real networks. The effects of key network aspects like imperfect channel, capture effect and unsaturated traffic condition were ignored. No analytical call capacity model was ever designed for multihop, multi-channel WLANs with multi-interface nodes. Most importantly, the voice quality degradation caused by the combined impact of the end-to-end delay and loss were not considered. Therefore, the research challenge of designing a precise analytical call capacity model

considering the voice quality and the above key network aspects remained open to date, which is the main driving factor behind this dissertation.

In this thesis, we addressed the above mentioned research challenges by developing a set of analytical call capacity models considering realistic traffic conditions (i.e., saturated and unsaturated) and real network aspects including imperfect channel and capture effect. We modeled the medium access control mechanism and the queue with attention to fine details so that any over or under estimation can be avoided. Both PCF and DCF mechanisms were considered; and additionally, multihop scenarios with uniform and non-uniform node distributions are incorporated for DCF based WLANs. Instigated by the low call capacity of IEEE 802.11 WLANs, we considered the availability of multiple channels and additionally introduce a novel, hybrid medium access mechanism where the client nodes volunteer to coordinate the channel access in the absence of an AP. The proposed mechanism offers a high call capacity like PCF while providing a wide coverage like DCF. The following is a brief summary of our key contributions in this dissertation.

- **Analytical Call Capacity Model Considering Voice Quality:** We developed analytical VoIP call capacity models for DCF and PCF based medium access considering voice quality guidelines. We use the ITU-T E-model to determine the voice quality impairments due to the end-to-end delay and loss and determine the impairment budget for voice calls in a 802.11 WLAN. Our proposed capacity models use the impairment budget as the limiting condition so that the WLANs are sufficiently provisioned to serve the expected call volume while maintaining the voice quality.

- **Enhanced Markov Model for DCF based Medium Access Mechanism:** The DCF based medium access mechanism defined in the IEEE 802.11 standards incurs a high delay and loss in channel access which degrade the voice quality. We model the finite retry BEB based DCF mechanism using a Markov model to accurately determine the delay and loss in the medium access so that the voice quality impairment can be precisely determined. Extensive simulations show close matches with our analytical results and validate both the Markov model and our call capacity model for DCF based medium access. The following are incorporated into a single model which, to the best of our knowledge, is not available in the current literature.

  - **Realistic Traffic Model:** Most WLANs operate under the unsaturated traffic condition for a considerable amount of time, which is also considered in our proposed Markov model. The Markov model, as shown using

simulations, can model unsaturated and saturated traffic condition closely.

- **Power Capture Effect:** Most modern WLAN devices support power capture which can bring a positive effect to the network performance. We consider the capture effect in a parameterized way which is primarily discussed for Rayleigh fading channels, although other fading models can easily be incorporated. Considering the capture effect in relation to collisions in the DCF based medium access allows our model to give a more realistic insight into the network performance.
- **Imperfect Channel:** An ideal channel is only possible in theory and every channel in the real world exhibits noise and interference that incur errors in packet transmissions. In order to model the real world network environment closely, we incorporate packet error rate in determining the success or failure of every packet transmission.

- **Multihop, Multichannel WLAN:** No existing work in the current literature provided an analytical call capacity model for multi-channel, multihop WLANs. In the multihop WLANs, the intermediate nodes forward voice packets to and receive from the nodes that are located far from the access point, and suffer from a higher packet arrival rate. We present a call capacity model for multihop, multi-channel WLANs considering the cumulative packet arrival rate in a multihop scenario. The impact of utilizing multiple channels and the formation of a critical zone due to this high traffic arrival rate are also considered in our call capacity model. We use the Markov model proposed for single hop WLAN to model the delay and loss in the medium access at each hop, therefore, our capacity model for multihop WLAN also considers the effects of an imperfect channel, realistic traffic and capture effect.

- **Modeling the Queue** From extensive simulations, we identify that the delay and loss in the queue incur a severe quality impairment to the voice streams. We use M/M/1/s_q and M/M/n_t/s_q systems to model nodes with single and multiple network interfaces, respectively, and determine the delay and loss in the queue which are used to determine the end-to-end delay and loss for the voice packets.

- **Capacity with the PCF based Medium Access Mechanism:** The existing analytical call capacity models for the PCF based medium access mechanism considered ideal channels only or incorrectly analyzed the delay in channel access with imperfect channels. As a result, the existing works provide inaccurate estimates of the voice capacity of PCF based WLANs. We use a Markov model to determine the delay and loss in the PCF based medium access mechanism which

is validated using extensive simulations. The presented Markov model allows a precise determination of the delay and loss so that an exact voice capacity estimation is possible. We also show that a higher call capacity can be attained with PCF than with DCF based medium access.

- **Hybrid Channel Access Scheme:** We propose a novel medium access mechanism that combines the best features of both DCF and PCF, and allows the PCF based time synchronized medium access to be used over multiple hops so that the spatial reuse and wide coverage like DCF can be attained. Our investigation using statistical human mobility data reveals that the scheme is feasible despite the mobility of mobile devices. We also illustrate the increase in the call capacity over multiple hops using our proposed scheme.
- **Detailed Capacity Analysis:** For all of our capacity models, for DCF and PCF in single hop or multihop WLANs using single or multiple channels and network interfaces, we analyze the impact of the voice codec, aggregation level, packet error rate, capture effect (only for DCF, since collisions do not occur in PCF) and data rate on the voice capacity. For multihop and multichannel scenario with multi-interface nodes, we additionally analyze the effect of multiple channels, network interfaces and multihop routes. Our detailed analyses reveal the underlying pattern of the effects of different network aspects on the voice capacity in the IEEE 802.11 WLANs.

7.2 Future Works

The voice call capacity models presented in this thesis can be extended in a number of ways, some of which are discussed below.

- As mentioned in Chapter 6, the DCF based medium access in the single hop WLANs gives rise to unbalanced traffic condition since the access point needs to transmit more packets than the client nodes. The PCF based medium access mechanism does not suffer from such a problem since every CFP enables transmissions of the same number of uplink and downlink packets. The unbalanced traffic condition in DCF can be addressed using a strategy similar to the vAP based medium access proposed in Section 6.6 where a non-vAP node can transfer all the accumulated packets using a RTS/CTS based channel reservation mechanism. In a DCF based single hop WLAN, once the AP gains access to the channel (possibly after some backoff), it can reserve the channel in a similar manner, and transfer all the downlink packets to the associated nodes after waiting for one SIFS between successive handshakes.

- Although we considered the capture effect in a parameterized way, we discussed its effect only for Rayleigh fading channels. Further analyses can be carried out for other fading models including Nakagami and Weibull fading.

- We identified that in multihop WLANs, the nodes in the outer rings suffer from a higher delay and loss than those nearer to the AP. The DCF based channel access mechanism can be extended to ensure fairness for all nodes in terms of the end-to-end delay and loss. A simple measure to accomplish this can be a distributed mechanism that selects appropriate backoff parameters, including the initial contention window size, its increment after each collision, the number of retry stages and the retry limit, etc., for each packet that belongs to a particular voice stream in order to maintain a given upper limit of the end-to-end delay and loss. The backoff parameters are to be chosen in a way to make the end-to-end delay and loss uniform for all nodes in the WLAN. In this manner, all nodes can be treated in a fair manner despite their location.

- If a packet suffers from a high end-to-end delay due to a high number of retransmissions in its multihop route (in a multihop WLAN), it might already be too late to be played out and the probability that the packet will be dropped by the dejitter buffer eventually increases. But such a packet still consumes bandwidth at each hop which is unnecessarily wasted. To reduce this wastage of the precious channel bandwidth, each packet could carry the number of retransmissions it faced at each hop; and using this historical data of previous packets, a stale packet identification criterion could be employed at each intermediate hop to identify and drop stale packets. This mechanism can substantially improve the voice performance of a nearly saturated WLAN.

- In the shared queue mechanism we introduced in Chapter 4, a network interface (whenever it becomes free) picks the next available packet from the queue. But routing protocols may require a packet to be forwarded using a given network interface. In such a case, there might be multiple packets buffered for one particular NI while another NI remains idle and has no packet to transmit. The performance of such a queue in light of routing protocols can be incorporated in the call capacity models.

- In all our capacity models presented in Chapters $3 \sim 6$, we consider homogeneous WLANs only where every node has the same number of NIs, uses the same codec, aggregation level, capture threshold, data transmission rate and suffers from the same packet error rate. Office and lab networks can be expected to be mostly homogeneous while public WLANs are usually heterogeneous and

therefore, these parameters can be different among the mobile nodes. In the absence of information on the expected combination of these parameters, we use homogeneous WLANs but our call capacity models can be extended to reflect the voice capacity of heterogeneous networks as well.

- In determining the value of the prioritization parameter ζ (Chapter 6), we consider a homogeneous WLAN and assign an equal priority to all the nodes who are bidding to become a vAP. The exact values v_i and weights w_i of the node resources are left open which can be determined using a comparative study of all resources of a node. The determination of these values, their weights and their effect on the call capacity can be determined in future works.

- In order to determine the maximum achievable call capacity, we considered only voice calls in the WLANs. But real networks additionally carry data packets as the users browse through the web, download songs and carry out online transactions, etc. Our call capacity models can be extended to incorporate a given volume of background data traffic.

- We considered a circular shape for the associated WLAN as used in previous works, including [150], in order to address a more generalized network scenario. But square and other irregular shapes for the WLAN can also be considered and their impact on the voice capacity can be determined.

- The CBR traffic used in our models for VoIP calls without silence suppression can easily be extended to incorporate the silence suppression based VBR traffic using the on-off speech model by Brandy [162].

Publications from this Work

The following papers are published from the contributions made in this thesis.

- **Conference Paper**

  1. M. A. R. Siddique and J. Kamruzzaman, "Increasing voice capacity over IEEE 802.11 WLAN," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–6.
  2. M. A. R. Siddique and J. Kamruzzaman, "Performance analysis of PCF based WLANs with imperfect channel and failure retries," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–6.
  3. M. A. R. Siddique and J. Kamruzzaman, "VoIP service over multihop 802.11 networks with power capture and channel noise," in *Proc. IEEE International Conference on Communications (ICC)*, 2010, pp. 1–6.
  4. M. A. R. Siddique and J. Kamruzzaman, "Voice over multi-channel multi-radio WLANs with power capture and imperfect channel," in *Proc. IEEE Wireless Communications & Networking Conference (WCNC)*, 2010, pp. 1–6. **Received best paper award.**
  5. M. A. R. Siddique and J. Kamruzzaman, "Performance analysis of m-retry BEB based DCF under unsaturated traffic condition," in *Proc. IEEE Wireless Communications & Networking Conference (WCNC)*, 2010, pp. 1–6.
  6. M. A. R. Siddique and J. Kamruzzaman, "VoIP capacity over PCF with imperfect channel," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–6.
  7. M. A. R. Siddique and J. Kamruzzaman, "VoIP call capacity over wireless mesh networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–6.

- **Technical Report**

  1. M. A. R. Siddique and J. Kamruzzaman, "VoIP call capacity of IEEE 802.11 based wireless mesh networks," Gippsland School of Information Technology, Faculty of Information Technology, Monash University, Australia, GSIT TECHNICAL REPORT, 1/2008, Feb. 2008.

- **Publication under review**

  1. M. A. R. Siddique and J. Kamruzzaman, "Quality VoIP calls over 802.11 Networks," IEEE Transactions on Mobile Computing, pp. 1–15, 2010, *under review.*

- **Citation**

  1. Mehmooda *et al.* [163] (detailed below) cited our work [23] which is a contribution from this dissertation.

     R. Mehmooda, R. Alturkia, and S. Zeadallyb, "Multimedia applications over metropolitan area networks (MANs)," Journal of Network and Computer Applications, pp. 1–12, 2010, *in press.*

Bibliography

[1] Cisco Public Information. (2010, Jun.) Cisco visual networking index: Forecast and methodology, 2009-2014. Cisco Visual Networking Index (VNI). [Online]. Available: http://ciscosystems.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[2] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[3] X. Ma, Y. Wu, Z. Niu, and T. Saito, "Performance analysis of the packetized voice transmission with PCF in an IEEE 802. 11 infrastructure wireless LAN," in *Proc. 9th Asia-Pacific Conference on Communications (APCC)*, 2003, pp. 571–575.

[4] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the levy-walk nature of human mobility." in *IEEE INFOCOM*, 2008, pp. 924–932.

[5] Comparison of WLAN, WPAN and WMAN technologies. www.NetworkDictionary.com. [Online]. Available: http://www.networkdictionary.com/Wireless/Comparison-of-WLAN.php

[6] R. Beuran, "VoIP over wireless LAN survey," Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, Research report IS-RR-2006-005, Apr. 2006.

[7] S. Mathiyalakan, "VoIP adoption: Issues & concerns," *Communications of the IIMA*, vol. 6, no. 2, pp. 19–24, 2006.

[8] D. Pogue. (2008, Dec.) Cool phone tricks. TED Conferences, LLC. [Online]. Available: http://www.ted.com/index.php/talks/david_pogue_on_cool_phone_tricks.html

[9] T. Burton. (2008, Feb.) Twenty percent annual growth for VoIP. The Fiercemarkets Network. [Online]. Available: http://www.fiercevoip.com/story/twenty-percent-annual-growth-for-voip/2008-02-25

[10] J. Edwards. (2008, Mar.) What the recession will do to VoIP. www.voip-news.com. [Online]. Available: http://www.voip-news.com/feature/recession-will-do-voip-032408/

[11] A. Nusca. (2009, Oct.) In recession, VoIP makes strong case for cutting your home landline. ZDNet, CBS Interactive. [Online]. Available: http://www.zdnet.com/blog/btl/in-recession-voip-makes-strong-case-for-cutting-your-home-landline/25192/

[12] I. Ahmad, J. Kamruzzaman, and S. Aswathanarayaniah, "An improved preemption policy for higher user satisfaction," in *Proc. The IEEE 19th International Conference on Advanced Information Networking and Applications (AINA 2005)*, 2005, pp. 749–754.

[13] S. Garg and M. Kappes, "Can I add a VoIP call?" in *Proc IEEE International Conference on Communications (ICC)*, 2003, pp. 779–783.

[14] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in *Proc. IEEE International Conference on Communications (ICC)*, 2004, pp. 196–201.

[15] F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyanathan, A. Dutta, P. Agrawal, T. Kodama, Y. Katsube, and T. Technol, "Voice performance in WLAN networks—an experimental study," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2003, pp. 3504–3508.

[16] A. Zahedi and K. Pahlavan, "Capacity of a wireless lan with voice and data services," *IEEE Transactions on Communications*, vol. 48, no. 7, pp. 1160–1170, 2000.

[17] T. Kawata, S. Shin, A. G. Forte, and H. Schulzrinne, "Using dynamic PCF to improve the capacity for voip traffic in IEEE 802.11 networks," in *IEEE Wireless Communicationa and Networking Conference (WCNC)*, 2005, pp. 1589–1595.

[18] *The E-model, a computational model for use in transmission planning*, ITU-T Recommendation G.107 Std., 2005.

[19] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, "Unsaturated throughput analysis of IEEE 802.11 in presence of non ideal transmission channel

and capture effects," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1276–1286, 2008.

[20] M. Veeraraghavan, N. Cocker, and T. Moors, "Support of voice services in IEEE 802.11 wireless LANs," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, 2001, pp. 488–497.

[21] M. A. Zaharia and S. Keshav, "Fast and optimal scheduling over multiple network interfaces," University of Waterloo, Tech. Rep., May 2007.

[22] L. Angrisani, A. Napolitano, and A. Sona, "VoIP over IEEE 802.11 wireless networks: Experimental analysis of interference effects," in *Proc. International Symposium on Electromagnetic Compatibility - EMC Europe*, 2008, pp. 1–6.

[23] M. A. R. Siddique and J. Kamruzzaman, "VoIP call capacity over wireless mesh networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–6.

[24] ——, "Performance analysis of m-retry BEB based DCF under unsaturated traffic condition," in *Proc. IEEE Wireless Communications & Networking Conference (WCNC)*, 2010, pp. 1–6.

[25] ——, "Voice over multi-channel multi-radio wlans with power capture and imperfect channel," in *Proc. IEEE Wireless Communications & Networking Conference (WCNC)*, 2010, pp. 1–6.

[26] ——, "Voip service over multihop 802.11 networks with power capture and channel noise," in *Proc. IEEE International Conference on Communications (ICC)*, 2010, pp. 1–6.

[27] ——, "Performance analysis of PCF based WLANs with imperfect channel and failure retries," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–6.

[28] ——, "Performance analysis of PCF based WLANs with imperfect channel and failure retries," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–6.

[29] ——, "Increasing voice capacity over IEEE 802.11 WLAN using virtual access points," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–6.

[30] VoIP makes sense during the recession. CompareBusinessProducts.com. [Online]. Available: http://www.comparebusinessproducts.com/business-voip/news-on-business-voip/voip-makes-sense-during-the-recession.aspx

[31] A. Cox. (2010, Jun.) Mobile voice— whispers of change? Juniper Research Ltd. [Online]. Available: http://www.juniperresearch.com/shop/products/whitepaper/pdf/Whitepaper-MobileVoice-sec.pdf

[32] C. Gallen. (2010, Apr.) 487 million users to enjoy hd-quality mobile voice in 2015. Allied Business Intelligence, Inc. [Online]. Available: http://www.abiresearch.com/press/1645-487+Million+Users+to+Enjoy+HD-Quality+Mobile+Voice+in+2015

[33] T. Czech. (2010, Feb.) VoIP penetration forecast to reach 79% of US businesses by 2013. Biz-News.com. [Online]. Available: http://voip.biz-news.com/news/2010/02/05/0001/

[34] I. Elwood. (2006) Billions in business market growth expected - VoIP news. http://www.voip-news.com/news/billions-projected.

[35] R. Hof. (2005, Sep.) Why ebay is buying skype—businessweek. Bloomberg L.P. [Online]. Available: http://www.businessweek.com/the_thread/techbeat/archives/2005/09/why_ebay_is_buy.html

[36] BBC ©MMX. (2005, Sep.) Ebay to buy Skype in $2.6bn deal. [Online]. Available: http://news.bbc.co.uk/1/hi/business/4237338.stm

[37] A. Smith. (2002, Aug.) Demystifying convergence and VoIP. CISCO press. http://www.ciscopress.com/articles/article.asp?p=28472.

[38] A. Mondal, C. Huang, J. Li, M. Jain, and A. Kuzmanovic, "A case for WiFi relay: Improving VoIP quality for WiFi users," in *Proc. IEEE International Conference on Communications (ICC)*, 2010, pp. 1–5.

[39] Infocom Telecommunications Consultancy. (2008, Feb.) Nearly 88 million VoIP subscriptions by 2012 in western europe—the Netherlands to show the highest VoIP penetration of nearly 70%. Stuttgart, Germany. [Online]. Available: http://www.infocom-de.com/pressarchives/press_060208.html

[40] G. Kim. (2009, Mar.) VoIP in U.S. and Europe in 2012: A tale of two markets. Metaswitch networks. [Online]. Available: http://blogs.metaswitch.com/gk/2009/03/voip-in-us-and-europe-in-2012-a-tale-of-two-markets.html

[41] Internet World Stats. (2010, Jun.) Internet usage statistics—the internet big picture—world internet users and population stats. Miniwatts Marketing Group. [Online]. Available: http://www.internetworldstats.com/stats.htm

[42] H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, *SIP: Session Initiation Protocol*, Network Working Group—Request for Comments: 3261 Std., 2002. [Online]. Available: http://tools.ietf.org/html/rfc3261

[43] *Cisco IP Telephony Solution Reference Network Design Guide*, Cisco Systems, Inc. Std., 2003.

[44] S. Karapantazis and F.-N. Pavlidou, "VoIP: A comprehensive survey on a promising technology," *Computer Networks*, vol. 53, no. 12, pp. 2050–2090, 2009.

[45] Wikipedia, the free encyclopedia. (2010, Aug.) Code-excited linear prediction. Wikimedia Foundation, Inc. [Online]. Available: http://en.wikipedia.org/wiki/Code-excited_linear_prediction

[46] *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)*, ITU-T Recommendation G.729 Std., Jan. 2007.

[47] *Pulse Code Modulation (PCM) of Voice Frequencies*, ITU-T Recommendation G.711 Std., Nov. 1988.

[48] *Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*, ITU-T Recommendation G.723.1 Std., May 2006.

[49] *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*, Recommendation G.726 Std., Dec. 1990.

[50] Wikipedia, the free encyclopedia. (2010, May) G.726. [Online]. Available: http://en.wikipedia.org/wiki/G.726/

[51] *Methods for subjective determination of transmission quality*, ITU-T Recommendation P.800 Std., Aug. 1996.

[52] *Amendment 2: New Appendix II — Planning examples regarding delay in packet-based networks*, ITU-T Recommendation G.108 (1999) — Amendment 2 Std., Mar. 2004.

[53] *Application of the E-model: A planning guide*, ITU-T Recommendation G.108 Std., Sep. 1999.

[54] *New Appendix I: The relationship between and interaction of talker echo and absolute delay*, ITU-T Recommendation G.108 (1999) — Amendment 1 Std., Sep. 2003.

[55] *Transmission impairments due to speech processing*, ITU-T Recommendation G.113 Std., Nov. 2007.

[56] *Transmission impairments due to speech processing — Appendix I: Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl*, ITU-T Recommendation G.113 — Appendix I Std., May 2002.

[57] R. Cole and J. Rosenbluth, "Voice over IP performance monitoring," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 2, pp. 9–24, Apr. 2001.

[58] *The E-model, a computational model for use in transmission planning*, ITU-T Recommendation G.107 Std., 1998.

[59] L. Carvalho, E. Mota, R. Aguiar, A. Lima, J. De Souza, and A. Barreto, "An E-Model implementation for speech quality evaluation in VoIP systems," in *Proc. 10th IEEE Symposium on Computers and Communications (ISCC)*, 2005, pp. 933–938.

[60] *Packet-based multimedia communications systems*, ITU-T Recommendation H.323 Std., Dec. 2009.

[61] A. Markopoulou, F. Tobagi, and M. Karam, "Assessing the quality of voice communications over internet backbones," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 747–760, 2003.

[62] Coda Research Consultancy. (2010, Feb.) Wi-Fi enabled mobile phone handsets in the US, 2010-2015. Reportlinker.com. [Online]. Available: http://www.reportlinker.com/p0177156/Wi-Fi-enabled-mobile-phone-handsets-in-the-US-2010-2015.html

[63] T. He, S. Chan, and C. Wong, "Homemesh: a low-cost indoor wireless mesh for home networking," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 79–85, 2008.

[64] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications*, IEEE Std 802.11-1997 Std., Jun. 1997.

[65] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHZ Band*, IEEE Std 802.11a-1999 Std., Sep. 1999.

[66] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: Higher-speed Physical Layer Extension In The 2.4 GHz Band*, IEEE Std 802.11b-1999 Std., Sep. 1999.

[67] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*, IEEE Std 802.11e-2005 Std., Nov. 2005.

[68] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band*, IEEE Std 802.11g-2003 Std., Jun. 2003.

[69] *Supplement To IEEE Standard For Information Technology- Telecommunications And Information Exchange Between Systems- Local And Metropolitan Area Networks- Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications: AMENDMENT 6: Medium Access Control (MAC) Security Enhancements*, IEEE Std 802.11i-2004 Std., Oct. 2005.

[70] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LANs," in *Proc. of the Conference on Communications Architectures, Protocols and Applications (ACM SIGCOMM)*, 1994, pp. 212–225.

[71] P. Karn, "MACA—a new channel access method for packet radio," in *Proc. ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, 1990, pp. 134–140.

[72] K. J. Biba, *A Hybrid Wireless MAC Protocol Supporting Asynchronous and Synchronous MSDU Delivery Services*, IEEE P802.11/91-92 Std., Sep. 1991.

[73] J. Geier. (2002, Nov.) 802.11 medium access methods. QuinStreet Inc. [Online]. Available: http://www.wi-fiplanet.com/tutorials/article.php/1548381/80211-Medium-Access-Methods.htm

[74] J. L. Stanford. (2007, Jan.) The Spectec SDW-820 SDIO WiFi card review. Gear Diary. [Online]. Available: http://www.geardiary.com/2007/01/12/the-spectec-sdw-820-sdio-wifi-card-review/

[75] H. Wu, S. Cheng, Y. Peng, K. Long, and J. Ma, "IEEE 802.11 distributed coordination function (DCF): analysis and enhancement," in *Proc. IEEE International Conference (ICC)*, 2002, pp. 605–609.

[76] M. I. Abu-Tair, G. Min, Q. Ni, and H. Liu, "Adaptive medium access control for VoIP services in IEEE 802.11 WLANs," in *Proc. 4th IEEE International Conference on Circuits and Systems for Communications (ICCSC)*, 2008, pp. 487–491.

[77] I. H. Lin and J. Y. Pan, "Throughput analysis of a novel backoff algorithm for IEEE 802.11 WLANs," in *Proc. Wireless Telecommunications Symposium*, 2005, pp. 85–90.

[78] P. Chatzimisions, V. Vitasas, A. Boucouvalas, and M. Tsoulfa, "Achieving performance enhancement in IEEE 802.11 WLANs by using DIDD backoff mechanism," *International Journal of Communication Systems*, vol. 20, no. 1, pp. 23–41, 2007.

[79] I. Vukovic, N. Smavatkul, M. Inc, and I. Arlington Heights, "Saturation throughput analysis of different backoff algorithms in IEEE 802. 11," in *Proc. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2004, pp. 1870–1875.

[80] I. Vukovic and N. Smavatkul, "Delay analysis of different backoff algorithms in IEEE 802.11," in *Proc. 2004 IEEE 60th Vehicular Technology Conference (VTC2004-Fall)*, 2004, pp. 4553–4557.

[81] M. A. Bender, M. Farach-Colton, S. He, B. C. Kuszmaul, and C. E. Leiserson, "Adversarial analyses of window backoff strategies," in *Proc. 18th International Parallel and Distributed Processing Symposium ((IPDPS)*, 2004, pp. 203–210.

[82] Z. Cao, R. P. Liu, X. Yang, and Y. Xiao, "Modeling IEEE 802.11 DCF system dynamics," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010, pp. 1–5.

[83] Y. Barowski, S. Biaz, and P. Agrawal, "Towards the performance analysis of IEEE 802.11 in multi-hop ad-hoc networks," in *Proc. 2005 IEEE Wireless Communications and Networking Conference (WCNC)*, 2005, pp. 100–106.

[84] Y. S. Liaw, A. Dadej, and A. Jayasuriya, "Performance analysis of IEEE 802.11 DCF under limited load," in *Proc. Asia Pacific Conference on Communications*, 2005, pp. 759–763.

[85] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs," in *Proc. 7th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 1996, pp. 392–396.

[86] G. Bianchi, "IEEE 802.11-saturation throughput analysis," *IEEE Communications Letters*, vol. 2, no. 12, pp. 318–320, 1998.

[87] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF performance analysis," *IEEE Communications Letters*, vol. 9, no. 8, pp. 765–767, 2005.

[88] E. Ziouva and T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: throughput and delay analysis," *Computer Communications*, vol. 25, no. 3, pp. 313–321, 2002.

[89] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement," in *Proc. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 2002, pp. 599–607.

[90] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "Performance analysis of IEEE 802.11 DCF in presence of transmission errors," in *Proc. IEEE International Conference on Communications (ICC)*, 2004, pp. 3854–3858.

[91] ——, "Influence of channel BER on IEEE 802.11 DCF," *Electronics Letters*, vol. 39, no. 23, pp. 1687–1689, 2003.

[92] ——, "Packet delay analysis of IEEE 802.11 MAC protocol," *Electronics Letters*, vol. 39, no. 18, pp. 1358–1359, 2003.

[93] ——, "IEEE 802.11 packet delay-a finite retry limit analysis," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2003, pp. 950–954.

[94] J. Jun and M. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Communications*, vol. 10, no. 5, pp. 8–14, 2003.

[95] Z. Hadzi-velkov and B. Spasenovski, "Capture effect in IEEE 802.11 basic service area under influence of rayleigh fading and near/far effect," in *Proc. The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2002, pp. 172–176.

[96] ——, "On the capacity of IEEE 802.11 DCF with capture in multipath-faded channels," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 191–199, 2002.

[97] *One-way transmission time*, ITU-T Recommendation G.114 Std., Feb. 1996.

[98] B. Langenhoven, "VoP over WiFi finds niche uses, adds to unified communications ecosystem but DECT still rules wireless office telephony," *Quantum—The specialist Journal for Electronics and Communications Professionals*, vol. 4, no. 10, p. 18, 2010.

[99] K. Medepalli, P. Gopalakrishnan, D. Famolari, T. Kodama, T. Technol, and N. Piscataway, "Voice capacity of IEEE 802.11 b, 802.11 a and 802.11 g wireless LANs," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2004, pp. 1549–1553.

[100] *Artifical Conversational Speech*, ITU-T Recommendation P.59 Std., May 1993.

[101] "Spectral link voice priority." [Online]. Available: http://www.spectralink.com/products

[102] D. Malone, P. Clifford, and D. J. Leith, "On buffer sizing for voice in 802.11 WLANs," *IEEE Communications Letters*, vol. 10, no. 10, pp. 701–703, 2006.

[103] P. Dini, O. Font-Bach, and J. Mangues-Bafalluy, "Experimental analysis of VoIP call quality support in IEEE 802.11 DCF," in *Proc. 6th International Symposium on Communication Systems, Networks and Digital Signal Processing (CNSDSP)*, 2008, pp. 443–447.

[104] S. Shin and H. Schulzrinne, "Measurement and analysis of the VoIP capacity in IEEE 802.11 WLAN," *IEEE Transactions on Mobile Computing*, vol. 8, no. 9, pp. 1265–1279, Sep. 2009.

[105] ——, "Experimental measurement of the capacity for VoIP traffic in IEEE 802.11 WLANs," in *Proc. 26th IEEE International Conference on Computer Communications (INFOCOM)*, 2007, pp. 2018 –2026.

[106] S. Armenia, L. Galluccio, A. Leonardi, and S. Palazzo, "Transmission of VoIP traffic in multihop ad hoc IEEE 802.11 b networks: experimental results," in *Proc. First International Conference on Wireless Internet (WICON)*, 2005, pp. 148–155.

[107] Y. Liu, Y. Xiong, Y. Yang, P. Xu, and Q. Zhang, "An experimental study on multi-channel multi-radio multi-hop wireless networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2005, pp. 311–315.

[108] D. Niculescu, S. Ganguly, K. Kim, and R. Izmailov, "Performance of VoIP in a 802.11 wireless mesh network," in *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, 2006, pp. 1–11.

[109] S. Ganguly, V. Navda, K. Kim, A. Kashyap, D. Niculescu, R. Izmailov, and S. Hong, "Performance optimizations for deploying VoIP services in mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2147–2158, 2006.

[110] P. Falconio, J. Mangues-Bafalluy, M. Cardenete-Suriol, and M. Portoles-Comeras, "Performance of a multi-interface based wireless mesh backbone to support VoIP service delivery," in *Proc. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2006, pp. 159–165.

[111] P. Falconio, J. Mangues-Bafalluy, M. Cardenete-Suriol, and R. Cusani, "Assessment of the perceived VoIP quality in a multi-interface mesh network with different traffic loads," in *Proc. IST Mobile & Wireless Communications Summit (Mobilesummit)*, 2006, pp. 1–6.

[112] D. Chen, S. Garg, M. Kappes, and K. Trivedi, "Supporting VBR VoIP traffic in IEEE 802.11 WLAN in PCF mode," in *Proc. of OPNETWork*, 2002, pp. 1–6.

[113] S. Guha, N. Daswani, and R. Jain, "An experimental study of the Skype peer-to-peer VoIP system," in *Proc. International workshop on Peer-To-Peer Systems (IPTPS)*, Feb. 2006, pp. 1–6.

[114] Y. Kim and Y. Suh, "Adaptive polling MAC schemes for IEEE 802.11 wireless LANs," in *IEEE Vehicular Technology Conference (VTC)*, 2003, pp. 2528–2532.

[115] W. Quan and D. M. Hui, "Improving the performance of WLAN to support VoIP application," in *Proc. 2005 2nd International Conference on Mobile Technology, Applications and Systems*, 2005, pp. 1–5.

[116] *One-way transmission time*, ITU-T Recommendation G.114 Std., May 2003.

[117] S. Simoens and D. Bartolome, "Optimum performance of link adaptation in hiperlan/2 networks," in *Proc. IEEE VTS 53rd Vehicular Technology Conference (VTC) Spring*, vol. 2, 2001, pp. 1129–1133.

[118] T. Kawata and H. Yamada, "Adaptive multi-rate VoIP for IEEE 802.11 wireless networks with link adaptation function," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2006, pp. 357–361.

[119] H.-Y. Hsieh, Y.-E. Lin, and H.-P. Lin, "Enhancing VoIP service for ubiquitous communication in a campus WLAN with partial coverage," *Computer Networks*, vol. 52, no. 13, pp. 2489–2504, 2008.

[120] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, 1994, pp. 1–9.

[121] C. Sreenan, J. Chen, P. Agrawal, and B. Narendran, "Delay reduction techniques for playout buffering," *IEEE Transactions on Multimedia*, vol. 2, pp. 88–100, 2000.

[122] L. Atzori and M. Lobina, "Speech playout buffering based on a simplified version of the ITU-T E-model," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 382–385, 2004.

[123] W. Wang, S. C. Liew, and V. Li, "Solutions to performance problems in VoIP over a 802.11 wireless LAN," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 1, pp. 366–384, 2005.

[124] S. Yun, H. Kim, H. Lee, and I. Kang, "100+ VoIP calls on 802.11b: The power of combining voice frame aggregation and uplink-downlink bandwidth control in wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 689–698, May 2007.

[125] V. Jacobson, *Compressing TCP/IP Headers for Low-Speed Serial Links*, Request for Comments: 1144 Std., 1990.

[126] C. Bormann *et al.*, *RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed*, Request for Comments: 3095 Std., 2001.

[127] P. Fortuna and M. Ricardo, "Header compressed VoIP in IEEE 802.11," *IEEE Wireless communications*, vol. 16, no. 3, pp. 69–75, 2009.

[128] S. Casner and V. Jacobson, *Compressing IP/UDP/RTP Headers for Low-Speed Serial Links*, Request for Comments: 2508 Std., 1999.

[129] L. Zhitao, Y. Min, and X. Huimin, "Performance improvement of VoIP over WLAN via packet segmentation strategy," in *Proc. 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, 2008, pp. 1–4.

[130] R. O. Baldwin, I. V. N. J. Davis, S. F. Midkiff, and R. A. Raines, "Packetized voice transmission using RT-MAC, a wireless real-time medium access control protocol," *ACM SIGMOBILE Mobile Computing and Communications Review (SIGMOBILE)*, vol. 5, no. 3, pp. 11–25, 2001.

[131] D. Gao, J. Cai, C. H. Foh, C.-T. Lau, and K. N. Ngan, "Improving WLAN VoIP capacity through service differentiation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 465–474, Jan. 2008.

[132] I. Dangerfield, D. Malone, and D. Leith, "Understanding 802.11e voice behaviour via testbed measurements and modeling," in *Proc. Workshop on Wireless Network Measurement (WiNMee)*, 2007.

[133] K. Stoeckigt and H. Vu, "VoIP capacity analysis in ieee 802.11 WLAN," in *Proc. IEEE 34th Conference on Local Computer Networks (LCN)*, 2009, pp. 116–123.

[134] S. Tsao and C. Huang, "An energy-efficient transmission mechanism for VoIP over IEEE 802.11 WLAN," *Wireless Communications and Mobile Computing*, vol. 9, no. 12, pp. 1629–1644, 2009.

[135] C. Zhu, H. Yu, X. Wang, and H.-H. Chen, "Improvement of capacity and energy saving of VoIP over IEEE 802.11 WLANs by a dynamic sleep strategy," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–5.

[136] A. Chan and S. C. Liew, "Performance of VoIP over multiple co-located IEEE 802.11 wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 8, no. 8, pp. 1063–1076, Aug. 2009.

[137] Z. Chen, L. Wang, F. Zhang, X. Wang, and W. Chen, "VoIP over WLANs by adapting transmitting interval and call admission control," in *Proc. IEEE International Conference on Communications (ICC)*, 2008, pp. 3242–3246.

[138] C. Ortiz, J.-F. Frigon, B. Sanso, and A. Girard, "Effective bandwidth evaluation for VoIP applications in IEEE 802.11 networks," in *Proc. International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2008, pp. 926–931.

[139] J. Pechiar, G. Perera, and M. Simon, "Effective bandwidth estimation and testing for Markov sources," *Performance Evaluation*, vol. 48, no. 1-4, pp. 157–175, 2002.

[140] F. Kelly, "Notes on effective bandwidths," *Stochastic networks: theory and applications*, vol. 4, pp. 141–168, 1996.

[141] P. Dini, N. Baldo, J. Nin-Guerrero, J. Mangues-Bafalluy, S. Addepalli, and L. L. Dai, "Distributed call admission control for VoIP over 802.11 WLANs based on channel load estimation," in *Proc. IEEE International Conference on Communications (ICC)*, 2010, pp. 1–6.

[142] K. Yasukawa, A. Forte, and H. Schulzrinne, "Distributed delay estimation and call admission control in IEEE 802.11 WLANs," in *Proc. IEEE International Conference on Network Protocols (ICNP)*, 2007, pp. 334–335.

[143] J. J. Huang and S. C. Chang, "A hybrid channel scan mechanism for improved VoIP handoff performance in 802.11 WLANs," in *Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2008, pp. 1–5.

[144] J. Huang, Y. Chen, S. Chang, and H. Ferng, "An efficient channel scan scheduling algorithm for VoIP handoffs in WLANs," in *Proc. IEEE 65th Vehicular Technology Conference (VTC)—Spring*, 2007, pp. 1340–1344.

[145] J. Okech, Y. Hamam, and A. Kurien, "A cross-layer adaptation for VoIP over infrastructure mesh network," in *Proc. Third International Conference on Broadband Communications, Information Technology Biomedical Applications*, 2008, pp. 97–102.

[146] X-Fabric. (2008, May) E-model tutorial. [Online]. Available: http://www.itu.int/ITU-T/studygroups/com12/emodelv1/tut.htm

[147] *Estimates of Ie and Bpl parameters for a range of CODEC types*, Study Group 12 — Delayed Contribution 106 Std., Jan. 2003.

[148] C.-H. Ng and B.-H. Soong, *Queueing modelling fundamentals with applications in communication networks*. John Wiley and Sons, 2008.

[149] W. Xiuchao, "Simulate 802.11b channel within ns2," National University of Singapore, Singapore, Tech. Rep., 2004.

[150] C. Burmeister, U. Killat, and K. Weniger, "Power control and rate adaptation in multi-hop access networks-is it worth the effort?" in *Proc. the 20th international teletraffic conference on Managing traffic performance in converged networks*, 2007, pp. 767–778.

[151] J. Huang, L. Wang, and C. Chang, "Coverage and capacity of a wireless mesh network," in *Proc. 2005 International Conference on Wireless Networks, Communications and Mobile Computing*, 2005, pp. 458–463.

[152] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[153] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking (Prentice Hall Communications Engineering and Emerging Technologies Series)*. Prentice Hall PTR, 2007.

[154] M. A. Y. Khan and D. Veitch, "Isolating physical PER for smart rate selection in 802.11," in *Proc. the 28th Conference on Computer Communications (INFOCOM 2009)*, 2009, pp. 1080–1088.

[155] B. Sikdar, "An analytic model for the delay in IEEE 802.11 PCF MAC-based wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1542–1550, 2007.

[156] A. Lindgren. (2003, Oct.) IEEE 802.11 PCF support for ns-2.1b8. [Online]. Available: http://www.ietf.org/mail-archive/web/manet/current/msg02887.html

[157] G. Nair. (2008, Nov.) MAC 802.11 point coordinator function. [Online]. Available: http://software.intel.com/en-us/articles/mac-80211-point-coordinator-function/

[158] M. Cloran, "Simulation of IEEE 802.11 PCF function in GloMoSim," Master's thesis, School of Electrical Engineering, Dublin City University, Apr. 2004.

[159] J. Chittamuru, A. Ramanathan, and M. Sinha. (2002, Jun.) Simulation of point coordination function for IEEE 802.11 wireless LAN using Glomosim. University of Massachusetts. Amherst, whitepaper.

[160] C. Marin, Y. Leprovost, M. Kieffer, and P. Duhamel, "Robust MAC-Lite and header recovery based improved permeable protocol layer scheme," in *Proc. IEEE 10th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'08)*, 2008, pp. 496–501.

[161] O. Alay, T. Korakis, Y. Wang, and S. Panwar, "An experimental study of packet loss and forward error correction in video multicast over IEEE 802.11b network," in *Proc. 6th IEEE Consumer Communications and Networking Conference (CCNC 2009)*, 2009, pp. 1–5.

[162] P. T. Brandy, "Model for generating on-off speech patterns in two-way conversation," *The Journal of the Acoustical Society of America*, vol. 46, p. 109, 1969.

[163] R. Mehmooda, R. Alturkia, and S. Zeadallyb, "Multimedia applications over metropolitan area networks (MANs)," *Journal of Network and Computer Applications*, pp. 1–12, 2010, (in press).

[164] (1989 to date) The network simulator— ns-2. Information Sciences Institute, University of Southern California. [Online]. Available: http://www.isi.edu/nsnam/ns

Appendix A

Network Simulations

A.1 Simulations with the NS-2 simulator

To validate the Markov model and call capacity model introduced in Chapter 3 of this thesis, we carried out extensive simulations in Network Simulator NS-2 [164] which is widely used by network researchers. Its development is supported by DARPA[1] through SAMAN[2] and by NSF[3] through CONSER[4], both in collaboration with other researchers including ACIRI[5]. SAMAN, CONSER and ACIRI are research projects contributing to the development of NS-2. Individual researchers also contributed valuable additions to NS-2 where the code on wireless part is contributed primarily by the UCB Daedelus[6] and CMU Monarch projects[7] and Sun Microsystems[8].

We used two versions (2.32 and 2.28) of NS-2 in validating the Markov model and call capacity model presented in Chapter 3. We used similar logical arguments in developing the capacity model for multihop, multi-channel WLANs presented in Chapters 4 and similar agreement with simulations can be expected. For the simulations in Chapters 3, the IEEE 802.11 nodes were statically deployed randomly and uniformly. NOAH routing protocol[9] was used to waive the overhead of routing packets. To validate the Markov model (please see Chapter 3) for DCF based medium access, we additionally developed an Exponential traffic generator. A separate statistics data collector is used to boost up the simulation speed. The associated codes of these de-

[1] http://www.darpa.mil
[2] http://www.isi.edu/saman/index.html
[3] http://www.nsf.gov/
[4] http://www.isi.edu/conser/index.html
[5] http://www.icir.org
[6] http://daedalus.cs.berkeley.edu/
[7] http://www.monarch.cs.rice.edu/cmu-ns.html
[8] http://www.oracle.com/us/sun/index.html
[9] http://icapeople.epfl.ch/widmer/uwb/ns-2/noah

velopments are posted in the NS-2 mail groups[10] for comments and public use. Each simulation result was deduced from simulation runs of considerable length (mentioned in the respective chapter) to ensure statistical significance. We performed the simulations for both 802.11a with FHSS and 802.11b with DSSS based PHY layers. The parameters used in the simulation are mentioned in the respective chapter.

A.2 Simulations with *atisim* simulator

The original NS-2 simulator does not support the PCF based medium access. Although, a patch for NS-2.1b8 to allow the PCF based medium access was reported by Anders Lindgren [156], it is no longer supported (the project discontinued) and can not be used with the later versions of NS-2 (the current version is 2.34 at the time of writing this thesis). While used in conjunction with the targeted version NS-2.1b8a, the PCF implementation in [156] was found to incorrectly poll and wrongly respond to association requests. PCF implementation is also reported for GloMoSim simulator[11] in [157–159], however, when the author of this thesis contacted the authors of [157–159], some authors [158, 159] did not reply and the other [157] did not have the code any more. PCF based medium access is currently supported only by the OPNET simulator[12] which is sold commercially. At the time of writing this dissertation, it is identified by exhaustive search that no current open source network simulator supports the PCF based medium access mechanism. Therefore, the author of this thesis designed and developed a custom network simulator following a similar approach that is used in NS-2 in terms of node architecture. This custom simulator is titled as *atisim* and is under process to be published as a stand-alone open source simulator.

The client node architecture used in the *atisim* simulator is shown in Fig. A.1 where the application layer, queue and the client version of the PCF compatible 802.11 MAC layer are shown. The MAC layer is connected to the channel through a PHY layer (not shown) and follows the PCF mechanism as outlined in the IEEE 802.11 standard [64]. The RTP, UDP and IP layers can be plugged in or removed as their primary purpose is to manage sessions and packet fragmentation. For small packets used in VoIP calls, packet fragmentation can be ignored. However, we consider these layers in our simulations due to the overhead incurred by the protocol headers.

The architecture of the access point version of the IEEE 802.11 compliant PCF MAC is illustrated in Fig. A.2. The access point maintains applications and queues for each client node. Although not shown in Fig. A.2 for space constraint, the

[10] http://www.isi.edu/nsnam/ns/ns-lists.html
[11] http://pcl.cs.ucla.edu/projects/glomosim
[12] http://www.opnet.com

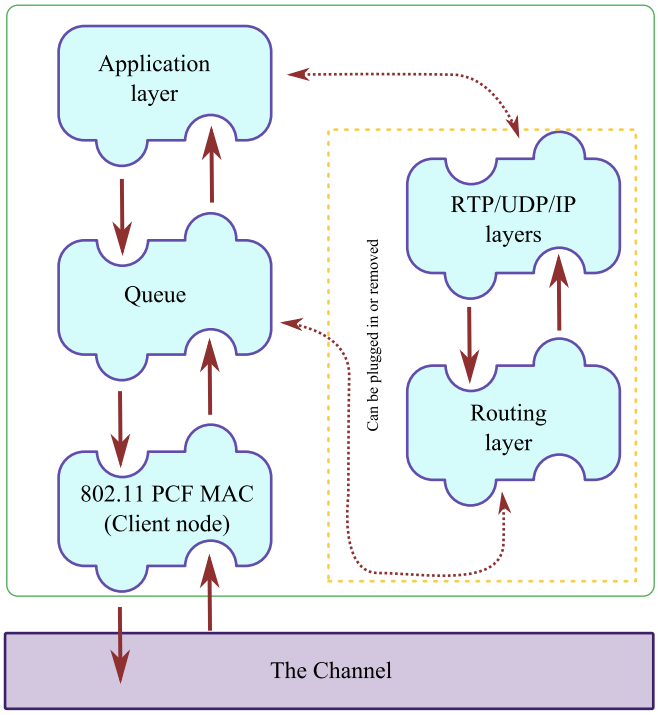


Figure A.1: Client node architecture used in *atisim* simulator showing different protocol layers which can be plugged in or removed.

RTP/UDP/IP layers are considered in our simulations. The AP MAC transmits Beacon frames to acquire control of the channel and coordinates channel access by polling each node.

A detail description of the *atisim* simulator is out of the scope of this thesis. However, we present the class diagrams of the components of a client and access point node in Figs. A.3 and A.4, respectively, for interested readers. The scheduling mechanism is at the heart of any simulation system. The class diagram of the scheduling mechanism used in *atisim* is shown in Fig. A.5. Readable names are used for the classes and their attributes, and standard UML notations are used for easy understanding.

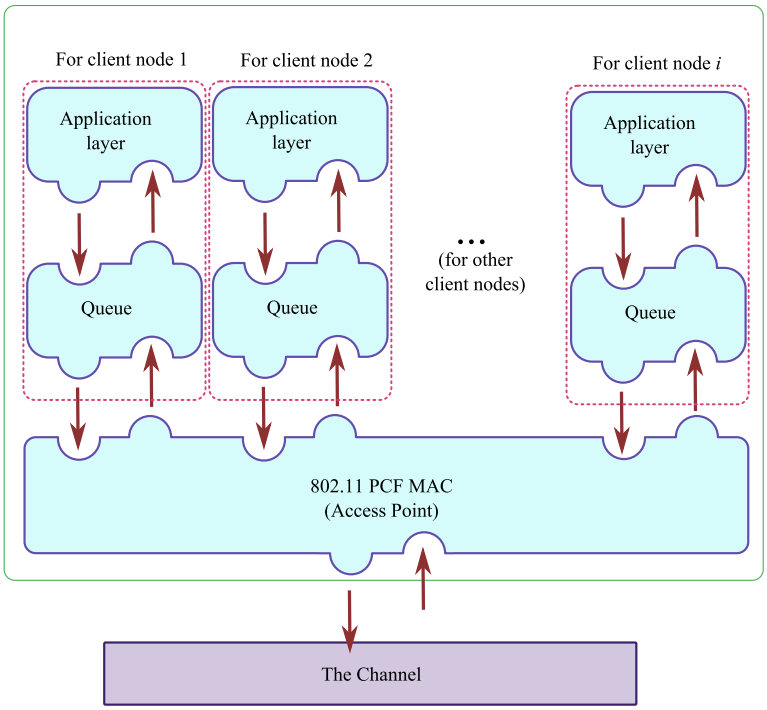


Figure A.2: Access Point architecture used in the *atisim* simulator showing the protocol layer stacks for each client node connected to the access point of the IEEE 802.11 PCF MAC layer.

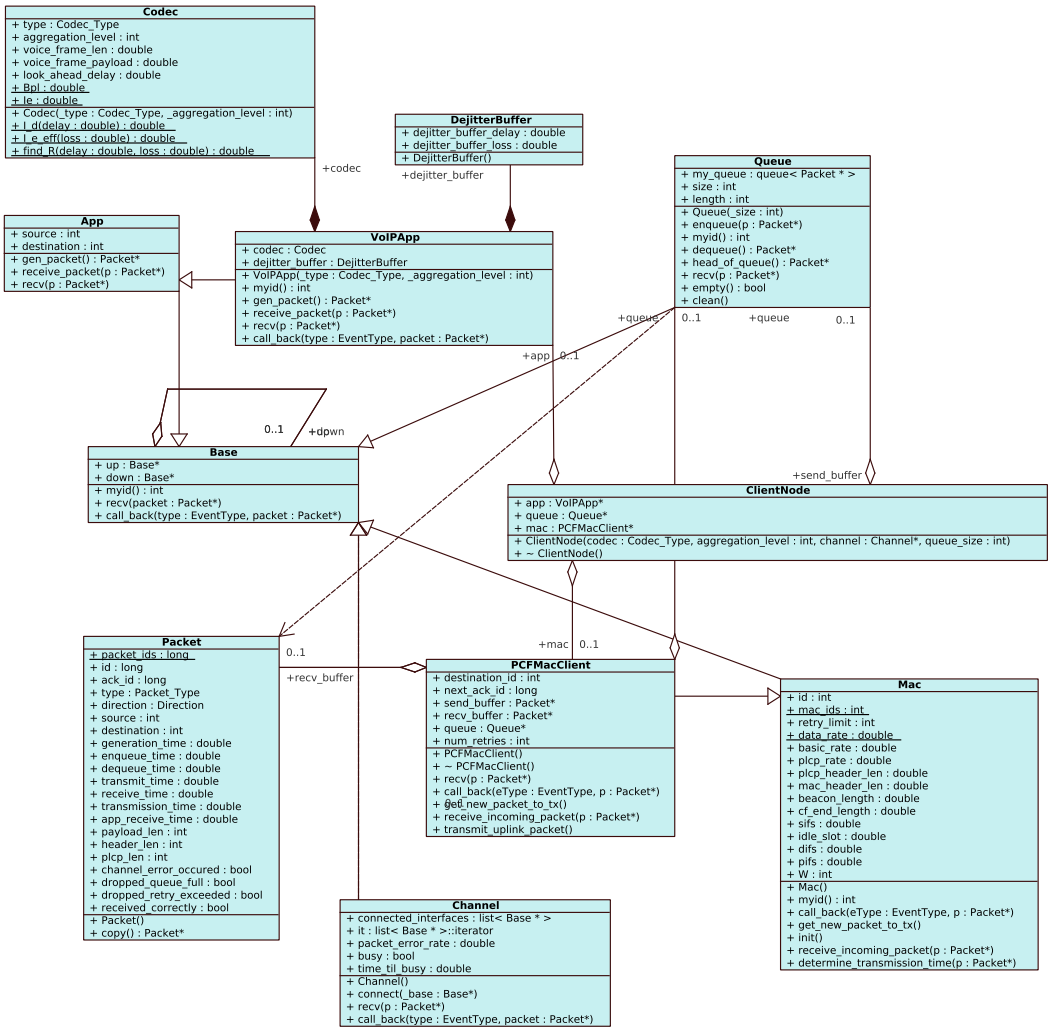


Figure A.3: Class diagram of the components used in a ***client*** node. Readable names are used for classes and their attributes, and standard UML notations are used to show the relationships between different components for easy understanding.

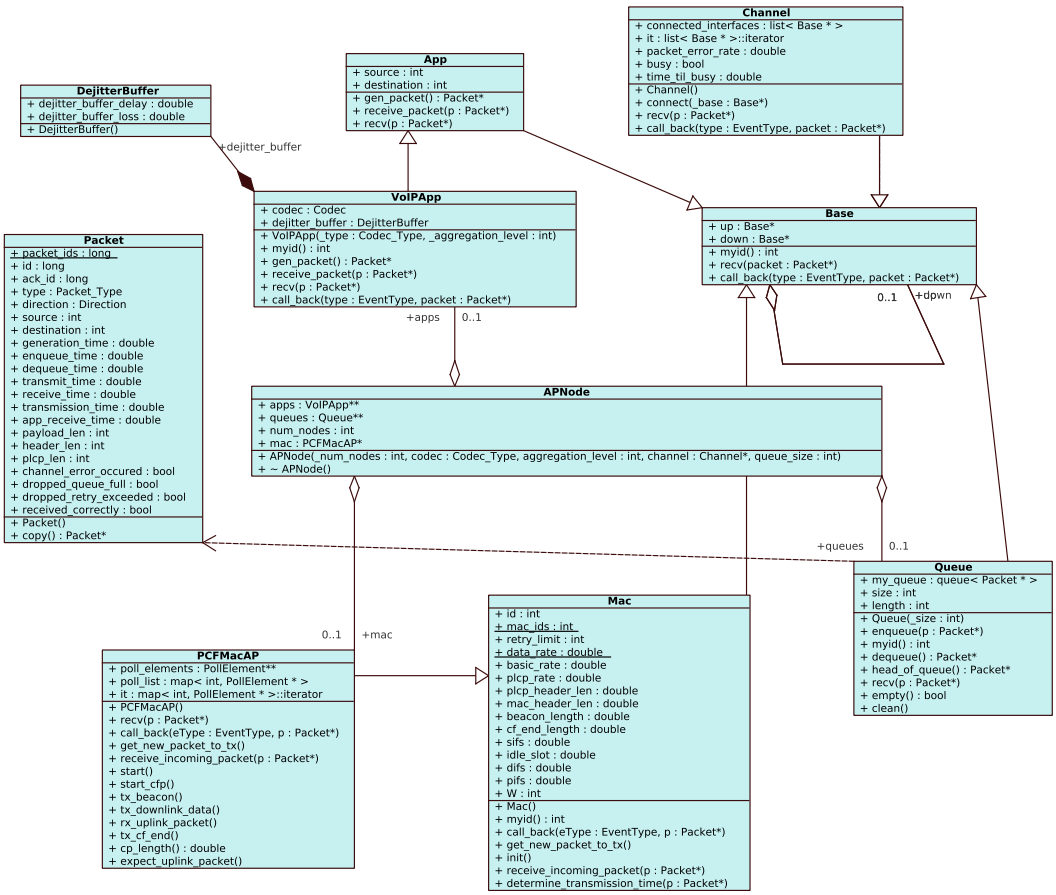


Figure A.4: Class diagram of the components used in an ***access point***. Readable names are used for classes and their attributes, and standard UML notations are used to show the relationships between different components for easy understanding.

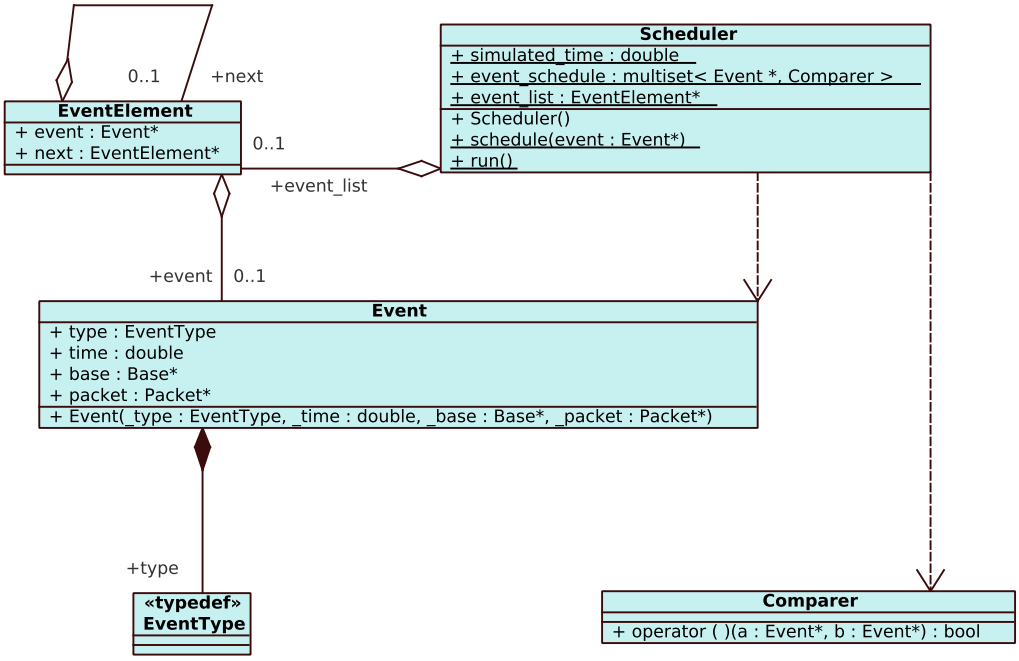


Figure A.5: Class diagram of the components used in the ***scheduling mechanism*** of the ***atisim*** simulator. Readable names are used for the classes and their attributes, and standard UML notations are used to show the relationships between different components for easy understanding.