Titre: Title:	Quality of Service Provisioning for Voice Application over Wlans
Auteur: Author:	Saeed Ghazanfari Rad
Date:	2010
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Ghazanfari Rad, S. (2010). Quality of Service Provisioning for Voice Application over Wlans [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/305/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/305/
Directeurs de recherche: Advisors:	Brunilde Sanso, & Jean-François Frigon
Programme: Program:	Génie électrique

UNIVERSITÉ DE MONTRÉAL

QUALITY OF SERVICE PROVISIONING FOR VOICE APPLICATION OVER WLANS

SAEED GHAZANFARI RAD
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE
MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE ÉLECTRIQUE)
AVRIL 2010

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

QUALITY OF SERVICE PROVISIONING FOR VOICE APPLICATION OVER WLANS

présenté par : M. GHAZANFARI RAD Saeed.
en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury constitué de :

M. NERGUIZIAN Chahé, Ph.D., président.
Mme. SANSÒ Brunilde, Ph.D., membre et directrice de recherche.
M. FRIGON Jean-François, Ph.D., membre et co-directeur de recherche.
M. GIRARD André, Ph.D., membre.

Je dédie ce mémoire,
à mes très chers parents.

Acknowledgments

I would like to express my sincere gratitude to my supervisors Dr. Brunilde Sansò and Dr. Jean-François Frigon for their continuous guidance, suggestions and encouragement during my graduate studies at École Polytechnique de Montréal. I would also like to thank my committee members Dr. Chahé Nerguizian and Dr. André Girard for their helpful comments and suggestions.

I would like to thank my colleagues and friends : Dr. Mohammad Torabi, Armelle Gnassou, Arash Azarfar, Mohammad Ali Torabi, Masood Khosroshahy, Mohamed Jihed Gafsi, Ben-Wah Kuang, Diego Enrique Perea and Vida Vakilian.

The financial support of this research was provided via a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Bell University Labs program.

Finally, but most importantly, I am deeply indebted to my beloved family especially my parents for their unconditional and continuous love and encouragement through these many years.

Résumé

Aujourd'hui, les réseaux locaux sans fils WLANs (en anglais Wireless Local Area Networks) sont de plus en plus déployés en raison de leur facilité d'installation et de leur faible coûts. Le standard IEEE (en anglais Institute of Electrical and Electronics Engineers) 802.11 définit les caractéristiques des réseaux locaux sans fil WLANS. Alors que la première norme proposée pouvait soutenir un débit de données jusqu'à 2 Mbps, dans les versions récentes de la norme 802.11, des débits de données pouvant atteindre jusqu'à 200 Mbps seront pris en charge dans les réseaux sans fil de prochaine génération. Garantir les besoins de QoS (en anglais Quality of Service) est un défi considérable pour les réseaux WLAN, en particulier pour les applications multimédia. Vu que largeur de bande disponible dans les réseaux sans fil est limitée, la bande passante supportant la QoS ne peut être facilement augmentée. Cependant, les protocoles efficaces capables de satisfaire la qualité de service doivent être conçus pour améliorer l'utilisation des ressources dans les réseaux. Le protocole de la couche MAC (en anglais Medium Access Control) affecte fondamentalement les paramètres QoS. Le contrôle d'admission est également un élément essentiel pour la qualité de service dans les réseaux locaux sans fil.

Ca projet a pour objectif de modéliser et analyser la couche MAC et contrôler l'admission dans des réseaux locaux sans fil 802.11 avec infrastructure basée sur le mécanisme DCF (en anglais Distributed Coordination Function). Bien que notre objectif général est de garantir les besoins de QoS des applications multimédias sur les réseaux sans fil, nous avons répondu à plusieurs questions importantes telles que la modélisation de la couche MAC, l'évaluation de la QoS et le contrôle d'admission.

La première contribution de cette recherche est de proposer un cadre analytique qui prend en compte la direction du trafic en mode infrastructure non saturé. Contrairement aux recherches antérieures, dans l'analyse proposée la probabilité de collision d'un paquet transmis par chaque station sans fil en liaison ascendante est différente de la probabilité de collision pour les paquets qui sont transmis à partir du point d'accès en liaison descendante. Ce modèle de distinction entre les modèles backoff par station en liaison ascendante et en liaison descendante est capable d'exprimer la performance MAC en termes de nombre de stations sans fil et plusieurs paramètres système tels

que la taille de la fenêtre de contention, le nombre maximum d'étapes backoff, la taille du tampon à la couche MAC, ainsi que les paramètres de trafic tels que les durées de conversation, le silence et le taux d'arrivée. Contrairement aux études précédentes, nous appliquons deux groupes d'équations, un groupe est défini pour la station sans fil et l'autre pour le point d'accès. Il y a trois équations qui présentent la probabilité de transmission, la probabilité de collision et la probabilité d'être dans l'état occupé. Ces équations sont en fonction du nombre de stations sans fil, du taux d'arrivée du trafic et des paramètres du système tels que la taille de la fenêtre de contention et le nombre maximum de retransmissions.

La deuxième contribution est la dérivation et l'évaluation des paramètres de QoS. Après la résolution d'un ensemble d'équations non linéaires obtenues à partir du modèle proposé MAC et en adoptant un modèle de files d'attente M/M/1/K simple et précis pour une station, nous évaluons les paramètres de QoS (délai des paquets, la perte de paquets et le débit) de la VoIP (en anglais Voice over IP). Le temps de service de paquets qui est nécessaire pour résoudre le modèle de files d'attente M/M/1/K, est calculé dans l'analyse proposée. Nous observons que le multiplexage de la couche MAC du point d'accès entraîne une plus grande probabilité de collision et une augmentation du temps de service pour les paquets en liaison ascendante en comparaison avec la liaison descendante. Notre analyse révèle également une dégradation significative du débit descendant en raison du mécanisme asymétrique et un débordement de tampon à la couche MAC du point d'accès.

Dans le domaine du contrôle d'admission et la qualité de service, nous proposons une technique de contrôle d'admission simple centralisée basée sur un modèle pour le provisionnement des besoins de la QoS pour les applications voix. La technique de contrôle d'admission proposée est basée sur la métrique de la probabilité de coupure de délai des paquets qui joue un rôle important dans la réalisation des besoins de la qualité de service des applications sensibles au délai telles que la voix et la vidéo. Nous calculons la probabilité de coupure de délai de paquet par l'application du temps de service de paquet calculé dans le cadre d'analyse proposé dans le modèle de files d'attente M/M/1/K. Nous étudions également l'impact des caractéristiques du trafic et les paramètres du système sur le nombre maximal de connexions voix dans les réseaux 802.11b et 802.11g. L'avantage majeur de cette recherche est le cadre d'analyse qui peut théoriquement estimer la capacité de voix sans la nécessité de nombreuses simulations. Ceci sera utile de manière significative lorsque de nouveaux types de services seront déployés ou des modifications aux paramètres du réseau seront apportées. Pour valider le modèle théorique, nous avons effectué des simulations et trouvé qu'il y a une bonne concordance entre l'analyse et les résultats de simulation ns-2 qui vérifient l'exactitude du modèle proposé.

Abstract

Wireless Local Area Networks (WLANs) are widely deployed nowadays because of their low cost and convenient implementation. The Institute of Electrical and Electronics Engineers (IEEE) 802.11 series standards define the specifications for such networks. While the first proposed standard could support a data rate up to 2 Mbps, in the recent upgraded versions of the standard data rates up to 54 Mbps are achievable and up to 200 Mbps are said to be supported in next generation WLANs. An important concern in WLANs is the support of Quality of Service (QoS), specifically for multimedia applications. Because of the limited available bandwidth in wireless networks, bandwidth cannot be easily increased to support QoS. However, efficient protocols capable of providing QoS have to be designed to improve resource utilization in networks. The Medium Access Control (MAC) protocol crucially affects the QoS parameters. Admission control is also an essential element for QoS provisioning in WLANs.

Our research covers mathematical modeling and analysis of the MAC layer and admission control considering Distributed Coordination Function (DCF) in infrastructure mode of IEEE 802.11-based WLANs. While our general goal is to guarantee the QoS parameters of multimedia applications over WLANs, we address several important issues such as MAC layer modeling, QoS evaluation and admission control.

The first contribution of this research is to propose an analytical framework which takes into account the traffic direction in non-saturated infrastructure mode of WLANs. Unlike previous work, in the proposed analysis the collision probability of a packet transmitted by each wireless station in the uplink direction is different from the probability of collision for the packets transmitted from the access point in the downlink direction. Our model differentiates between per-station backoff models in the uplink and downlink and is capable of expressing the MAC performance in terms of several system parameters such as contention window size, maximum number of backoff stages, size of buffer at the MAC layer, traffic parameters such as talk and silent durations and arrival rate, as well as the number of wireless stations. In contrast to the previous studies, we apply two groups of equations, one group is defined for the wireless station and the other one for the access point. These equations represent

the transmission probability, probability of collision and the probability of being in the busy state in terms of the number of wireless stations, the traffic arrival rate and system parameters such as the size of the contention window and maximum number of retransmissions.

The second contribution is the derivation and evaluation of QoS metrics. After solving a set of nonlinear equations obtained from the proposed MAC model and by adopting a simple and accurate M/M/1/K queueing model for a tagged station, we evaluate the QoS metrics (packet delay, packet loss and throughput) of VoIP which is one of the fastest growing applications over WLANs. The packet service time which is required to solve the M/M/1/K queueing model, is calculated in the proposed analysis. We observe that the multiplexing at the MAC layer of the access point results in higher collision probability and service time for packets in the uplink direction in comparison with the downlink direction. Our analysis also reveals a significant downlink throughput degradation due to the asymmetric contention mechanism and a buffer overflow at the MAC layer of the access point.

In the area of admission control and QoS provisioning, we propose a simple centralized model-based admission control technique for provisioning QoS requirements to voice applications. The proposed admission control technique is based on the packet delay outage probability metric which plays an important role in fulfilling the QoS requirements of delay sensitive applications such as voice and video. We compute the packet delay outage probability by applying the packet service time calculated in the proposed analytical framework in the M/M/1/K queueing model. We also investigate the impact of the traffic specifications and system parameters on the maximum number of voice connections in 802.11b and 802.11g networks. The major advantage of this research is the analytical framework which can theoretically estimate the voice capacity without the need of numerous simulations. This is significantly helpful when we deploy new traffic specifications or different system parameters according to the new versions of the 802.11 networks. To validate the theoretical model, we performed simulations and found that there is a good agreement between the analysis and ns-2 simulation results which verifies the accuracy of the proposed model.

Condensé

Les réseaux sans fil jouent un rôle important dans les systèmes de communication d'aujourd'hui comme une alternative aux réseaux câblés, pour offrir un accès à Internet presque n'importe où et n'importe quand. Actuellement, il existe différents types de réseaux sans fil tels que les réseaux locaux sans fil, WMANs (en anglais Wireless Metropolitan Area Networks) et WWANs (en anglais Wireless Wide Area Networks). Dans cette recherche, nous étudions les performances de WLANs qui peuvent être définis comme des réseaux sans fil qui couvrent une petite zone géographique (par exemple, les centres commerciaux, aéroports, etc.) et donnent accès à différents utilisateurs finaux tels que les ordinateurs portables, les téléphones cellulaires, les assistants numériques personnels PDA (en anglais Personal Digital Assistants), etc.

Garantir les besoins de QoS dans les réseaux locaux sans fil est un sujet d'intérêt récent à cause de la croissance exceptionnelle des réseaux locaux sans fil ces dernières années. La technique la plus populaire pour les réseaux locaux est basée sur le standard IEEE série 802.11, et depuis sa proposition plusieurs chercheurs se sont concentrés sur sa modélisation et l'évaluation de ses performances. Il existe deux méthodes d'accès dans IEEE 802.11 nommée DCF (en anglais Distributed Coordination Function) et PCF (en anglais Point Coordination Function). PCF est un mécanisme d'accès centralisé qui donne une partie fixe de ressources pour les différents flux de trafics. Toutefois, PCF est rarement utilisé en raison de sa complexité. Alors que PCF est optionnel, DCF est le mécanisme d'accès fondamental et le plus populaire dans WLAN 802.11. DCF est un protocole de distribution qui ne garantit pas la qualité de service pour le trafic multimédia et ne supporte que le trafic de type meilleur effort. Puisque l'avenir du succès des réseaux locaux sans fil dépend fortement de leur capacité à garantir les besoins de qualité de service à différents flux de trafic, la compréhension, la modélisation et la conception de protocoles efficaces sont extrêmement importantes. Ce qui motive l'orientation de ce travail pour la modélisation, l'analyse et la simulation de la couche MAC et d'enquêter sur la façon dont elle affecte les paramètres de QoS offerts par les réseaux locaux sans fil IEEE 802.11 en mode infrastructure basée sur le mécanisme DCF.

Ces dernières années, une extension des travaux de recherches traitant les techniques de contrôle d'admission dans les réseaux locaux sans fil a été réalisée, principalement en raison de l'impact de manière significative du contrôle d'admission sur la qualité de service. Les méthodes de contrôle d'admission peuvent être divisées en deux grandes catégories, les techniques basées sur des modèles et les techniques basées sur la mesure. Dans le contrôle d'admission basées sur un modèle, la décision de refuser ou d'accepter une nouvelle connexion est basée sur l'évaluation analytique des performances du réseau. Toutefois, dans les techniques basées sur la mesure, l'état du réseau est mesuré en continu pour prendre la décision de refuser ou d'accepter une nouvelle connexion. Dans cette étude, nous adressons un schéma de contrôle d'admission basé sur un modèle pour des réseaux WLANs 802.11.

Ce projet de recherche porte sur la garantie des besoins de QoS pour des usagers dans des réseaux locaux sans fil en mode infrastructure basée sur le mécanisme d'accès DCF à la couche MAC qui fournit un service meilleur effort et ne prend pas en charge des applications multimédias évoluées caractérisées par des contraintes de QoS particulières. En raison de la largeur de bande limitée disponible dans les réseaux sans fil, l'augmentation de la bande passante pour la qualité de service n'est pas facile et trop coûteuse. C'est la raison principale qui motive un grand nombre de chercheurs à concevoir des protocoles efficaces qui permettent le provisionnement des besoins de qualité de service pour différentes applications, tout en améliorant l'utilisation des ressources du réseau.

L'objectif général de cette recherche est la garantie des besoins de QoS pour les applications multimédia. La couche MAC affecte de manière significative les performances du réseau. Modélisation, analyse et simulation du protocole MAC 802.11 et l'obtention d'une bonne compréhension de la façon dont il fonctionne sont donc essentiels pour améliorer les performances du réseau. Notre premier objectif est de modéliser le mécanisme d'accès DCF à la couche MAC du WLAN qui fonctionne en mode infrastructure où les noeuds sans fil ne peuvent communiquer que par l'intermédiaire du point d'accès central. Le contrôle d'admission joue également un rôle important dans la réalisation des besoins de qualité de service des applications multimédias sur les réseaux sans fil. Notre deuxième objectif est alors de proposer un schéma de contrôle d'admission à base de modèle pour garantir les besoins de QoS des applications différentes.

Dans cette thèse, nous proposons un nouveau cadre d'analyse pour évaluer le temps de service de paquets et le débit basé sur la modélisation de la couche MAC pour les réseaux locaux sans fil non-saturés dans le mode infrastructure. Ce modèle est capable de différencier entre la liaison ascendante et descendante du trafic qui est nésessaire pour étudier l'effet de multiplexage à la couche MAC du point d'accès. Il y a trois équations non linéaires attribuées à chaque station dans le domaine sans fil. Ces équations représentent la probabilité de collision, la probabilité de transmission et la probabilité d'être dans l'état occupé en termes de nombre de stations sans fil, du taux d'arrivée du trafic et des paramètres système tels que la taille de la fenêtre de contention, le nombre des étapes du backoff et la taille du tampon de la couche MAC. En résolvant simultanément les équations non linéaires en utilisant des méthodes numériques, on trouve les probabilités de collision vu par les paquets en liaison ascendante et descendante. On trouve aussi d'autres paramètres importants tels que le temps de service de paquets et le débit. Des simulations ns-2 sont utilisées pour vérifier l'exactitude du modèle proposé. Nous obtenons aussi les paramètres de QoS tels que le délai moyen de paquets et le débit du réseau d'après un modèle simple M/M/1/K qui a été adopté pour analyser les files d'attente à la couche MAC.

Finalement, nous proposons une méthode de contrôle d'admission centralisée à base de modèle qui peuvent être mises en oeuvre au point d'accès pour garantir les paramètres QoS. Le contrôle d'admission est basée sur la définition de la probabilité de coupure de délai de paquet qui est la probabilité que le paquet ait un délai moyen qui dépasse un seuil prédéfini. Pour les applications sensibles au délai telles que la VoIP, il est très important que les paquets ne subissent pas un délai de bout en bout qui dépasse le délai acceptable. Le schéma de contrôle d'admission proposé prédit la probabilité de panne avant d'admettre ou de refuser une nouvelle connexion de voix. Selon le modèle, en admettant que le nouvel appel viole les contraintes de QoS, le point d'accès refuse le nouvel appel. L'implémentation de schéma de contrôle d'admission à base de modèle au point d'accès permettra d'éviter la dégradation de la qualité de la voix et la congestion du réseau.

La thèse est organisée comme suit. Le Chapitre 1 présente les objectifs et les principales contributions de ce mémoire. Après un bref aperçu des principes des réseaux locaux sans fil, des couches MAC et PHY dans le Chapitre 2, nous présentons le modèle du système, l'architecture des réseaux câblés et sans fil et les modèles des trafics (données et voix) au Chapitre 3. Nous portons notre attention aux réseaux constitués de stations sans fil et d'un point d'accès relié à un réseau filaire. Nous considérons un réseau câblé/sans fil dans lequel les connexions vocales existent entre chaque station sans fil dans le domaine sans fil et une station filaire dans le domaine filaire. Par conséquent, le nombre de flux de voix active est égale au nombre de stations sans fil. Le trafic voix est modélisé par un processus `on-off` qui comprend deux états.

Les revues de la littérature liées à l'analyse des performances de couche MAC et d'amélioration des performances sont présentées dans le Chapitre 4. En fait, les travaux antérieurs sont divisés en deux grandes catégories. La premiére contient des articles liés à la modélisation de la couche MAC en mettant l'accent sur la modélisation de la procédure de backoff et le déploiement des modèles de files d'attente différents pour obtenir un temps de service de paquets et le débit du réseau. La deuxième catégorie décrit des travaux antérieurs dans le domaine de l'amélioration des performances en utilisant différentes méthodes telles que la différenciation du trafic, la modification du protocole et les techniques de contrôle d'admission.

Nous présentons le nouveau modèle d'analyse de la couche MAC et la dérivation de paramètres de QoS dans le Chapitre 5. Contrairement aux recherches antérieures, nous proposons une analyse plus détaillée par la différence entre le modèle backoff pour la liaison ascendante et descendante. Bien que chaque station sans fil et le point d'accès utilisent la même procédure de backoff pour accéder à la bande passante partagée, la probabilité de transmission par paquet est différente parce qu'ils n'ont pas le même taux d'arrivée et la période d'inactivité dans chaque station sans fil est beaucoup plus importante en comparaison avec les périodes d'inactivité dans le point d'accès. Cela affectera également la formulation du backoff en raison de l'existence de l'état de repos et de la probabilitié de transition qui dépendent de la quantité du taux d'arrivée du la trafic. Contrairement aux études précédentes, nous appliquons deux groupes d'équations, un groupe est défini pour la station sans fil et l'autre pour le point d'accès. Il y a trois équations non linéaires pour chaque groupe de stations. Ces équations représentent la probabilité de transmission, la probabilité de collision et la probabilité d'être dans l'état occupé en termes de nombre de stations sans fil, du taux d'arrivée trafic et de paramètres du système tels que la taille de la fenêtre de contention, le nombre de étapes de backoff et la taille du tampon à la couche MAC. Nous utilisons l'approximation d'arrivée de Poisson pour trouver la probabilité qu'au moins un paquet est au tampon de la couche MAC pour être transmis. Cette hypothèse est bonne puisque l'intervalle de l'arrivée constante du paquet de voix durant la période d'activité est de l'ordre de quelques millisecondes et le temps d'un slot moyen est de l'ordre de microsecondes. L'hypothèse d'arrivée de Poisson est plus réaliste au tampon de la couche MAC du point d'accès car il est la superposition de n flux de trafics de n sources indépendantes de la voix. Nous calculons plusieurs paramètres importants tels que la probabilité de collision, le temps de service de paquets, le temps d'attente des paquets et le débit après l'application de méthodes numériques pour résoudre une série d'équations non-linéaires basés sur le modèle proposé. Ce modèle est capable d'évaluer le mécanisme asymétrique de contention à la couche MAC de réseaux locaux sans fil IEEE 802.11. Le multiplexage du trafic de la liaison descendante à l'AP ne penche pas en faveur du trafic de liaison ascendante à partir de stations sans fil. Cependant, il provoque de plus grands probabilité de collision et de temps de service par paquets dans la liaison ascendante. Il affaiblit considérablement le débit descendant en raison du débordement de tampon.

Nous utilisons des simulations ns-2 pour vérifier l'exactitude du cadre analytique proposé. Enfin, une simple méthode de contrôle d'admission centralisé basée sur ce modèle est proposée pour garantir les besoins de QoS des applications vocales existantes. La technique du contrôle d'admission proposée est basée sur la métrique de probabilité de coupure du délai de paquet. Cette métrique joue un rôle important pour remplir les besoins de qualité de service des applications sensibles au délai telles que la voix et la vidéo. Nous calculons la probabilité de coupure du délai des paquets en appliquant le temps de service calculé dans le cadre d'analyse proposé dans le modèle de files d'attente M/M/1/K. En appliquant cette technique de contrôle d'admission basée sur un modèle, le point d'accès peut prédire la probabilité de coupure du délai de paquet avant d'admettre un nouvel appel. Autrement dit, le point d'accès refuse ou admet la nouvelle connexion après avoir résolu le modèle proposé pour l'analyse de la couche MAC et avoir évalué la probabilité de coupure du délai de la voix. Le contrôle d'admission peut être mis en oeuvre sur le point d'accès pour prévenir la congestion du réseau et la dégradation de la qualité vocale. Nous étudions également l'impact des caractéristiques du trafic et des paramètres du système sur le nombre maximal de connexions vocales dans les réseaux 802.11b et 802.11g.

Nous concluons la recherche dans le Chapitre 6. Les futurs travaux sur l'analyse de la couche MAC, peuvent conduire à la coexistence de différentes applications telles que les données, voix et vidéo sur des réseaux sans fil. Cela peut se faire en modifiant le cadre d'analyse proposé pour y inclure une nouvelle série d'équations adaptée au type de trafic ainsi que de nouvelles expressions de la probabilité d'être dans l'état occupé en fonction de la modélisation du trafic de la nouvelle application ou de la nouvelle norme de codec. Il serait aussi intéressant de modifier le modèle de Markov de backoff afin de considérer les erreurs induites au canal sans fil. Un paquet est retransmis s'il est corrompu peu importe la source de corruption que ce soit à cause des erreurs induites par le canal ou à cause d'une collision. En déployant des modèles plus complexes de file d'attente, nous pouvons étendre le cadre d'analyse théorique proposé à trouver le nombre maximum de flux de données, de voix ou de vidéo supporté par le réseau WLAN 802.11. Enfin, nous pouvons étudier la taille optimale de la fenêtre de contention, le nombre optimal de retransmissions maximales, les périodes `on` et `off` , etc. pour atteindre le débit maximum, tout en la satisfaisant les besoins de QoS pour différentes applications.

Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

AC	Access Categories
ACK	Acknowledgment
AIFS	Full Duplex
AP	Access Point
BSS	Basic Service Set
CAC	Connection Admission Control
CAF	Channel Access Functions
CARC	Call Admission and Rate Control
CBR	Constant Bit Rate
CCK	Complementary Code Keying
CFP	Contention Free Period
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTS	Clear to Send
CW	Contention Window
DCF	Distributed Coordination Function
DIFS	Distributed Interframe Space
DS	Distributed System
DSSS	Direct Sequence Spread Spectrum
EDCA	Enhanced Distributed Channel Access
ESS	Extended Service Set
FHSS	Frequency-Hopping Spread Spectrum
FSMC	Finite State Markov Chain
HCCA	Enhanced Distributed Channel Access
HCF	Hybrid Coordination Function
HR/DSSS	High-rate Direct Sequence Spread Spectrum
IP	Internet Protocol
MAC	Medium Access Control
NAV	Network Allocation Vector

OFDM	Orthogonal Frequency-Division Multiplexing
PC	Point Coordinator
PCF	Point Coordination Function
PDA	Personal Digital Assistant
PDF	Probability Distribution Function
PGF	Probability Generating Function
PHY	Physical
PIFS	Point Coordination Inter Frame Space
PLCP	Physical Layer Convergence Procedure
PN	Pseudonoise
PMD	Physical Media Dependent
QoS	Quality of Service
QSTA	Quality of Service station
RTS	Request to Send
SIFS	Short Interframe Space
SNR	Signal to Noise Ratio
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VBR	Variable Bit Rate
VMAC	Virtual medium access control
VoIP	Voice over Internet Protocol
VS	Virtual Source
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WWAN	Wireless Wide Area Network

List of Notations

λ	Arrival rate during "on" period
T	Constant interarrival time during "on" period
on	Average length of the "on" period
off	Average length of the "off" period
λ_u	Arrival average rate into the MAC layer buffer of the wireless station
λ_d	Arrival average rate into the MAC layer buffer of the access point
n	Number of wireless stations
p_u	Uplink collision probability
p_d	Downlink collision probability
τ	Probability that the wireless station transmits a packet
τ_u	Probability that the wireless station transmits a packet in the uplink stream
τ_d	Probability that the access point transmits a packet in the downlink stream
$s(t)$	Stochastic process that represents the backoff stage
$b(t)$	Stochastic process that represents the backoff time counter
m	Maximum backoff stage
m'	Maximum number of retransmissions
q	Probability that there is at least one packet in the buffer
q_u	Probability that there is at least one packet in the buffer of the wireless station
q_d	Probability that there is at least one packet in the buffer of the access point
W_0	Size of minimum contention window
W_i	Size of contention window for the i^{th} stage of the backoff procedure
W_m	Size of maximum contention window
σ	Slot time
P_{tr}	Probability that at least one station transmits
P_s	Probability of a successful transmission given that at least one station transmits
$P_{u,s}$	Probability that a transmission is successful in the uplink direction
$P_{d,s}$	Probability that a transmission is successful in the downlink direction
$SIFS$	Length of the short interframe space
$DIFS$	Length of the distributed interframe space
$E[P]$	Packet payload

H_{PHY}	Length of the physical header
H_{IP}	Length of the IP header
H_{MAC}	Length of the MAC header
δ	Propagation delay
R_b	Basic bit rate
R_c	Channel bit rate
$E[slot]$	Average slot time
T_c	Average time the channel is sensed busy during a collision
T_s	Average time the channel is sensed busy because of a successful transmission
$\bar{B}^i$	Average backoff defer period
$\bar{T}^i$	Time needed for a successful transmission
$1/\mu$	Average packet service time
$1/\mu_u$	Uplink average packet service time
$1/\mu_d$	Downlink average packet service time
ρ	Utilization factor
$\bar{N}$	Average number of packets in the buffer
K	Buffer size (in number of packets)
$\bar{d}$	Average packet delay
P_{loss}	Blocking probability
$\bar{d}_u$	Uplink average packet delay
$\bar{d}_d$	Downlink average packet delay
$\bar{w}_u$	Uplink average packet waiting time
$\bar{w}_d$	Downlink average packet waiting time
ρ_u	Uplink utilization factor
ρ_d	Downlink utilization factor
S_u	Overall uplink throughput for one AP
S_d	Overall downlink throughput for one AP
S_t	Overall network throughput for one AP
T_0	Acceptable threshold delay
X	Length of the "`on`" period
Y	Length of the "`off`" period
Z	Length of the interarrival time
α	Parameter of the exponential random variable X
β	Parameter of the exponential random variable Y

$E[N]$	Average number of packet arrivals during the "`on`" period
$E[Z]$	Average interarrival time
$u(t)$	Unit step function
$\delta(t)$	Unit impulse function

Chapter 1

Introduction

1.1 Statement of the research problem and motivations

Wireless networks play an important role in today's communication world as an alternative to wired networks for providing access to the Internet almost for any user, anywhere and any time. Currently, there exist different types of wireless networks such as WLANs, Wireless Metropolitan Area Networks (WMANs) and Wireless Wide Area Networks (WWANs). In this research, we study the performance of WLAN which can be defined as wireless networks that cover a small geographic area (e.g., building, shopping centers, airports, etc.) and provide access to different end users such as laptops, cell phones, Personal Digital Assistants (PDAs), etc.

Providing QoS in WLANs is a recent topic of interest because of the outstanding growth of WLANs in recent years. The most popular technique for LANs is based on the IEEE 802.11 standard series and since its proposal several researchers have focused on its modeling and performance evaluation. There are two access methods in IEEE 802.11 named distributed coordination function (DCF) and point coordination function (PCF). PCF is a centralized access mechanism which provides fixed portion of resources for different traffic streams. However, PCF is rarely used because of its complexity. While PCF is optional, DCF is the fundamental and most popular access mechanism in 802.11 WLANs. DCF is a distributed protocol that does not guarantee QoS for multimedia traffic, but only supports best effort service. Since the future success of WLANs depends significantly on their capability to provision the acceptable QoS requirements to different traffic streams, the understanding, modeling and design of efficient protocols is critically important. This motivates the orientation of this work towards the modeling, analysis and simulation of the MAC layer and investigating how it affects the QoS parameters of the infrastructure mode IEEE 802.11-based WLANs in the DCF mode.

In recent years, there has also been extensive work dealing with the admission control techniques in WLANs mainly due to the significant impact of admission control on QoS provisioning. Admission control methods can be divided into two major categories, model-based and measurement-based techniques. In model-based admission control, admitting or denying a new connection is based on the analytical assessment of the network's performance. However, in measurement-based scheme, the network condition is continuously measured to make the decision of denying or accepting a new connection. In this study, we address a model-based admission control scheme for 802.11 WLANs.

1.2 Objectives

This research project is concerned with service guarantee in infrastructure mode of WLANs deploying DCF access mechanism at the MAC layer which provides best effort service and does not support advanced multimedia applications with their specific requirements. Because of the limited available bandwidth in wireless networks, increasing the bandwidth for QoS provisioning is not easy and too costly. This is the main reason which motivates a lot of researchers to design efficient protocols capable of provisioning QoS requirements for different applications, while improving the resource utilization for the network.

The general goal of our research is to provision QoS for multimedia applications. As we discussed earlier the MAC layer significantly affects the network performance. Modeling, analysis and simulation of the 802.11 MAC protocol and obtaining a good understanding of how it works are therefore vital to improve the network performance. Our first objective is to model DCF access mechanism at the MAC layer of WLAN which operates in the infrastructure mode where wireless nodes can only communicate through the central access point. We also stated that admission control plays an important role in fulfilling the QoS requirements of multimedia application over WLANs. Our second objective is then to investigate a model-based admission control scheme to guarantee the QoS requirements of different applications.

1.3 Contributions

In this thesis, we propose a novel analytical framework to evaluate packet service time and throughput based on the MAC layer modeling for unsaturated WLANs in the infrastructure mode. This model is capable of differentiating between uplink and downlink traffic which is useful to study the multiplexing effect at the access point MAC layer. There are three nonlinear equations assigned to each tagged station in the wireless domain. These equations represent the transmission probability and the probability of being in the busy state in terms of the number of wireless stations, the traffic arrival rate and system parameters such as contention window size, number of backoff stages and buffer size at the MAC layer. By solving the set of non-linear equations using numerical methods, we find the collision probabilities seen by the packets in the downlink and uplink directions. We also find other important parameters such as packet service time and throughput. ns-2 simulations are used to verify the accuracy of the proposed model. We also derive the QoS parameters such as average packet delay and throughput from solving a simple M/M/1/K model which has been adopted for queueing analysis at the MAC layer.

Finally we propose a centralized model-based admission control method which can be implemented at the access point to guarantee the QoS metrics. The admission control is based on the definition of the packet outage probability that is the probability that the packet experiences an average delay which exceeds a predefined threshold delay. For delay-sensitive applications such as VoIP, it is very important that packets do not experience an end-to-end delay which exceeds the acceptable delay. The proposed admission control scheme predicts the outage probability before admitting or denying a new voice connection. If according to the model, admitting the new call results in violating the QoS constraints, then the access point will deny the new call. Implementing the proposed model-based admission control scheme at the access point will prevent voice quality degradation and network congestion.

1.4 Outline of the thesis

The remainder of the research is organized as follows. After a brief overview of the principles of WLANs, MAC and PHY layers in Chapter 2, we describe the system model, wired/wireless network architecture and models for data and voice traffic in

Chapter 3. Literature review and previous work related to the MAC layer performance analysis and performance enhancement are presented in Chapter 4. Basically, previous work is divided into two major categories. The first one includes articles related to the MAC layer modeling with a focus on the modeling of the backoff procedure and the deployment of different queueing models to obtain packet service time and network throughput. The second category describes previous work in the area of performance enhancement using different methods such as traffic prioritization, protocol modification and admission control techniques.

We present our novel model for MAC layer analysis and derivation of QoS parameters in Chapter 5. Based on the bi-dimensional backoff model in non-saturated mode, we obtain three non-linear equations for each tagged station. Knowing the arrival rate of voice packets to the MAC layer of the access point and a tagged wireless station, we develop a set of non-linear equations which can be solved using numerical methods. Several important unknown variables such as packet collision probability, transmission probability and probability of a successful transmission can be found after solving the corresponding equations. Next we derive the expressions for packet service time, average slot time and network throughput in the uplink and downlink direction. M/M/1/K queueing model is also applied to obtain the average packet delay and packet outage probability. The later one is then used for our proposed model-based admission control scheme.

Finally, we conclude the research in Chapter 6.

Chapter 2

802.11 Wireless Local Area Networks

We briefly overview the principles of WLANs in this chapter. While there exist different designs for WLANs, we study the infrastructure mode of WLANs in this thesis. Different access mechanisms at the MAC layer of WLANs such as DCF, PCF and EDCA according to the 802.11 and 802.11e standards are discussed in this chapter. We also study the specifications of physical layer of IEEE 802.11 standard and 802.11 frame format.

2.1 Wireless Local Area Networks

IEEE 802.11-based WLANs are widely used for wireless connectivity in homes, offices and public areas because of their simple deployment, low cost and capability of supporting high data rates. The original IEEE 802.11 standard published in 1997 could support data rate of 2 Mbps [1]. A data rate of 11 Mbps is supported by the IEEE 802.11b and of 54 Mbps by the IEEE 802.11a and IEEE 802.11g, deploying different transmission techniques such as Orthogonal Frequency-Division Multiplexing (OFDM) in the PHY layer. IEEE 802.11e is concerned with traffic prioritization by introducing an enhanced channel access function in the MAC layer [2].

A traditional centralized IEEE 802.11-based local area network consists of one access point (AP) and its associated end users. This setup is also known as infrastructure mode or Basic Service Set (BSS) and can only cover a small area. To extend the coverage area, another setup is defined by connecting multiple BSSs to form an Extended Service Set (ESS) mesh network. The IEEE 802.11s defines MAC and physical specification for the ESS mesh network [3]. Figure 2.1 shows the general implementation of 802.11 wireless networks. For the first basic service set (BSS 1), station 2 (STA 2) is used as the access point and all communication between mobile stations within BSS 1, are relayed through STA 2. A distributed system (DS) is used to connect BSS 1 and BSS 2 to cover a larger area. When a frame is transmitted from STA 1 in BSS 1, to STA 4 in BSS 2, the DS will deliver the frame to STA 3 which is serving STA 4 in BSS 2.

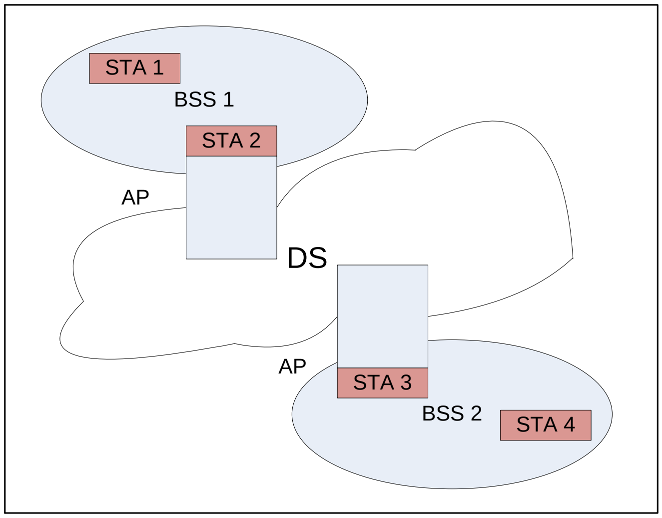

Figure 2.1 General architecture of 802.11 networks.

There are many challenges in ESS mesh networks like resource management, scalability, throughput capacity, throughput fairness, power management, MAC and PHY layer protocols, admission control, congestion control, routing, reliability, robustness and security. In this research we concentrate on answering the questions of how the MAC layer parameters affect the performance of BSS network and how to guarantee the QoS parameters of multimedia applications using admission control algorithms.

2.2 IEEE 802.11 layer reference model

Deploying radio waves as the transmission medium in 802.11 networks, makes the physical layer more complex. The IEEE 802.11 standard use the MAC and physical layers shown in Figure 2.2. The physical layer includes two sublayers known as Physical Layer Convergence Procedure (PLCP) Sublayer and Physical Media Dependent (PMD) Sublayer. The PLCP sublayer, sets up the MAC frames according to the medium which allows MAC to perform with minimum dependence on the physical characteristics of the medium. The PMD sublayer is responsible for transmitting the frames received from the PLCP sublayer.

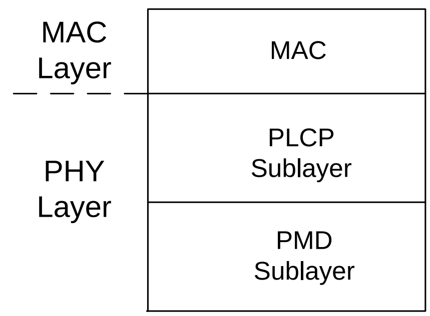

Figure 2.2 Portion of IEEE 802.11 reference model.

2.3 MAC layer

Unlike wireline networks, we cannot rely on the collision detection based access mechanism in wireless networks. In the IEEE 802.11 networks, there are two types of MAC layer access mechanisms based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA): Distributed Coordination Function (DCF) and Point Coordination Function (PCF).

In DCF mode, a station senses the medium and if it is idle for a period called Distributed Interframe Space (DIFS) it starts transmitting the frame, otherwise it waits until the medium is idle for a DIFS and then starts a backoff. The backoff period is uniformly chosen between 0 and the contention window (CW). The size of the contention window depends on the number of retransmissions. We define CW_{min} as the minimum value of the contention window (CW), and m as the maximum number of transmissions. The contention window size equals CW_{min} in the first attempt and is doubled after each transmission failure up to the maximum size of the contention window which is $CW_{max} = 2^m CW_{min}$. The station decrements its backoff counter as long as the channel is sensed idle and starts transmission when the backoff counter reaches zero. If the channel is sensed busy while the backoff is decremented, then the backoff is frozen. The backoff starts counting down again after the channel is sensed idle for a DIFS. DCF access mechanism is shown in Figure 2.3. Station 1 starts transmitting the packet after the channel is sensed idle for a DIFS. When the destination receives the packet from Station 1, it transmits an acknowledgment (ACK) after a Short Interframe Space (SIFS) interval. Station 2 has a packet for transmission and starts the backoff procedure after sensing the channel idle for a DIFS. Station 2 chooses randomly a backoff time which is equal to 8 and starts counting down the backoff counter. However, after sensing the channel busy, backoff is frozen at 4 and starts decrementing again after sensing the channel idle for a DIFS. The packet from station 2 is transmitted after the backoff time reaches zero.

A collision is inferred when there is no acknowledgment (ACK) from the receiver. There are two types of access in the DCF mode: basic access and Request to Send/ Clear to Send (RTS/CTS) access. In the basic mode as shown in Figure 2.3, an acknowledgment is sent upon successful reception of the packet and the mechanism is known as a two-way handshaking. However, in the RTS/CTS mode as shown in Figure 2.4 a four-way handshaking mechanism is used and before sending the data information, an RTS frame is sent by the transmitter and the destination responds with a CTS upon successful reception of the RTS. CTS and RTS include information about the duration of the data frames followed by them. As it can be seen in Figure 2.4, during a transmission from station 1, all other stations update a network allocation vector (NAV) which includes the information about the period of time during which, the channel will remain busy.

In the basic access mechanism the problem of the hidden terminal exists because the neighbors of the receiver may transmit when they do not sense a transmission and a collision will occur. This problem is solved in the RTS/CTS mechanism because the channel is reserved according to the information carried by CTS/RTS and the neighbors do not send a packet because they use a virtual carrier sense mechanism.

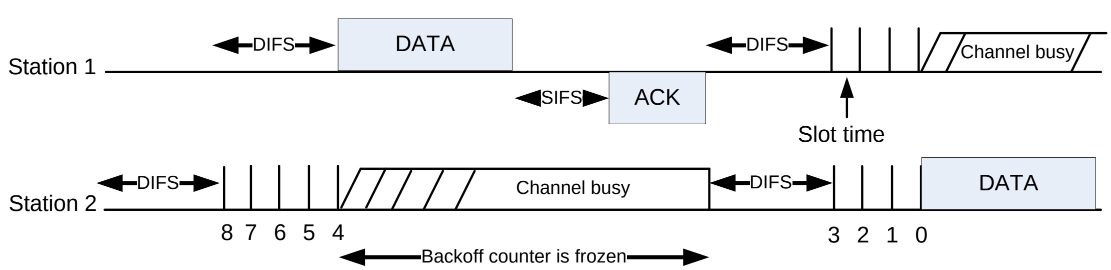

Figure 2.3 Channel access using basic mechanism.

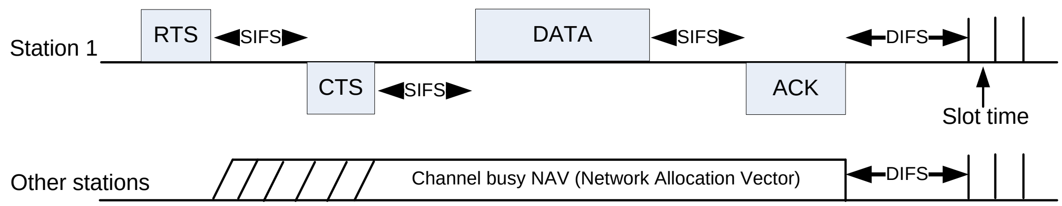

Figure 2.4 Channel access using RTS/CTS mechanism.

Another optional method to access the channel is Point Coordination Function PCF. PCF is a contention-free centralized mechanism used in the infrastructure configuration. The AP acts as a point coordinator and polls its associated stations. PCF access method is shown in Figure 2.5. A Contention Free Period (CFP) is the time during which a point coordination function is used to give the transmission right to the stations of the basic service set. The CFP starts with a beacon frame and ends with a CF-END frame. Access point periodically broadcasts beacons to announce the parameters of the contention free period such as the maximum duration of the contention free period and the number of time units remaining in the current contention free period. As it can be seen in Figure 2.5, $DATA + POLL$ frame is used by the access point to send data to a mobile station and request data from the mobile station which is in the polling list. The mobile station will transmit the $DATA + ACK$ frame after SIFS interval. This would be followed by a $DATA+ACK+POLL$ frame used by the access point to acknowledge the receipt of data, transmit the new data and announce the next station in the polling list. PC polls the next station in the polling list if it does not receive an ACK after a Point Coordination Inter Frame Space (PIFS) interval. During CFP, all stations know that the medium is busy and set their Network Allocation Vector (NAV). Data is transferred between CFP-Pollable stations and AP according to the polling list of the Point Coordinator (PC). There is also a SIFS interval between all transmissions. The contention-free period is finished when the access point transmits the CF-END frame to release stations from PCF mode and begin the contention period using DCF access mode. Because of its complexity and implementation problems, PCF is rarely supported by commercial devices.

Although IEEE 802.11-based wireless LANs with DCF access mechanism are widely used because of their low cost, convenience and easy deployment, they can

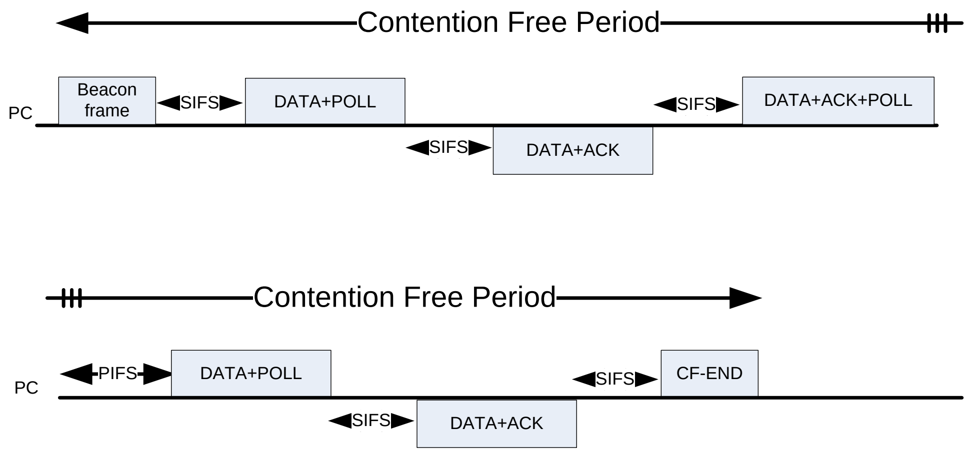

Figure 2.5 Channel access using PCF machanism.

only support best effort traffic. With the increasing emergence of multimedia applications, supporting QoS in such networks becomes an important task. Hybrid Coordination Function (HCF) is defined by the enhancement task group (TGe) to support QoS in IEEE 802.11-based WLANs [2]. HCF integrates Enhanced Distributed Channel Access (EDCA) which is a distributed channel access similar to DCF and HCF controlled channel access (HCCA) which is a centralized mechanism similar to PCF.

Traffic transmission in DCF is based on a first come first serve policy and there is no priority among the different stations. Therefore, the AP, which might require more throughput, has the same priority as the other stations. However, EDCA considers service differentiation using four different queues for their corresponding access categories (ACs). Therefore, there are four different channel access functions (CAFs) in each QoS station (QSTA).

Each CAF works like a DCF with its particular contention window size and short interframe space. A CAF_i senses the channel for its arbitration interframe space $AIFS_i$ and transmits if the channel is sensed idle for a period equal to $AIFS_i$. If the channel is busy, CAF_i continues sensing the channel and when the channel is sensed idle for $AIFS_i$, it starts the backoff process with initial backoff counter chosen randomly among a uniformly distributed variable in the range of 0 and CW_i. Similar to the DCF, the initial window size is CW^i_{min} and is doubled after each transmission attempt until it reaches the maximum size of the window CW^i_{max}. There is also another parameter known as the Transmission Opportunity (TXOP) limit which is adjusted by each CAF.

A difference between DCF and CAF_i is that CAF_i is permitted to transmit consecutive frames after winning the contention to access the channel provided that the transmission duration does not exceed the $TXOP_i$ limit. The EDCA channel access scheme which includes four access categories, is shown in Figure 2.6.

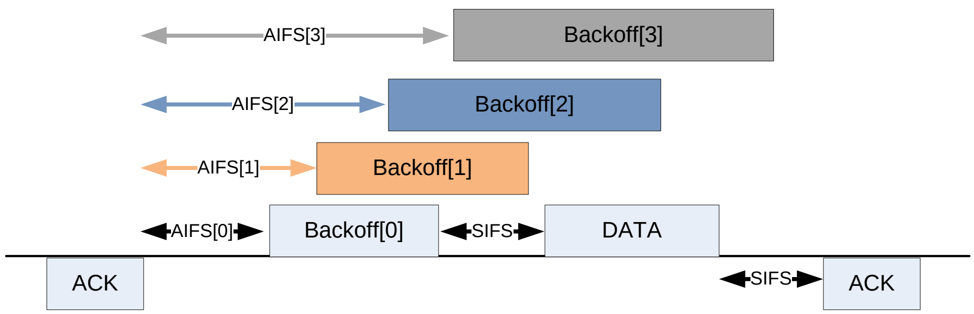

Figure 2.6 Channel access using EDCA machanism.

2.4 Physical layer

The specifications of the physical layer of the IEEE 802.11 standard can be found in [1]. Five different methods for frame transmission are defined in the standard. Frequency-Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), High-rate Direct Sequence Spread Spectrum (HR/DSSS), Orthogonal Frequency-Division Multiplexing (OFDM) and Infrared (IR). The first three spread spectrum techniques operate in the 2.4 GHz band, OFDM operates in 2.4 GHz and 5 GHZ band and IR operates in 300-428 GHz. The carrier signal of FHSS hops from one frequency to another among many frequency channels. Using a hopping sequence, a pseudorandom hopping pattern is generated which includes all frequency channels.

A transmission mode in the IEEE 802.11 networks defines the sender characteristics such as the modulation type, the encoding rate and the transmission rate. Table 2.1 includes eight different PHY modes existing in the 802.11a/g standard.

The maximum rate of data transmission is 2 Mbps when using the FHSS technique. The data transmission rate is 2 Mbps in DSSS where the original data signal is multiplied by a pseudonoise (PN) sequence which typically has a higher rate than the

Table 2.1 PHY layer modes of the IEEE 802.11a/g standard.

Mode	Modulation	Coding rate	Data rate
1	BPSK	1/2	6 Mbps
2	BPSK	3/4	9 Mbps
3	QPSK	1/2	12 Mbps
4	QPSK	3/4	18 Mbps
5	16-QAM	1/2	24 Mbps
6	16-QAM	3/4	36 Mbps
7	64-QAM	2/3	48 Mbps
8	64-QAM	3/4	54 Mbps

data signal. HR/DSSS which uses complementary code keying (CCK) as spreading phase sequence in IEEE 802.11b provides higher data rate of up to 11 Mbps. Higher data rate of up to 54 Mbps is provided in the IEEE 802.11a/g when OFDM modulation is used instead of spread spectrum techniques and a large number of orthogonal sub-carriers are used to carry the data signal. There are two supported data rates of 1 and 2 Mbps for IEEE 802.11 infrared communication using light waves as the transmission medium.

Chapter 3

Network and Voice Traffic Modeling

In this chapter, we present the models for the network and the voice traffic. The network includes wired and wireless domains. We consider a typical wired/wireless network in which voice connections exist between each wireless station in the wireless domain and a wired station in the wired domain. Therefore, the number of active voice flows is equal to the number of wireless stations. Voice traffic is modeled using a two-state `on-off` process.

3.1 Wired/Wireless network

A typical diagram of the network under consideration in this research is shown in Figure 3.1. All wireless nodes communicate only with one single access point that is connected to the wired network. The access point performs as a gateway that provides Internet access for the wireless end users in its coverage range. A wired server is used to direct all traffic from wired nodes to the access point and vice-versa. In the case of voice only traffic there is a two directional voice flow between each pair of wired/wireless nodes. All wireless nodes are in the coverage range of the access point and they operate according to the infrastructure mode.

A voice connection is created between one node in the wired domain and another node in the wireless domain. All traffic in the wireless domain is transmitted through the access point. All packets from any node in the wireless domain transmitted to any of the nodes of the wired domain pass through the access point and vice-versa. So the arrival rate into the access point in the downlink direction is n times the arrival rate into any wireless station in the uplink direction. In other words, all stations are using the same channel for uplink and downlink transmission and there is no priority for the access point. This results in a large buffer delay in the downlink which is

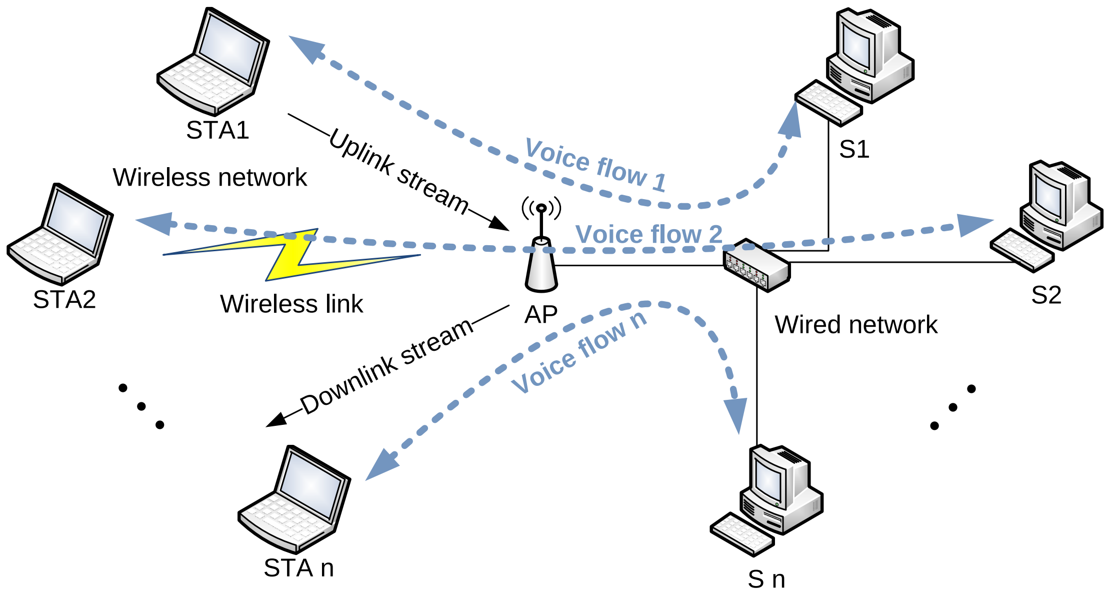

Figure 3.1 A typical wired/wireless Network.

the bottleneck of the system and limits the voice capacity. A simple queueing model for the wireless domain of the system is shown in Figure 3.2 where each node is represented by the MAC level queue.

3.2 Voice traffic

Our focus is mainly on how WLANs can support delay-sensitive voice application. Different QoS parameters for voice application includes delay, jitter and packet loss. Packet delay variations or jitter degrades the voice quality. Playout buffer is used to mitigate delay variations. A packet that experience a large delay is useless even if it is received by the destination. We define packet outage probability which is the probability that a packet arrives later than an appropriate delay bound. A 75 ms threshold is considered. We analytically obtain the collision probability, packet service time, packet delay, outage probability and throughput for uplink and downlink flows.

Voice traffic can be modeled using a two state `on-off` process as shown in Figure 3.3. The voice traffic source can be in the `on` (active) or `off` (silent) states. `on` and `off` periods are independently and exponentially distributed with parameters α

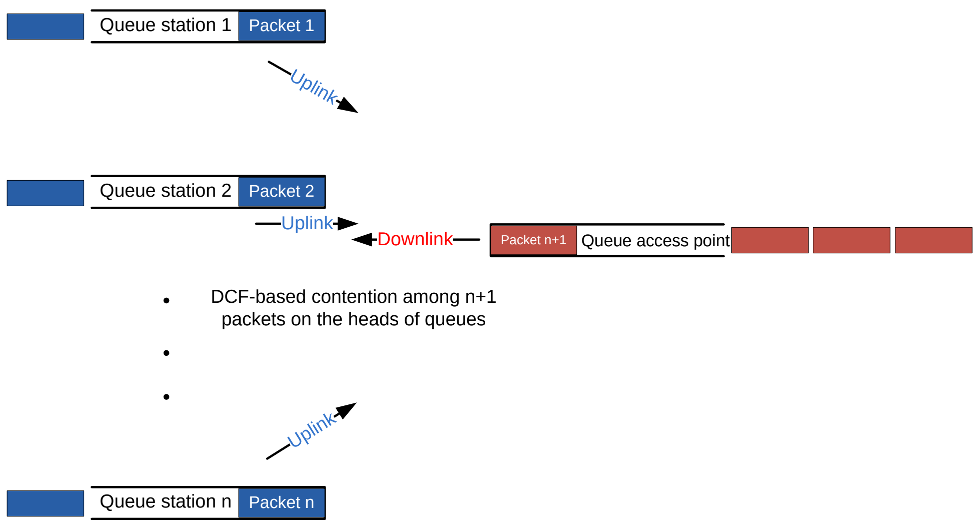

Figure 3.2 A queueing model of the wireless domain.

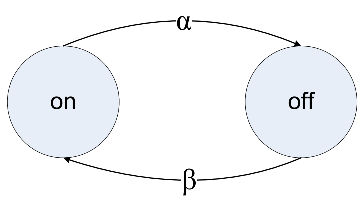

Figure 3.3 Voice traffic model.

and β. In this study, a voice source spends 400 ms in `on` state and 600 ms in `off` state.

We assume that a G.711 codec standard is used for converting the analogue waveform to a digital signal. Voice is sampled 8000 times per second. Since there are 8 bits per sample, the voice sampling rate equals 64 kb/s. Frames are generated for a sample period of 20 ms (50 packets per second) or 10 ms (100 packet per second). Denote $E_{10\,ms}[P]$ and $E_{20\,ms}[P]$ as the packet payload for 10 ms and 20 ms intervals. We can find the packet payload for 10 ms and 20 ms intervals as following:

$$E_{10\,ms}[P] = \frac{64000(bps)}{100(packet\ per\ second)} = 640(bits/packet) = 80(bytes/packet) \quad (3.1)$$

$$E_{20\,ms}[P] = \frac{64000(bps)}{50(packet\ per\ second)} = 1280(bits/packet) = 160(bytes/packet). \quad (3.2)$$

3.3 Interarrival time distribution of `on-off` voice traffic

In what follows we are proposing an approximation to calculate the probability that the buffer is in the idle state. Figure 3.4 shows the packet arrival process from a single source in the time domain. Denote X and Y as two independent random variables that represent the length of the `on` and `off` periods respectively. X and Y are exponentially distributed with parameters α and β and their probability density functions are given by:

$$f_X(x) = \alpha\ e^{-\alpha x} \quad , x \geq 0 \quad (3.3)$$

$$f_Y(y) = \beta\ e^{-\beta y} \quad , y \geq 0 \quad (3.4)$$

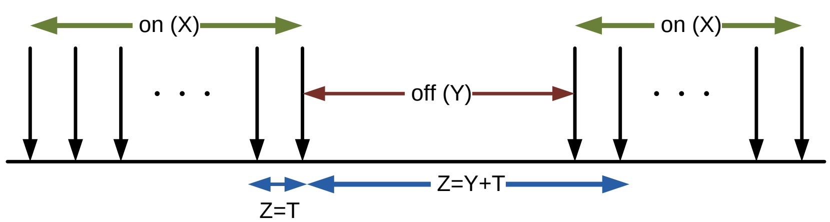

Figure 3.4 Packet arrival process from a single source.

During the `on` period, packets arrive at the fixed intervals of T second or at a constant rate of $\lambda = 1/T$ packet per second. The number of arrivals N during the `on` period can be approximated by a geometric random variable with the following distribution:

$$P(N = k) = p^k(1-p) \quad , k = 1, 2, ..., N \quad (3.5)$$

where p is the probability that the next interarrival time is T. We can find the average number of packet arrivals during `on` period as follows:

$$E[N] = \sum_{k=1}^{\infty} k \ P(N=k) = \frac{1}{1-p} \tag{3.6}$$

Since the average length of the `on` period is $1/\alpha$ and the arrival rate during the `on` period is λ, we have that $E[N] = \lambda/\alpha$ or $E[N] = 1/\alpha T$. From this, we have that $1 - p = \alpha T$.

Denote Z as the random variable that represents the length of the interarrival time of the packets. Since with probability p the next interarrival time is T and with probability $1-p$ the next interarrival time is exponentially distributed with parameter β, the probability density function of the packet interarrival time Z is given by [4]:

$$f_Z(z) = p\delta(z-T) + (1-p)\beta e^{-\beta(z-T)}u(z-T) \quad , z \geq 0 \tag{3.7}$$

where $u(t)$ is the unit step function and $\delta(t)$ is the unit impulse function. The average interarrival time is given by:

$$E[Z] = \int_0^{\infty} f(z)\, dz = \frac{(1-p) + \beta T}{\beta} = \frac{\alpha T + \beta T}{\beta} = \frac{\alpha + \beta}{\beta}\ T \tag{3.8}$$

Let q be the probability that at least one packet arrives during T_0 second. We can find q from the following equation.

$$q = \frac{T_0}{E[Z]} = \frac{\beta}{\alpha + \beta}\ \frac{T_0}{T} \tag{3.9}$$

If we use a Poisson process with parameter $\lambda_u = \lambda\, on\,/(on + off)$ or $\lambda_u = \lambda\, \beta/(\alpha + \beta)$ as the approximation of the arrival distribution, then we can find the probability q_u that at least one packet arrives during T_0 as,

$$q_u = 1 - e^{-\lambda_u\ T_0} = 1 - e^{-\frac{\beta}{\beta+\alpha}\ \lambda\ T_0} = 1 - e^{-\frac{\beta}{\beta+\alpha}\ \frac{T_0}{T}} \tag{3.10}$$

If $\beta\, T_0\,/[(\beta + \alpha)\, T] \to 0$ then Equation (3.10) is a good approximation for Equation (3.9).

The superposition of n traffic streams from n single and independent sources of voice, can be modeled by a Poisson process with parameter $\lambda_d = n\, \lambda\, on\,/(on + off)$

that can also be written as $\lambda_d = n\,\lambda\,\beta\,/(\alpha+\beta)$. As a result the probability q_d that at least one packet arrives during T_0 from the superposition of n voice streams is given by:

$$q_d = 1 - e^{-\lambda_d\ T_0} = 1 - e^{-\frac{n\beta}{\beta+\alpha}\ \frac{T_0}{T}} \tag{3.11}$$

We use Equations (3.10) and (3.11) in Section (5.1) to find the probability that a station exits from the idle state of the backoff model. As previously stated, the results of this section are approximations that can be used for the `on-off` model of the voice traffic. If there are other types of traffic such as data or video flows in the network, then we need to modify such approximations. This can be done according to the mathematical modeling of each type of traffic.

Chapter 4

MAC Layer Modeling and Performance Analysis

In this chapter, we give a brief description of the previous work related to this research. We classify previous work in the area of performance analysis and QoS support into two categories. The first one is concerned specifically with the performance analysis in terms of throughput and delay. Then, we explain the work that relates to the enhancement of network performance. We also review the research work addressing admission control techniques.

4.1 Performance analysis

There has been a lot of research in the field of performance analysis mostly trying to study throughput and delay characteristics. The first which related to the modeling of IEEE 802.11 considered constant or geometrically distributed backoff window [5] and [6]. Exponential backoff limited to two stages was considered in [7] and [8]. In [8], the performance of the DCF access scheme is evaluated through the measurement of the system throughput and access delay. Access delay is defined as the interval from the time when the packet gets at the head of the line in the transmission queue until the successful transmission of its first bit. The authors considered a DCF scheme based on the adaptive contention window mechanism. In this scheme, the optimal backoff window is dynamically selected according to the estimate of the number of contending stations.

It is worth mentioning that references [5] and [6] consider simplifications in the backoff procedure. Although reference [7] considers exponential backoff process, it is limited to two stages. In [9], Bianchi proposed an accurate model for MAC performance analysis. Bianchi's model is based on a two-dimensional discrete time Markov chain. Because of its simplicity and detailed analysis of backoff process, Bianchi model has received notice from other researchers. The assumptions in Bianchi's work are that the channel conditions are ideal, the network includes a finite number of terminals and the network is saturated. In other words, all wireless stations always have a packet to transmit. Although Bianchi's model is simple and accurate, it does not take into account errors due to the wireless channel. Throughput analysis in a lossy channel with rate adaptation in the physical layer has been found by modifying Binachi's model and considering the failure due to the channel noise in the calculation of the probability of losing a packet in [10]. In the IEEE 802.11, the number of retransmissions can be more than the maximum number of backoff stage. In Bianchi's model the number of retransmissions is unlimited. In [11], a limited number of retransmissions is considered. A Markovian state dependent single server queue to analyze the throughput and the mean packet delay is proposed in [12]. In [13], the 802.11 throughput analysis using a p-persistent backoff mechanism to approximate the original backoff in the protocol was studied.

An exact analysis of the throughput behavior is presented in [14]. After some mathematical analysis, it is shown that the throughput is a linear function of the arrival rate, number of stations and average payload.

The authors in [15] obtained the packet delay considering time spent for deferring transmission, unsuccessful transmission due to collision and final packet's successful transmission. But the average time slot is calculated for n competing stations. However, when one station defers packet transmission, the average time slot must be calculated according to the remaining $n - 1$ stations. A successful transmission may occur in different stages. Using the probability of successful transmission at each stage as defined in [16] and considering $n - 1$ competing stations, the authors obtained the average time needed for a successful transmission.

In [17], the authors obtained the probability distribution of the MAC layer service time. MAC layer service time is the time interval from the moment the packet arrives at the head of the queue to the instant the packet is successfully acknowledged or dropped. The authors obtained the signal flow graph in the z domain based on the Bianchi's model. Finally, from the inverse of the Probability Generating Function (PGF), the Probability Distribution Function (PDF) of the service time is obtained. Also in [18] the QoS performance of the IEEE 802.11 is studied in terms of throughput, delay and delay variation, and packet loss rate.

In [19], a closed form for the PGF of the service time is obtained for finite load

IEEE 802.11 network. Service time distribution and queue characteristics are obtained using a G/G/1 queueing model at each node. Using a G/G/1 model for each node, arbitrary arrival rate and packet length distribution are considered and the effect of the number of nodes and the network load on queue length and delay are studied. The generating function method has also been used to find the access delay model for the EDCA mechanism in [20]. The shortcoming of the PGF-based analytical method in [17]-[20] is that the conversion of the PGF to the PDF is not straightforward and it requires numerical methods which makes them improper for resource allocation decisions.

In [21], the service time distribution in the IEEE 802.11 with RTS/CTS is investigated. It is shown that the service time is approximately geometrically distributed. The analysis of the number of packets in the queueing system and the queue length is based on the M/Geo/1 queueing system. The simple approximation of the service time introduced in this work would be applicable in developing statistical admission control and prediction of the buffer utilization.

4.1.1 Packet service time analysis

For the delay-sensitive applications such as VoIP, the evaluation of the delay performance is vital to design efficient algorithms to guarantee the QoS requirements of such applications. In Chapter 5, we compute the packet service time and average delay for voice application over DCF-based WLAN. In this section, we briefly discuss the principles of computing the packet service time according to the literature review. Several models have been proposed to analyze the packet service time in the IEEE 802.11. In [15] the packet service time is obtained considering times spent for deferring transmission, unsuccessful transmission due to collision and final packet's successful transmission. Suppose that there are m stages according to the DCF access mechanism. Each stage represents the size of the contention window which depends on the number of transmissions failed for the packet. After each unsuccessful transmission, the size of the contention window is doubled up to the maximum contention window. Denote $E[T]$ as the average time for a successful transmission. A successful transmission may occur in different stages. As a result, the following formula can be used to compute the average time needed for a successful transmission [16],

$$E[T] = \sum_{i=0}^{m} E[T_i]\, p_i \tag{4.1}$$

where $E[T_i]$ is the average time for a successful transmission at the i^{th} stage and p_i is the probability that the successful transmission takes place in the i^{th} stage. To determine $E[T_i]$, we need to find successful transmission time, collision time and backoff time. We can write $E[T_i]$ as,

$$E[T_i] = T_b + T_s + iT_c \tag{4.2}$$

where T_b is the the backoff time, T_s is the average time that the channel is sensed busy because of a successful transmission and T_c is the time that the channel is busy because of a collision. For the basic mode, we can write the successful transmission time $T_{s,basic}$ and collision time $T_{c,basic}$ as,

$$T_{s,basic} = DIFS + \frac{H_{PHY}}{R_b} + \frac{H_{MAC}}{R_c} + \frac{l}{R_c} + \delta + SIFS + \frac{ACK}{R_b} + \delta \tag{4.3}$$

$$T_{c,basic} = DIFS + \frac{H_{PHY}}{R_b} + \frac{H_{MAC}}{R_c} + \frac{l}{R_c} + \delta \tag{4.4}$$

where H_{PHY} is the length of the physical header, R_b is the basic bit rate, H_{MAC} is the length of the MAC header, R_c is the channel bit rate, l is the packet payload and δ is the propagation delay. For the RTS/CTS we can find the successful transmission time $T_{s,RTS/CTS}$ and collision time $T_{c,RTS/CTS}$ as follows.

$$\begin{aligned} T_{s,RTS/CTS} = {} & DIFS + \frac{RTS}{R_b} + SIFS + \delta + \frac{CTS}{R_b} + SIFS + \delta \\ & + \frac{H_{PHY}}{R_b} + \frac{H_{MAC}}{R_c} + \frac{l}{R_c} + SIFS + \delta + \frac{ACK}{R_b} + \delta \end{aligned} \tag{4.5}$$

$$T_{c,RTS/CTS} = DIFS + \frac{RTS}{R_b} + SIFS + \frac{CTS}{R_b} \tag{4.6}$$

The probability of successful transmission at the i^{th} stage is defined in [16] as follows.

$$p_i = \frac{p^i(1-p)}{1-p^{m+1}} \tag{4.7}$$

where $1 - p^{m+1}$ is the probability that the packet is not dropped and $p^i(1-p)$ is the probability that the packet is successfully transmitted with probability $(1-p)$ after it reaches the i^{th} stage with probability p^i. Denote n as the number of contending stations, W as the minimum contention window size, and m as the maximum number of backoff stage. We can find the probability that a station transmits a packet in a randomly chosen slot time τ and p the collision probability, from the following nonlinear equations [9]:

$$p = 1 - (1-\tau)^{n-1} \tag{4.8}$$

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)} \tag{4.9}$$

In the IEEE 802.11, the number of retransmissions can be more than the maximum number of backoff stages. In Bianchi's model [9], the maximum number of backoff stages is defined as the maximum number of retransmissions. In [11], more retransmissions are considered. Then the contention window size W_i for the i^{th} stage of the backoff procedure can be found as following.

$$W_i = \begin{cases} 2^i W & i \leq m \\ 2^m W & m \leq i \leq m' \end{cases} \tag{4.10}$$

where m' is the maximum number of retransmissions.

We can also find the probability that a station transmits a packet in a randomly chosen slot time τ from the following equation.

$$\tau = \begin{cases} \frac{2(1-2p)(1-p^{m+1})}{W(1-(2p)^{m+1})(1-p)+(1-2p)(1-p^{m+1})} & m \leq m' \\ \frac{2(1-2p)(1-p^{m+1})}{W(1-(2p)^{m'+1})(1-p)+(1-2p)[(1-p^{m+1})+W2^{m'}p^{m'+1}(1-p^{m-m'})]} & m \geq m' \end{cases} \tag{4.11}$$

After solving two nonlinear Equations (4.8) and (4.11), we can find the transmission probability and the collision probability. Then the probability that at least one station transmits P_{tr} and also the probability of a successful transmission given that at least one station transmits P_s can be found:

$$P_{tr} = 1 - (1-\tau)^n \tag{4.12}$$

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \tag{4.13}$$

We notice that provided that the packet has been transmitted at the i^{th} stage, the packet service time is obtained from the following equation:

$$E[T_i] = E[slot]\sum_{k=0}^{j}[\frac{W_k - 1}{2}] + iT_c + T_s \qquad i \in [0, m] \tag{4.14}$$

where the average length of a transmission slot can be written as:

$$E[slot] = (1 - P_{tr})\sigma + P_{tr}P_sT_s + P_{tr}(1 - P_s)T_c \tag{4.15}$$

where σ is the slot time, T_s is the average time the channel is sensed busy because of a successful transmission and T_c is the average time the channel is sensed busy during a collision.

Finally from Equation (4.1) we can write the packet service time as follows:

$$E[T] = \sum_{j=0}^{m} E[T_i]\left(\frac{p^i(1-p)}{1-p^{m+1}}\right) \tag{4.16}$$

Moreover, to find a more accurate result, we notice that while one station defers its transmission, there are $n-1$ stations competing for the channel [16]. As a result, there should be a slight change in the definition of $E[slot]$. The probability that at least one station transmits $P_{tr}^{'}$ and also the probability of a successful transmission given that at least one station transmits $P_s^{'}$ can be found:

$$P_{tr}^{'} = 1 - (1-\tau)^{n-1} \tag{4.17}$$

$$P_s^{'} = \frac{(n-1)\tau(1-\tau)^{n-2}}{P_{tr}^{'}} = \frac{(n-1)\tau(1-\tau)^{n-2}}{1-(1-\tau)^{n-1}} \tag{4.18}$$

Then, we find the slot length in which $n-1$ stations are competing:

$$E^{'}[slot] = (1 - P_{tr}^{'})\sigma + P_{tr}^{'}P_s^{'}T_s^{'} + P_{tr}^{'}(1 - P_s^{'})T_c \tag{4.19}$$

The average packet service time can be found:

$$E[T] = \sum_{j=0}^{m} \left(E^{'}[slot] \sum_{k=0}^{j} [\frac{W_k - 1}{2}] + iT_c + T_s \right) \left(\frac{p^i(1-p)}{1-p^{m+1}} \right) \tag{4.20}$$

4.1.2 Saturation throughput analysis

In addition to the evaluation of the packet service time and delay characteristics, we will also analyze the throughput performance of DCF in Chapter 5. In this section we briefly explain the most important model which has been proposed in [9] to compute the saturation throughput of DCF. Throughput S is defined as the amount of bits per second successfully transmitted. Using Equations (4.12), (4.13) and (4.15), throughput S is given by [9]:

$$S = \frac{P_s \; P_{tr} \; E[P]}{(1-P_{tr})\;\sigma + P_{tr}\; P_s\; T_s + P_{tr}\; (1-P_s)\; T_c} \tag{4.21}$$

where $E[P]$ represents the payload information transmitted in a slot time.

We note that there are are three assumptions in Bianchi's model:

1. The network is in the saturated mode which means that each station has always a packet to transmit.
2. The collision probability of a packet transmitted by each station is constant and independent regardless of the traffic arrival rate.
3. The channel is ideal.

Those limitations will be dealt with in Chapter 5.

4.2 Performance enhancement

Another research area considers enhancement of the quality of service in the IEEE 802.11 network. In [22], the average size of the contention window that leads to the near-optimum performance of the network in terms of maximum achievable throughput is obtained. It is shown that the network performance is significantly improved when the exponential backoff is replaced by an adaptive backoff mechanism.

Scaling the contention window and using different interframe spacing to support service differentiation is proposed in [23]. Although changing the size of the contention window according to the different users' priority improves the performance of User Datagram Protocol (UDP) flows, it is not efficient in the case of Transmission Control Protocol (TCP) flows. The second mechanism which adapts the interframe spacing helps improving the performance of UDP and TCP flows. Another approach based on different maximum frame size for different priorities results in the improvement of both TCP and UDP traffic.

Virtual medium access control (VMAC) was suggested in [24] to provide differentiated services for data, voice and traffic. In this mechanism, the channel is monitored and VMAC operates in parallel with MAC, but it does not really send packets. Also, a virtual source (VS) algorithm is proposed. In the VS algorithm, packets are generated by emulated applications and the delay curve is obtained based on the virtual application, interface queue and VMAC. Using delay curves, an application can adjust its traffic parameters in order to ensure a globally stable state for the network.

In [25], high priority is considered for real-time traffic. However, it imposes special requirements on high-priority traffic and is not fully compatible with the existing 802.11 standard. There are two shortcomings in [22]-[25]. First, they do not guarantee specified QoS parameters, but they provide service differentiation. Second, they impose different changes on the original access function as defined in the standard.

Different admission control techniques and the analysis of the maximum number of connections in the IEEE 802.11 network have been studied in several recent work. In [26], the behavior of Voice over Internet Protocol (VoIP) in 802.11 networks in terms of the maximum number of connections that a single AP can support is investigated analytically and experimentally. It is shown that 802.11 cannot support a large number of IP calls because of the inefficiency of the wireless channel. To improve, it is suggested to use larger payload sizes. However, this will result in degradation of delay, jitter and packet loss.

In [27], voice capacity of a WLAN in infrastructure mode with DCF is evaluated. The authors considered 802.11b, 802.11a and 802.11g and developed different simulation scenarios to study the effect of channel bandwidth, various codec packetization and different packet size on the capacity of the network. It is shown that under the constraints of delay budget and packet loss, 802.11a and 802.11g support more connections in comparison with 802.11b.

A threshold-based admission control is proposed in [28]. The admission control is implemented based on the traffic measurement in each station. In other words, each station monitors the wireless link. By calculating the relative occupied bandwidth or average collision and comparing them with the proper threshold values a decision is made whether a new connection is admitted or refused. The difficulty of finding optimal values of threshold is a challenge in this method.

The HARMONICA scheme is proposed in [29] to avoid congestion using a flexible admission control mechanism based on link layer quality indicator (LQI) which includes parameters such as drop rate, delay and throughput for each traffic class. AP sends the information about channel access parameters (CAP) to all other nodes to ensure that the QoS of different applications is supported while channel utilization is maximized. Based on the requirements of the flow and LQI parameters, a decision is made regarding admitting the new flow. But finding the optimal parameters of the channel access is not addressed and it may take a long time for the network to work in the optimal status.

Reference [30] proposed a Markov chain model-based admission control based on the predicted achievable throughput. The method assumes that the network is in the saturated mode, which is not a realistic assumption. In [31], a Connection Admission Control (CAC) algorithm similar to [30] and based on the non-saturation analysis was proposed. Developing admission control algorithms based on the decisions of the end users rather than the resources, as presented in [30] and [31], would be a challenge. A tutorial of various admission control algorithms is discussed in [32].

A call admission and rate control (CARC) is proposed in [33] based on the idea of regulating the arriving traffic of the network in order to make sure that the network is working at the optimal point. In this method the busyness channel ratio is defined as the portion of the time during which the channel is not idle. The algorithm uses busyness channel ratio and the specification of real-time flow and a coordinator node to reject or accept the new stream.

A connection admission control, based on the theory of effective bandwidth was proposed in [34]. It is shown that the admission region is approximately linear and therefore a simple and effective admission control algorithm can be obtained by using the concept of effective bandwidth. In [35], the impact of wireless channel parameters on the QoS parameters of the VoIP traffic was studied. A Finite State Markov Chain (FSMC) is used to model the time-varying fading channel and the modulation and coding in the physical layer is performed based on the channel condition. It is found that even in the case of a non-ideal channel, the admission region is bounded with a linear function. Simulation results reveal that the increase of SNR or Doppler frequency results in the increase of the maximum number of VoIP in the network. [34] and [35] consider two types of traffic voice and data in the experiments as well as DCF channel access in the infrastructure mode of IEEE 802.11 networks and do not include a theoretical analysis of the measurement-based results.

From the literature review, we observe that the previous work do not consider the difference between the amount of traffic entering into the queue of the MAC layer of a wireless station and the access point. In the next chapter, we propose a new model which is capable of differentiating between the uplink and downlink traffic. We evaluate the DCF performance through the calculation of several important parameters such as the probability of collision, packet service time, average delay and throughput. In the category of performance enhancement, We propose a novel model-based admission control scheme to guarantee the QoS requirements of the voice application. The proposed admission control scheme deploys the packet delay outage probability which is found by expressing the waiting time distribution of the voice packet at the MAC layer buffer of the access point. The packet service time that is found according to the proposed performance analysis is used to estimate the waiting time distribution to propose the admission control scheme in the category of performance enhancement.

Chapter 5

Models and Performance Measures for Voice Traffic

In this chapter we propose a novel analytical framework for MAC layer analysis by differentiating between the backoff model for the uplink and the downlink direction. First, we represent the backoff model in the non-saturated mode based on a two-dimensional Markov chain in Section (5.1). We explain the simulation set-up and the confidence interval in Section (5.2). After obtaining a set of nonlinear equations in Section (5.1), we use numerical methods to solve them to investigate important parameters such as collision probability, packet service time and throughput. In Section (5.3), we compare the values of collision probability obtained from the proposed analysis with those obtained from ns-2 simulations. We represent the packet service time analysis and simulation results in Section (5.4). After analysis of the M/M/1/K queueing model in Section (5.5), we will discuss the average packet delay and the simulation results in Section (5.6). Finally, we obtain the packet delay outage probability and propose a model-based admission control scheme in Section (5.8). The analysis provided in this chapter considers the existence of only voice application in the network. The system model and voice traffic model are the ones we discussed in Chapter 3.

5.1 MAC layer analysis for voice traffic

Assuming that the average duration of talk and silent periods for a voice connection are `on` and `off` respectively, we can express the arrival average rate into the MAC layer buffer of the access point, λ_d, and of the wireless station, λ_u, as follows:

$$\lambda_u = \frac{on}{on + off}\ \lambda \tag{5.1}$$

$$\lambda_d = n \, \frac{on}{on + off} \, \lambda \tag{5.2}$$

where n is the number of wireless stations and λ is the packet arrival rate during a talk period. λ depends on the sample period for which the voice frames are generated.

Suppose that each wireless station transmits a packet in the uplink direction with probability τ_u and the access point transmits a packet in the downlink stream with probability τ_d. Then the probability that a packet transmitted in the uplink sees a collision can be found as

$$p_u = 1 - (1 - \tau_u)^{n-1}(1 - \tau_d) \tag{5.3}$$

A packet sent from the access point sees a collision if at least one of the n wireless stations transmits. As a result, the downlink collision probability is given by:

$$p_d = 1 - (1 - \tau_u)^n \tag{5.4}$$

The backoff procedure in the non-saturated mode can be modeled using a two-dimensional Markov model as shown in Figure 5.1. Each state $\{s(t), b(t)\}$ of the backoff model, includes two elements. $s(t)$ is the stochastic process that represents the backoff stage $[0, 1, ..., m]$ while $b(t)$ is the stochastic process that represents the backoff time counter $[0, 1, ..., W_i - 1]$. W_i is the size of contention window for the i^{th} stage of the backoff procedure. W_0 is the size of minimum contention window at the first transmission attempt. W_m is the size of maximum contention window for the maximum backoff stage m. For each stage of the backoff i^{th}, we can find the size of contention window as $W_i = 2^i W_0$. The average slot time that includes the average time of a successful transmission, collision or idle slots is used as the time unit of the backoff model. Since after each slot time, the backoff counter is decremented with probability 1, we can write:

$$P\{i, k | i, k+1\} = 1 \qquad k \in [0, W_i - 2] \qquad i \in [0, m] \tag{5.5}$$

The idle state in the backoff model, represents the empty buffer in the non-saturated mode. Denote q as the probability that there is at least one packet in the buffer and p as the probability of collision. After a successful transmission, the

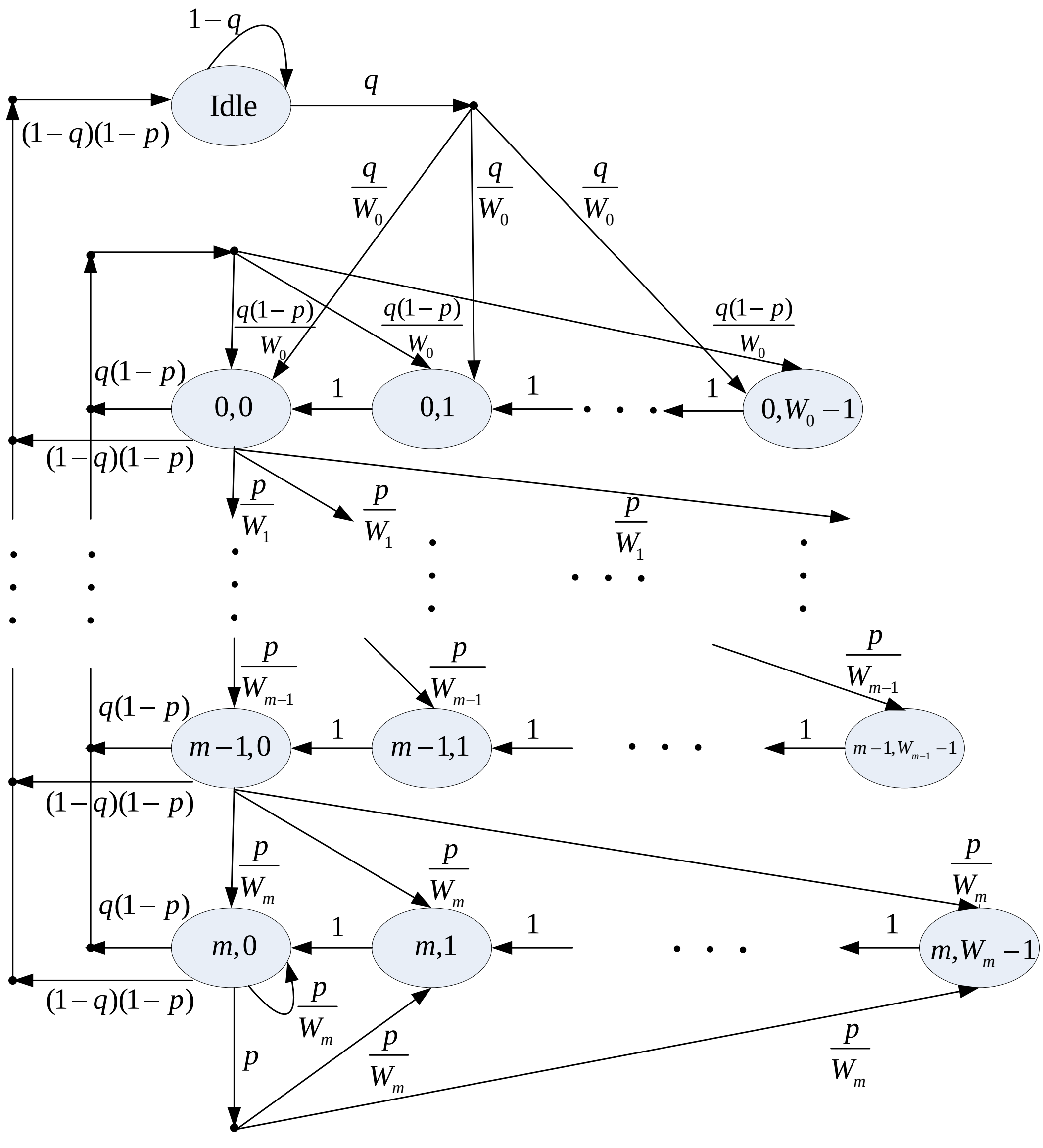

Figure 5.1 Backoff modeling in non-saturated mode.

station starts a new transmission attempt at the 0^{th} backoff stage, provided that the buffer is not empty. As a result, we can write:

$$P\{0, k|i, 0\} = \frac{q\,(1-p)}{W_0} \qquad k \in [0, W_0 - 1] \qquad i \in [0, m] \tag{5.6}$$

If there is a collision at the $(i-1)^{th}$ stage, then the station randomly chooses a new contention window of the next backoff stage which is distributed uniformly in the range of $[0, W_i]$. Therefore, the following equation holds.

$$P\{i, k|i-1, 0\} = \frac{p}{W_i} \qquad k \in [0, W_i - 1] \qquad i \in [1, m] \tag{5.7}$$

We assume that when a collision happens after the maximum backoff stage m, then the station starts to retransmit the packet by choosing the maximum contention window W_m. We can write:

$$P\{m, k|m, 0\} = \frac{p}{W_m} \qquad k \in [0, W_m - 1] \tag{5.8}$$

There are two situations which make the station being in the idle state. First, the station has no packet to transmit and at the beginning of the next slot time, the buffer is still empty. This can be written as the following equation.

$$P\{Idle|Idle\} = 1 - q \tag{5.9}$$

Second, after a successful packet transmission, there is no packet to transmit at the buffer. We can write this event as follows:

$$P\{Idle|i, 0\} = (1-q)(1-p) \qquad i \in [0, m] \tag{5.10}$$

If the station is in the idle state and a packet arrives into the buffer, then the station starts to transmit the new packet by choosing randomly the contention window at the 0^{th} backoff stage. So, we can write:

$$P\{0, k|Idle\} = \frac{q}{W_0} \qquad k \in [0, W_0 - 1] \tag{5.11}$$

Denote $b_{i,k} = \lim_{t\to\infty} P\{s(t) = i, b(t) = k\}, i \in [0, m], k \in [0, W_i - 1]$ as the stationary distribution of the Markov chain. After solving in the stationary state, we can find a closed-form solution for this Markov chain. First, we can write the following relations.

$$b_{i-1,0}\, p = b_{i,0} \longrightarrow b_{i,0} = p^i b_{0,0}, \;\; \forall\; i \in [1, m-1] \tag{5.12}$$

$$b_{m-1,0}\, p = (1-p) b_{m,0} \longrightarrow b_{m,0} = \frac{p^m}{1-p} b_{0,0}, \;\; i = m \tag{5.13}$$

The probability b_I of being in the idle state in the stationary state is given by:

$$b_I = (1-q) b_I + (1-q)(1-p) \sum_{i=0}^{m} b_{i,0} \longrightarrow b_I = \frac{(1-q)(1-p)}{q} \sum_{i=0}^{m} b_{i,0} \tag{5.14}$$

We can find the stationary probabilities for each $k \in [1, W_i - 1]$ as follows:

$$b_{i,k} = \frac{W_i - k}{W_i} \begin{cases} q(1-p) \sum_{i=0}^{m} b_{i,0} + q b_I, & i = 0 \\ p b_{i-1,0}, & i \in [1, m-1] \\ p(b_{m-1,0} + b_{m,0}), & i = m \end{cases} \tag{5.15}$$

After substituting Equation (5.14) in (5.15), by using the normalization condition $\sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} = 1$ and remembering the fact that $\sum_{i=0}^{m} b_{i,0} = b_{0,0}/1-p$, we can find the probability $b_{0,0}$ as follows.

$$b_{0,0} = \frac{2(1-p)(1-2p)q}{q[(1-2p)(W+1) + pW(1-(2p)^m)] + 2(1-q)(1-p)(1-2p)} \tag{5.16}$$

where W is the minimum size of the contention window and m is the maximum number of backoff stages.

We can now express the probability τ that the station transmits in a randomly chosen slot time as follows:

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{2(1-2p)q}{q[(1-2p)(W+1) + pW(1-(2p)^m)] + 2(1-q)(1-p)(1-2p)} \tag{5.17}$$

We consider a similar backoff procedure for both access point and wireless station since there is no priority for the access point over the wireless stations. The model represents the backoff procedure per station. Since we assume that DCF parameters such as the size of the contention window, the number of backoff stages, etc are not variable, the only difference among backoff model per station for the access point and the wireless station is the amount of packet arrival which is related to the probability that a station has a packet waiting to be transmitted or stays in the idle state. Let q_u be the probability that there is at least one packet in the buffer of the wireless station, p_u be the probability of collision seen by the packets transmitted by the wireless station in the uplink direction, q_d be the probability that there is at least one packet in the buffer of the access point and p_d be the probability of collision seen by the packets transmitted by the access point in the downlink direction. After solving the backoff model for the wireless station and the access point, we can find the probability that the wireless station transmits in a randomly chosen slot time τ_u and the probability that the access point transmits in a randomly chosen slot time τ_d as following:

$$\tau_u = \frac{2(1-2p_u)q_u}{q_u[(1-2p_u)(W+1)+p_uW(1-(2p_u)^m)]+2(1-q_u)(1-p_u)(1-2p_u)} \tag{5.18}$$

$$\tau_d = \frac{2(1-2p_d)q_d}{q_d[(1-2p_d)(W+1)+p_dW(1-(2p_d)^m)]+2(1-q_d)(1-p_d)(1-2p_d)} \tag{5.19}$$

In Section (3.3), we have shown that the average interarrival time $E[Z]$ for a single `on-off` source is given by:

$$E[Z] = \frac{\alpha+\beta}{\beta}\,T \tag{5.20}$$

where α is the parameter of the exponential random variable that represents the length of the `on` period, β is the parameter of the exponential random variable that represents the length of the `off` period and T is the constant interarrival time during `on` period. Let q be the probability that at least one packet arrives from a single `on-off` source at the MAC layer buffer of the wireless station during the slot time $E[slot]$ second. We can find q from the following equation.

$$q = \frac{E[slot]}{E[Z]} = \frac{\beta}{\alpha+\beta}\,\frac{E[slot]}{T} \tag{5.21}$$

If we use a Poisson process with parameter $\lambda_u = \lambda\,on\,/(on+off)$ or $\lambda_u = \beta\,\lambda\,/(\alpha+\beta)$ as the approximation of the arrival distribution for the wireless station, then we can find the probability q_u that at least one packet arrives during $E[slot]$ as,

$$q_u = 1 - e^{-\lambda_u \; E[slot]} = 1 - e^{-\frac{\beta}{\beta+\alpha} \; \frac{E[slot]}{T}} \tag{5.22}$$

We note that since $E[slot]$ is in the order of microsecond and T is in the order of millisecond, thus we have,

$$\frac{\beta}{\beta+\alpha} \; \frac{E[slot]}{T} \rightarrow 0 \tag{5.23}$$

Therefore, Equation (5.22) is a good approximation for Equation (5.21).

$E[slot]$ is the average slot time and can be found from the following equation,

$$E[slot] = (1 - P_{tr})\sigma + P_{tr}P_sT_s + P_{tr}(1 - P_s)T_c \tag{5.24}$$

where P_{tr} is the probability that at least one station transmits and P_s is the probability that a successful transmission occurred. σ is the slot time, T_s is the average time the channel is sensed busy because of a successful transmission and T_c is the average time the channel is sensed busy during a collision. As shown in Figure 5.2, for the basic access mechanism we have the following expressions for T_s and T_c :

$$T_s = DIFS + \frac{H_{PHY}}{R_b} + \frac{H_{MAC} + H_{IP}}{R_c} + \frac{E[P]}{R_c} + \delta + SIFS + \frac{ACK}{R_b} + \delta \tag{5.25}$$

$$T_c = DIFS + \frac{H_{PHY}}{R_b} + \frac{H_{MAC} + H_{IP}}{R_c} + \frac{E[P]}{R_c} + \delta + ACK_{timeout} \tag{5.26}$$

where H_{PHY} is the length of the physical header, R_b is the basic bit rate, H_{IP} is the length of the IP header, H_{MAC} is the length of the MAC header, R_c is the channel bit rate, $E[P]$ is the packet payload and δ is the propagation delay. Since we choose $ACK_{timeout} = SIFS + ACK/R_b + \delta$, we can write $T_c = T_s$.

P_{tr} and P_s can be found from the following equations,

$$P_{tr} = 1 - (1 - \tau_u)^n(1 - \tau_d) \tag{5.27}$$

$$P_s = P_{u,s} + P_{d,s} \tag{5.28}$$

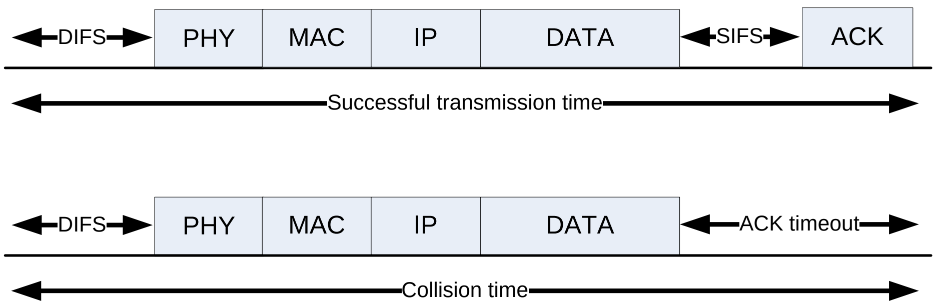

Figure 5.2 T_s and T_c for basic access mechanism.

where $P_{u,s}$ is the probability that a wireless station successfully transmits a packet in the uplink direction and $P_{d,s}$ is the probability that the access point successfully transmits a packet in the downlink direction. $P_{u,s}$ and $P_{d,s}$ can be found from the following equations.

$$P_{u,s} = \frac{n\tau_u(1-\tau_u)^{n-1}(1-\tau_d)}{P_{tr}} = \frac{n\tau_u(1-\tau_u)^{n-1}(1-\tau_d)}{1-(1-\tau_u)^n(1-\tau_d)} \tag{5.29}$$

$$P_{d,s} = \frac{\tau_d(1-\tau_u)^n}{P_{tr}} = \frac{\tau_d(1-\tau_u)^n}{1-(1-\tau_u)^n(1-\tau_d)} \tag{5.30}$$

The probability q_d that at least one packet from the superposition of n voice streams arrives during the slot time $E[slot]$ at the MAC layer buffer of the access point can be found as:

$$q_d = 1 - e^{-\lambda_d \ E[slot]} = 1 - e^{-\frac{n\beta}{\beta+\alpha} \ \frac{E[slot]}{T}} \tag{5.31}$$

By solving the nonlinear Equations (5.3), (5.4), (5.18), (5.19), (5.22) and (5.31) simultaneously, we can find the unknown variables p_u, p_d, τ_u, τ_d, q_u and q_d in order to assess important network parameters such as the packet service time, packet queueing delay and throughput.

5.2 Simulation using ns-2

For the sake of validating the proposed analytical framework, we perform simulations using ns-2 (version 2.33) [36]. The wired links are ideal with a delay of 0 ms and capacity of 100 Mbps. One access point and n wireless stations are placed over an area of $670 \times 670\, m^2$. We set up one voice connection between each wireless station in the wireless domain and every wired node in the wired domain. All voice flows carry UDP traffic with a rate of 64 kbps. The arrival distribution is exponential `on-off` traffic which is a realistic traffic model for voice flow. Each node in the `on` state generates packets every 20 ms (160 B) or every 10 ms (80 B). The average *on* and off periods are 400 ms and 600 ms respectively. We used AWK programming to analyze the trace files obtained from running ns-2 simulations using the OTcl file. We also wrote a SHELL script to evaluate the performance metrics for different scenarios. The evaluation of each parameter such as the probability of collision, the packet service time, the packet delay, the throughput and the packet delay outage probability is accomplished by averaging over M=30 sample scenarios while each scenario run time is 120 second. For each variable X, we have computed the mean $\bar{X}$, variance σ_X^2 and the 95 percent confidence interval $[X_l, X_u]$ from the following equations.

$$\bar{X} = \frac{1}{M}\sum_{i=1}^{M} X_i \tag{5.32}$$

$$\sigma_X^2 = \frac{1}{\sqrt{M-1}}\sum_{i=1}^{M} (X_i - \bar{X})^2 \tag{5.33}$$

$$X_l = \bar{X} - \frac{1.96 \times \sigma_X}{\sqrt{M}} \tag{5.34}$$

$$X_u = \bar{X} + \frac{1.96 \times \sigma_X}{\sqrt{M}} \tag{5.35}$$

Since the variance of the key parameters such as the probability of collision and the packet delay outage probability (particularly in the region of interest where the packet delay outage probability is less than 1 percent) is negligible, we only show the mean of the variables in the figures.

5.3 Probability of collision

When two packets are transmitted simultaneously a collision happens and both packets are lost. As we saw in the previous section, collision probability seen in the uplink and downlink direction p_u and p_d can be found analytically from Equations (5.3) and (5.4). In this section, we compare the simulation results with the values obtained from the analytical formulation. We obtained these analytical values by using the `fsolve` function in Matlab to find the unique root (zero) of the system of nonlinear equations. Therefore, we write the set of nonlinear equations in the form of $F(x) = 0$, where $x = [x(1), x(2), x(3), x(4), x(5), x(6)]$ is a vector that represents the unknown variables p_d, τ_d, q_d, p_u, τ_u and q_u. After substituting x in Equations (5.3), (5.4), (5.18), (5.19), (5.22), (5.24) and (5.31), we can find the set of six equations of $F(x)$ as follows:

$$x(1) + (1 - x(5))^n - 1 = 0 \tag{5.36}$$

$$\frac{2(1-2x(1))x(3)}{x(3)[(1-2x(1))(W+1) + x(1)W(1-(2x(1))^m)] + 2(1-x(3))(1-x(1))(1-2x(1))} - x(2) = 0 \tag{5.37}$$

$$e^{-\frac{n\beta}{\beta+\alpha}\frac{E[slot]}{T}} + x(3) - 1 = 0 \tag{5.38}$$

$$x(4) + (1 - x(5))^{n-1}(1 - x(2)) - 1 = 0 \tag{5.39}$$

$$\frac{2(1-2x(4))x(6)}{x(6)[(1-2x(4))(W+1) + x(1)W(1-(2x(4))^m)] + 2(1-x(6))(1-x(4))(1-2x(4))} - x(5) = 0 \tag{5.40}$$

$$e^{-\frac{\beta}{\beta+\alpha}\frac{E[slot]}{T}} + x(6) - 1 = 0 \tag{5.41}$$

The simulation parameters are summarized in Table 5.1. Figure 5.3 shows the probability of collision when the packet payload is 80 bytes and the number of wireless stations varies from 1 to 45. Although the collision probability seen from uplink and downlink are very close, we observe that a packet in the uplink sees a larger probability of collision in comparison with a packet in the downlink. This happens because of higher transmission probability from the access point in the downlink

Table 5.1 Simulation parameters.

Parameter	802.11b	802.11g
Packet payload	80 B or 160 B	80 B or 160 B
on period	400 ms	400 ms
off period	600 ms	600 ms
Channel bit rate	11 Mbps	54 Mbps
Basic bit rate	1 Mbps	6 Mbps
IP header	20 B	20 B
MAC header	34 B	34 B
PHY header	24 B	15 B
ACK	38 B	29 B
Slot time	20 μs	9 μs
SIFS	10 μs	10 μs
DIFS	50 μs	28 μs
CW_{min}	32	16
CW_{max}	1024	1024
Propagation delay	1 μs	1 μs
Maximum number of retransmission	6	7

direction. Indeed the voice traffic arrival rate is n times larger than the voice traffic arrival rate into a wireless station. As a result, packets transmitted from the access point reduce the chance of a successful transmission in the uplink direction.

Figure 5.4 shows the probability of collision when the packet payload is 160 bytes. It is seen that for given number of wireless stations, the collision probability decreases as we increase the packet payload. This is due to the fact that when we use 160 bytes voice packets instead of 80 bytes, we are decreasing the packet arrival rate from the voice traffic source during a slot time.

5.4 Packet service time

We define the service time of a packet as the time interval from the instant that the packet is ready to be transmitted in the head of the queue till the instant that the packet is either successfully transmitted or dropped. The MAC layer service

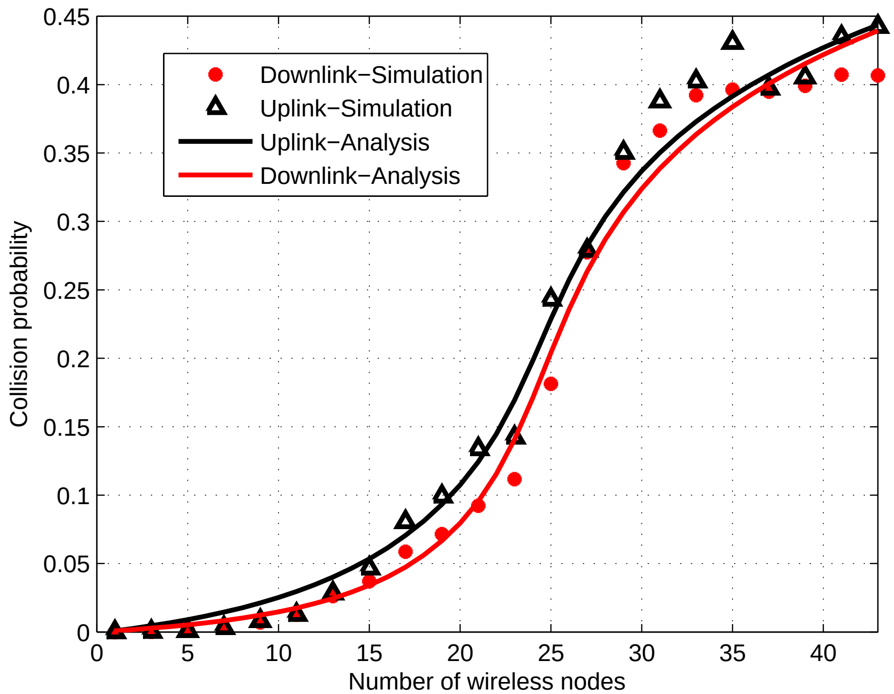

Figure 5.3 Collision probability for 80 bytes voice payload.

time includes backoff defer, i collision time and one final successful transmission time provided that after i failures due to collision, the transmission is successful. Denote $\overline{T}^i$ as the time needed for a successful transmission at the i^{th} backoff stage. We can find $\overline{T}^i$ as,

$$\overline{T}^i = \overline{B}^i + iT_c + T_s \tag{5.42}$$

Since the backoff counter in the j^{th} stage is uniformly chosen in the range of $[0, W_j]$, we can write the average backoff defer period $\overline{B}^i$ as:

$$\overline{B}^i = \sum_{j=0}^{i} \frac{W_j - 1}{2} \, E[slot] \tag{5.43}$$

Finally, we can find the average packet service time $1/\mu$:

$$\frac{1}{\mu} = \sum_{i=0}^{m} \frac{p^i(1-p)}{1-p^{m+1}} \, \overline{T}^i \tag{5.44}$$

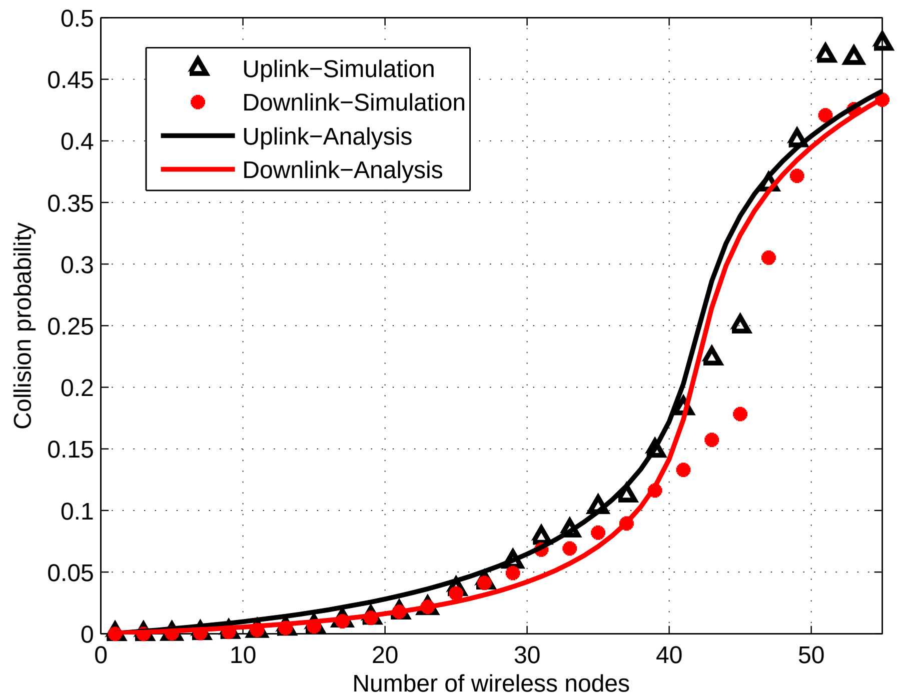

Figure 5.4 Collision probability for 160 bytes voice payload.

where p is the collision probability.

Let p_u be the collision probability corresponding to the packets transmitted in the uplink direction and p_d be the collision probability seen by the packets in the downlink direction. We can find the average packet service time for the packets served by the wireless station $1/\mu_u$ and the average packet service time for the packets transmitted by the access point $1/\mu_d$ from the following equations.

$$\frac{1}{\mu_u} = \sum_{i=0}^{m} \frac{p_u^i(1-p_u)}{1-{p_u}^{m+1}}\,\overline{T}^i \tag{5.45}$$

$$\frac{1}{\mu_d} = \sum_{i=0}^{m} \frac{p_d^i(1-p_d)}{1-{p_d}^{m+1}}\,\overline{T}^i \tag{5.46}$$

Figure 5.5 shows the packet service time as a function of the number of wireless stations when the packet payload is equal to 80 bytes. Packets in the downlink direction and sent from the access point to the wireless stations have a lower service

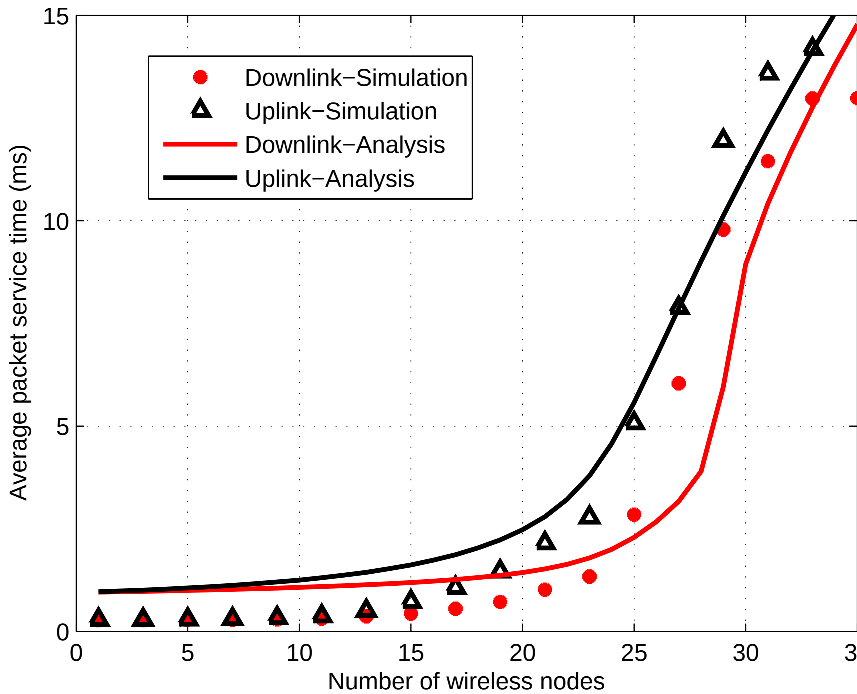

Figure 5.5 Average packet service time for 80 bytes voice payload.

time in comparison with the packets sent from the wireless stations to the access point in the uplink direction.

Figure 5.6 shows packet service as a function of the number of active voice flows when the packet payload is equal to 160 bytes. It is seen that the average packet service time remains almost negligible for a maximum number of 40 voice connections. But when we use smaller packet payload which is equal to 80 bytes, the maximum number of active voice flows for which the packet service time remains small (less than 5 ms), is equal to 25. Enlarging the packet payload will result in smaller service time and permits more voice flows over WLANs.

We notice that according to Equations (5.44)-(5.46), the packet service time depends on several parameters such as the average successful transmission time T_s, the average collision time T_c, the average defer time due to the backoff procedure $\overline{B}^i$, the probability of collision in the uplink direction p_u and the probability of collision in the downlink direction p_d. However, the collision probability is the key parameter for finding the packet service time. Therefore, backoff modeling based on the two-dimensional Markov chain as shown in Figure 5.1 and finding the collision probability p_u and p_d are the necessary steps for packet service time analysis. Packet service time is the key parameter for the queueing analysis in Section (5.6) to find the average packet delay. Unlike the complex two-dimensional Markov chain used for the back-off modeling, we can use a simple one-dimensional Markov diagram to evaluate the queue characteristics of a tagged station when there is only one type of traffic in the network and the packet payload is fixed.

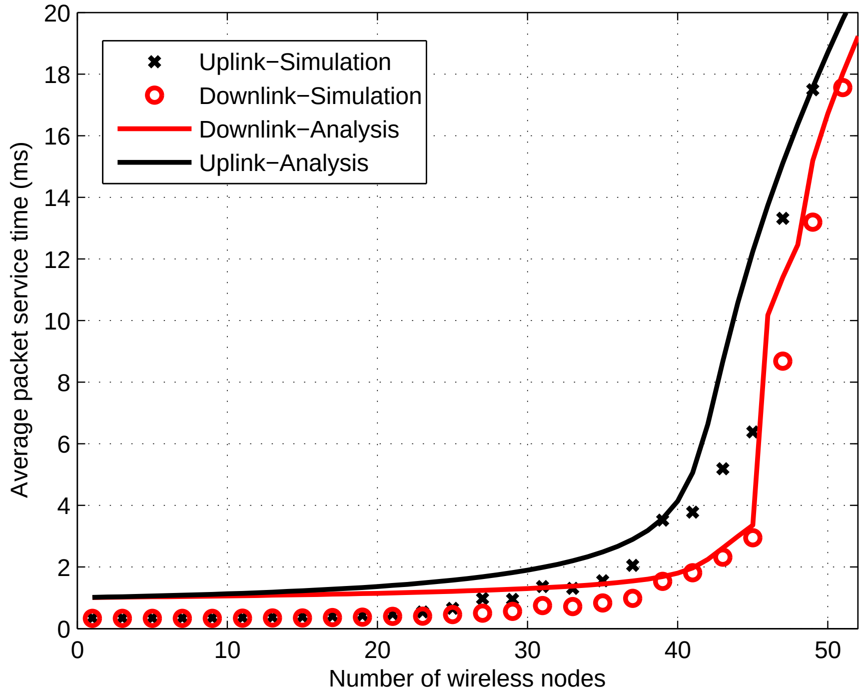

Figure 5.6 Average packet service time for 160 bytes voice payload.

5.5 Queue analysis for one type of traffic

To analyze delay and throughput, we evaluate queue parameters of a tagged station using a one-dimensional Markov chain. To apply a queuing model we need to figure out the queue specification such as arrival rate, service time and size of queue. In an M/M/1/K model as shown in Figure 5.7 where the arrival rate is Poisson and the service time is exponentially distributed, we can find steady-state probabilities using the product form solution [37]. Denoting the total number of packets in the system by the random variable N, the steady-state probability that there are n packets in the system is:

$$P(N = n) = \frac{1-\rho}{1-\rho^{K+1}}\ \rho^n \qquad n = 0, 1, \cdots, K \tag{5.47}$$

where $\rho = \lambda/\mu$, λ is the packet arrival, μ is the service parameter and K is the buffer size. Using little's theorem, we can find that the average delay is:

$$d = \frac{\bar{N}}{\lambda} = \frac{\rho + \rho^{K+1}[\rho K - K - 1]}{\lambda(1-\rho)(1-\rho^K)} \tag{5.48}$$

Finally, if we replace $n = K$ in Equation (5.47) we obtain blocking probability as:

$$P_{loss} = P(N = K) = \frac{1-\rho}{1-\rho^{K+1}}\ \rho^K \tag{5.49}$$

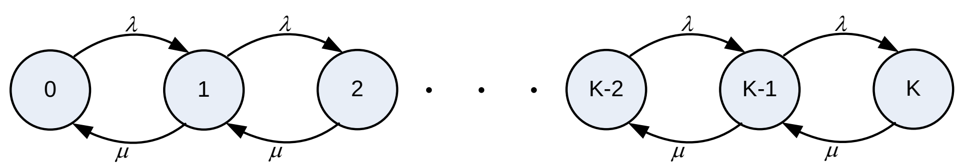

Figure 5.7 M/M/1/K Markov model for one class traffic.

Even though the M/M/1/K model is an approximation, it can help us find important parameters such as the average packet delay and the packet delay outage probability. Note that the Poisson arrival assumption is closer to reality in the access point where there is multiplexing of several traffic flows. We also note that the downlink characteristics are more vital because of the fact that the access point is the bottleneck of the 802.11 network.

5.6 Packet delay

Assuming that the wired domain of the system is ideal and the processing delay is negligible, we can write the end-to-end packet delay $\overline{d}$ as:

$$\overline{d} = \overline{w} + \frac{1}{\mu} \tag{5.50}$$

where $\overline{w}$ is the average waiting time for a packet in the queue at the MAC layer.

Let w_u be the average waiting time for the packets at the MAC layer of a wireless station, w_d be the average waiting time for the packets on the MAC layer at the access point, $\overline{d}_u$ be the average end-to-end packet delay in the uplink direction and $\overline{d}_d$ be the average end-to-end packet delay in the downlink direction. We can write:

$$\overline{d_u} = \overline{w_u} + \frac{1}{\mu_u} \tag{5.51}$$

$$\overline{d_d} = \overline{w_d} + \frac{1}{\mu_d} \tag{5.52}$$

Knowing the packet arrival rate into the buffer of the MAC layer of the wireless station λ_u and the packet arrival rate into the buffer of the MAC layer of the access point λ_d, we can use the M/M/1/K model to find the average waiting time as following:

$$\overline{w}_u = \frac{\rho_u + \rho_u^{K+1}[\rho_u K - K - 1]}{\lambda_u(1-\rho_u)(1-\rho_u^K)} \tag{5.53}$$

$$\overline{w}_d = \frac{\rho_d + \rho_d^{K+1}[\rho_d K - K - 1]}{\lambda_d(1-\rho_d)(1-\rho_d^K)} \tag{5.54}$$

where $\rho_u = \lambda_u/\mu_u$, $\rho_d = \lambda_d/\mu_d$, and K is the buffer size at the MAC layer which is set to 100 packets.

Average packet delay as a function of the number of wireless stations for 80 bytes packet payload is shown in Figure 5.8. Although the packet service time for downlink traffic is less than uplink as discussed in the previous section, the end-to-end delay of downlink packets is much larger than the uplink packets because of a large queueing delay at the MAC layer buffer of the access point. We observe that the downlink average packet delay increases exponentially while the uplink packet delay remains negligible for 18 simultaneous voice flows. When we use larger packet payload which is equal to 160 bytes, as shown in Figure 5.9, up to 21 voice connections can be admitted before the downlink delay starts to increase significantly and the uplink delay remains almost zero even if the number of wireless stations in the network is equal to 35.

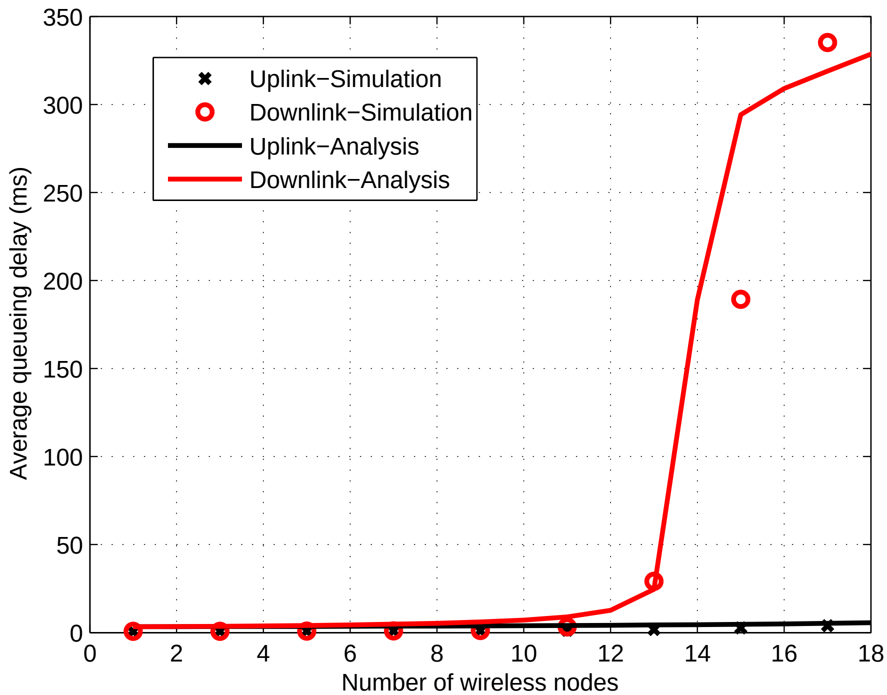

Figure 5.8 Average packet delay for 80 bytes voice payload.

5.7 Throughput

In this section, we represent the analytical and simulation results for the uplink, downlink and overall throughput of the network. To find the throughput of the downlink we investigate all possible events during two consecutive successful transmission from the access point. There may be collisions or successful transmission from other wireless stations in the uplink direction as well as idle slots. We define the throughput as the amount of bits per second successfully transmitted. As a result, the network

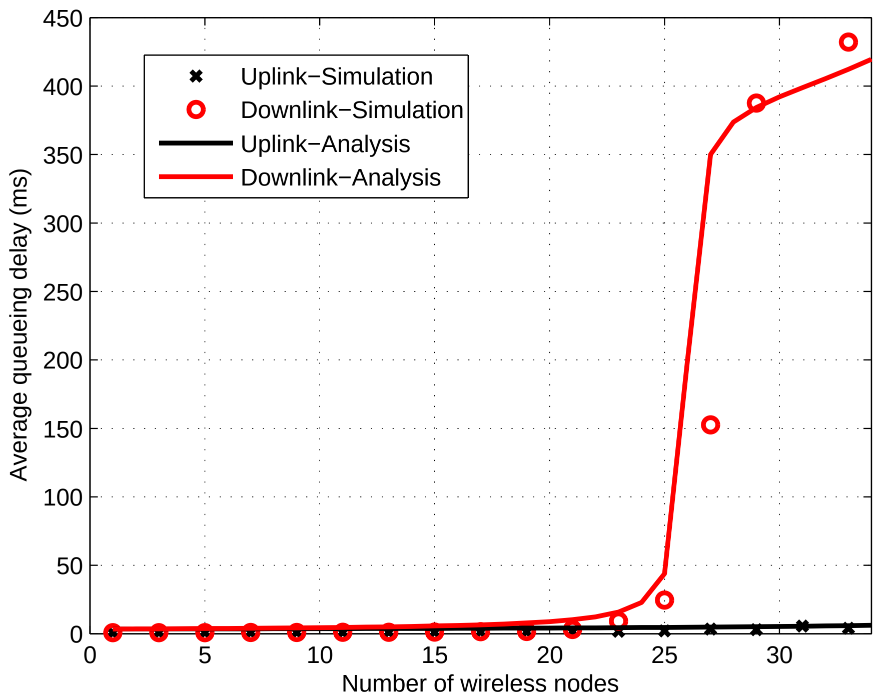

Figure 5.9 Average packet delay for 160 bytes voice payload.

downlink throughput S_d can be written as

$$S_d = \frac{P_{d,s}\ P_{tr}\ E[P]}{(1 - P_{tr})\ \sigma + P_{tr}\ P_s\ T_s + P_{tr}\ (1 - P_s)\ T_c} \tag{5.55}$$

where $E[P]$ represents the payload information transmitted in a slot time.

Between two successful transmission of a tagged wireless station, there may be successful transmissions from the access point or other wireless stations, collisions or idle slots. Network throughput in the uplink direction for all wireless stations can be written as following.

$$S_u = \frac{P_{u,s}\ P_{tr}\ E[P]}{(1 - P_{tr})\ \sigma + P_{tr}\ P_s\ T_s + P_{tr}\ (1 - P_s)\ T_c} \tag{5.56}$$

We can also find the total throughput or overall throughput S_t which is simply the sum of the uplink and downlink throughput:

$$S_t = S_d + S_u \tag{5.57}$$

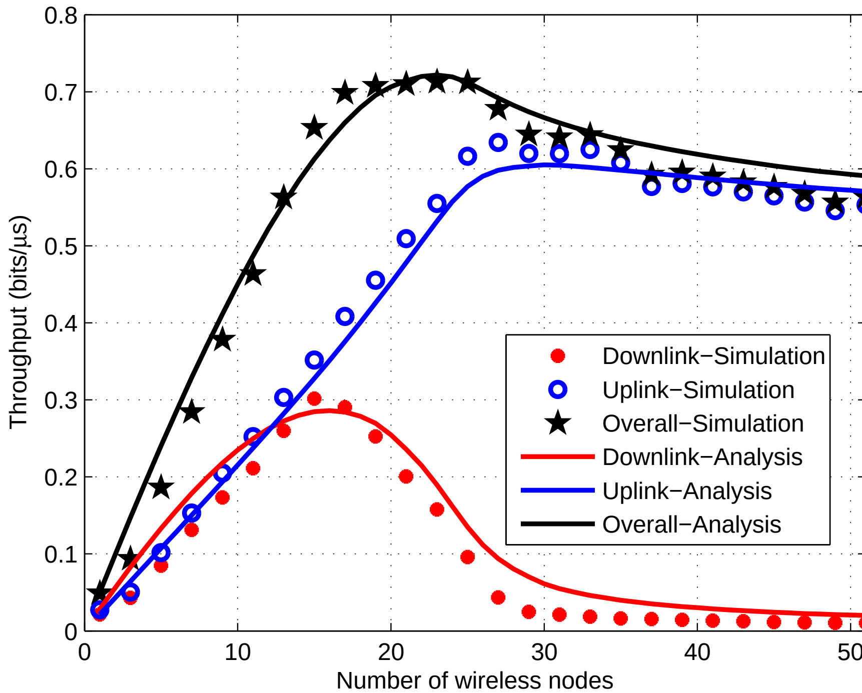

Figure 5.10 Throughput for 80 bytes voice payload.

Figure 5.10 shows throughput as a function of the number of wireless nodes. As it can be seen, throughput linearly increases while we add more voice connections. For both uplink and downlink there exists a particular number of wireless nodes for which the maximum throughput is achieved. For 80 bytes voice packets, the maximum throughput of 0.3 bits/μs for the access point is achieved when there are 15 voice connections in the network. Throughput decreases linearly when we admit more than 15 voice connections and it reaches approximately zero when there are 30 or more voice connections in the network. This significant degradation of throughput is a result of high collision probability and buffer overflow at the access point. For the uplink direction, a maximum throughput of 0.63 bits/μs is achieved when the number of wireless nodes equals 27 and throughput degrades when we increase the number of wireless nodes. The reduction of throughput is much slower in the downlink rather than in the uplink due the multiplexing of all flows at the access point. Figure 5.11 shows throughput for 160 bytes packets.

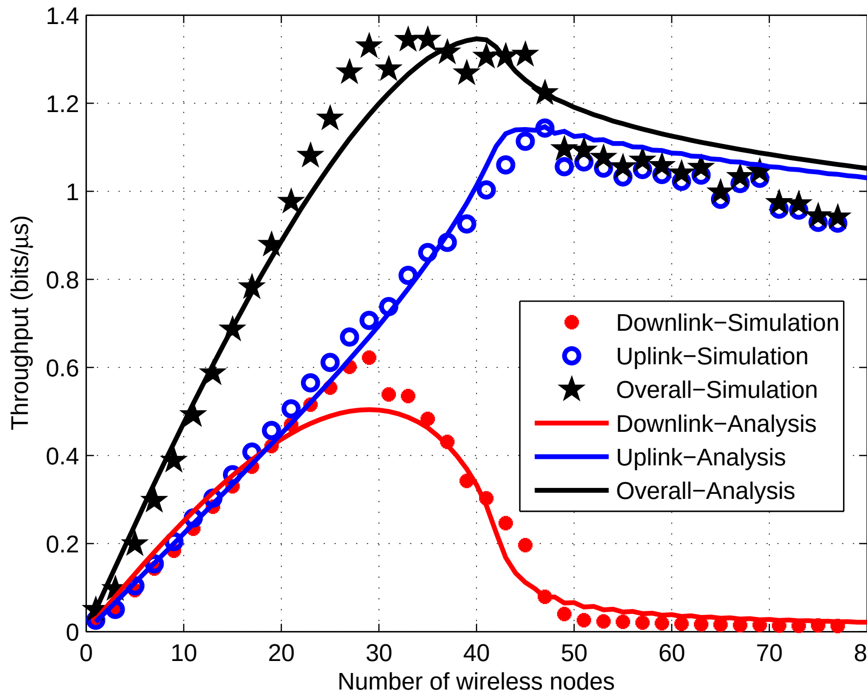

Figure 5.11 Throughput for 160 bytes voice payload.

5.8 Admission control

In this section, we propose a model-based admission control technique which can be deployed in the infrastructure mode of WLANs to guarantee the QoS for VoIP applications. This algorithm takes into account all system parameters such as contention window size, number of backoff stages, buffer size at the MAC layer, traffic characteristics such as average packet arrival rate and packet payload, and the number of active voice flows in the network. The decision of admitting or blocking an application is based on the packet outage probability. We define packet outage probability as the probability that a packet in the downlink direction experiences a delay which exceeds a minimum acceptable value T_0 as the quality requirement for voice flow. Because of the traffic multiplexing at the AP, there would be a large waiting time for packets to reach the head of the queue at the MAC layer which is not desirable for delay sensitive applications such as VoIP. As we already adopted an M/M/1/K model for queueing analysis, we can show that the waiting time is exponentially distributed with parameter $\mu(1-\rho)$. Therefore, we can find the outage probability for the wireless station $P(d_u > T_0)$ and the outage probability for the access point $P(d_d > T_0)$ as:

$$P(d_u > T_0) = \rho_{u,e}\ e^{-\mu_u(1-\rho_{u,e})T_0} \tag{5.58}$$

$$P(d_d > T_0) = \rho_{d,e}\ e^{-\mu_d(1-\rho_{d,e})T_0}, \tag{5.59}$$

where $\rho_{u,e} = \lambda_{u,e}/\mu_u$ and $\rho_{d,e} = \lambda_{d,e}/\mu_d$. For a finite buffer we can define the effective arrival rate into the queues of a tagged wireless station and the access point as $\lambda_{u,e} = \lambda_u(1-P_{u,b})$ and $\lambda_{d,e} = \lambda_d(1-P_{d,b})$, where $P_{u,b}$ and $P_{d,b}$ are the blocking probabilities at the MAC layer of the wireless station and the access point. We compute $P_{u,b}$ and $P_{d,b}$ using Equation (5.49).

When a new voice connection request is received, the admission control unit of the access point predicts the packet outage probability after admitting the new voice request. If packet outage probability is smaller than the predefined threshold T_0, then it would admit the new user. If packet outage probability is larger than the threshold it means that a new connection will degrade the QoS requirements of the existing flows. Therefore, the new connection is blocked until the network exits from the heavy loaded status.

To find the maximum number of voice connections that can be provided with the acceptable QoS requirements, we use the metric introduced in Equation (5.59). Accepting a new call will increase the number of contending stations n by 1. This will increase the average arrival rate at the MAC layer buffer of the access point, the probability of collision, the packet service time, the packet average delay and the packet delay outage probability as defined in Equation (5.59). Therefore accepting a new call will result in the performance degradation for all existing calls over 802.11 network and there exists a critical point for which the performance (the packet average delay) violates the acceptable threshold delay. This will occur at the MAC layer buffer of the access point because of the multiplexing of several traffic streams in the downlink direction.

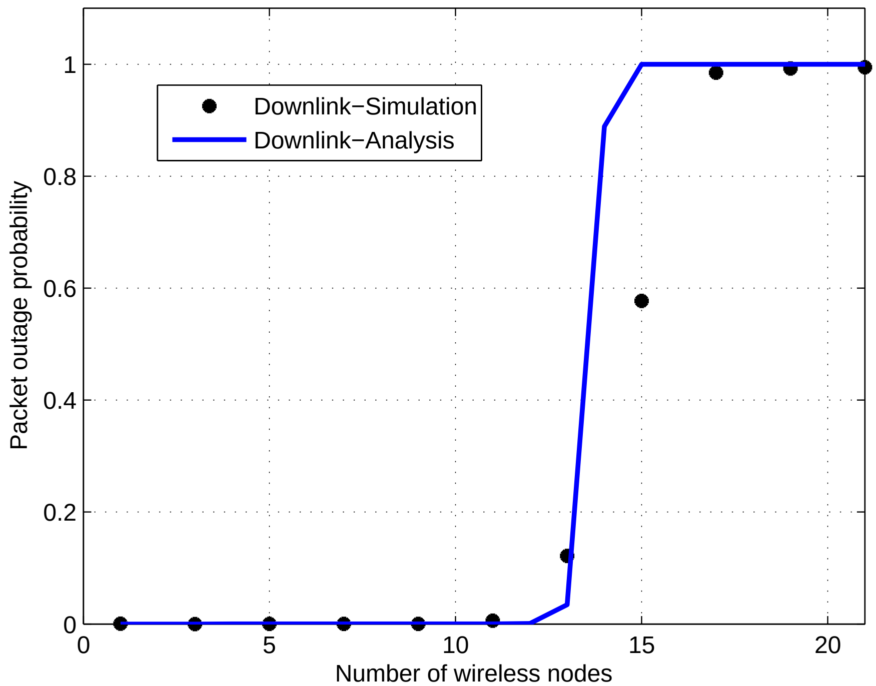

Figure 5.12 Packet delay outage probability for 80 bytes voice payload.

Figure 5.12 shows the outage probability as a function of the number of wireless stations or the number of active voice flows in the 802.11b network. A 75 ms delay bound as the threshold T_0 is considered. As it can be seen, when we admit more than 12 voice connections, the packet outage probability in downlink starts to increase significantly which degrades the QoS parameters of all active voice flows. Using the admission control technique, the maximum number of voice sources for this specific scenario should not exceed 12 in order to avoid QoS degradation of active flows. As shown in Figure 5.13, the maximum number of voice connections for the same scenario with a different packet payload which is equal to 160 bytes should not exceed 21 for fulfilling the QoS requirements.

To investigate the effect of the system parameters and traffic characteristics on the maximum number of voice connections over 802.11 network, we plot the packet delay outage probability in Figure 5.14 as a function of the number of wireless stations for different traffic specifications in the IEEE 802.11g network. We use two different packetization intervals 10 ms and 20 ms according to the G.711 codec standard. We also define parameter γ as the ratio of the average *on* period to the average $on + off$

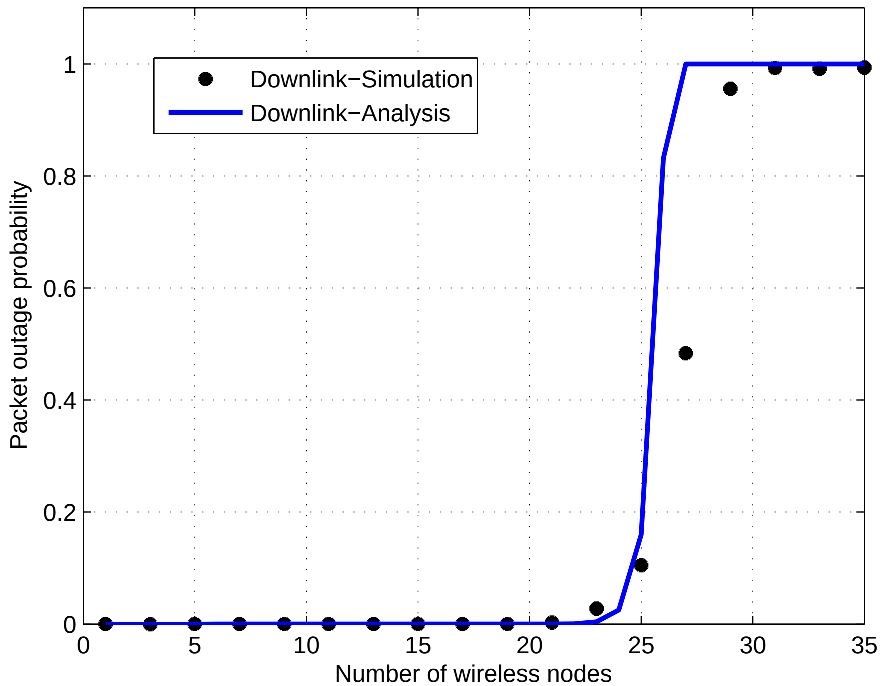

Figure 5.13 Packet delay outage probability for 160 bytes voice payload.

period. Therefore, we can find the parameter γ from the following equation.

$$\gamma = \frac{on}{on + off} = \frac{\beta}{\alpha + \beta} \tag{5.60}$$

If the average `on` and `off` periods are equal to 400 ms and 600 ms respectively, then $\gamma = 0.4$. If we set $\gamma = 1$, the average `off` period β tends to zero. In other words, $\gamma = 1$ represents the constant bit rate (CBR) traffic modeling of the voice application with $\lambda_u = \lambda$ and $\lambda_d = n\,\lambda$. It is seen from Figure 5.14 that when we increase the packetization interval from 10 ms to 20 ms, we can provide more voice connections. When we choose $\gamma = 1$, the maximum number of voice connections decreases in comparison with $\gamma = 0.4$. This is mainly due to the direct dependency of the traffic load at the access point on traffic specifications according to Equation (5.2). Choosing a packetization interval of 10 ms rather than 20 ms is equivalent to doubling the packet arrival rate λ during the `on` period.

In addition to the traffic characteristics, the system parameters also affect the

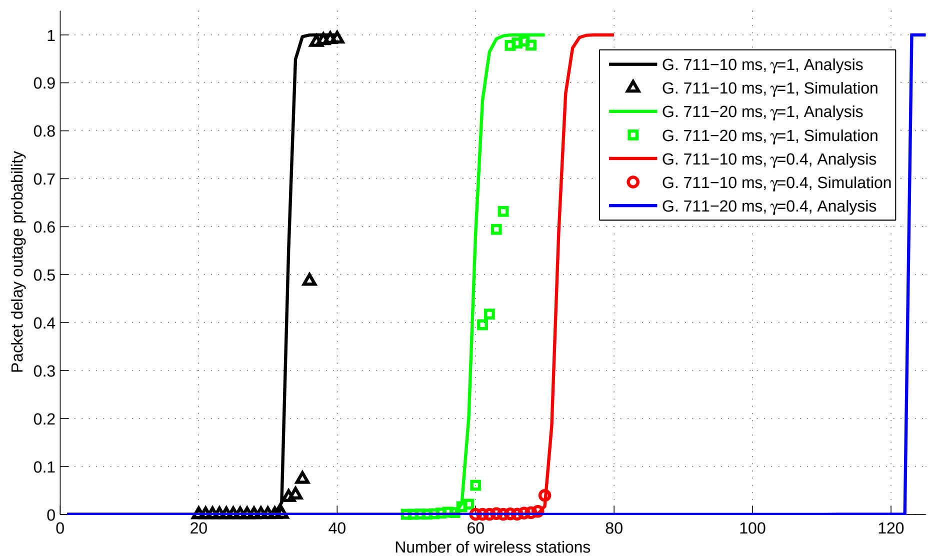

Figure 5.14 Packet delay outage probability as a function of number of wireless stations in 802.11g network.

maximum number of voice connections in the 802.11 network. As it can be seen in Table 5.1, The channel bit rate is increased from 11 Mbps to 54 Mbps and the basic bit rate is increased from 1 Mbps to 6 Mbps in the 802.11g standard. The PHY header length is decremented from 24 B to 15 B and the ACK length is decremented from 38 B to 29 B. There are smaller values for slot time and DIFS and the minimum contention window is decreased by half. As a result, the average time for a successful transmission T_s and the average time that the channel is sensed busy because of a collision T_c are decremented according to Equations (5.25) and (5.26). Therefore, the average packet service time is significantly decreased in 802.11g network and this will permit the 802.11g network to provide the minimum acceptable QoS requirements to more active flows in comparison with the 802.11b network.

Voice capacity equals to the maximum number of wireless stations n that satisfies the QoS requirement which is defined as $P(d_d > T_0) < 0.01$ and guarantees that the packet delay outage probability at the MAC layer buffer of the access point is less than 0.01 . Figure 5.15 shows the packet delay outage probability as a function of number of wireless stations and the region that satisfies the QoS requirements in the IEEE 802.11g networks. We have also performed extensive simulations to validate

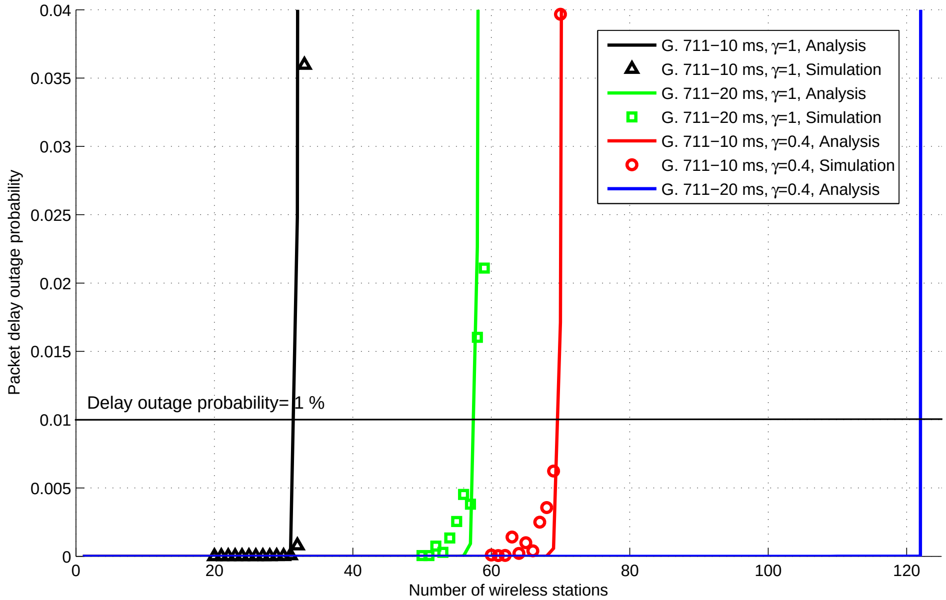

Figure 5.15 Packet delay outage probability as a function of number of wireless stations in 802.11g network and the region that satisfies the QoS requirements.

Table 5.2 Voice capacity of 802.11b and 802.11g networks for different packetization intervals (10 ms and 20 ms) and different `on` and `off` periods ($\gamma = 1$ and $\gamma = 0.4$).

Packetization Interval (ms)	γ	Capacity 802.11b		Capacity 802.11g	
		Proposed Analysis	Simulation	Proposed Analysis	Simulation
10	1	5	5	31	32
20	1	10	10	57	57
10	0.4	12	11	69	69
20	0.4	21	22	122	$\times$

the analytical results. Table 5.2 summarizes the voice capacity for several traffic specifications in 802.11b and 802.11g networks. The major advantage of our research is the analytical framework which can theoretically estimate the voice capacity without the need of numerous simulations. This would be considerably helpful when we apply new traffic specifications or different system parameters based on the new versions of the 802.11 networks.

Chapter 6

Conclusions and Future Work

In this thesis, we studied QoS provisioning for voice application over WLANs. Since the MAC layer significantly affects the QoS parameters of different applications, we examined the performance of the DCF access mechanism at the MAC layer of IEEE WLANs. We have considered a wired/wireless network in which voice connections exist between each wireless station in the wireless domain and a wired station in the wired domain. Voice traffic is modeled using the two-state `on-off` process which is a realistic model according to the existence of the active and silent intervals as defined in the G.711 codec standard.

We have proposed a model-based admission control scheme based on the packet delay outage probability metric which plays an important role in fulfilling the QoS requirements of delay sensitive applications such as voice and video. We have found the packet delay outage probability by modeling the MAC layer buffer of the access point using an M/M/1/K queueing model. To solve the M/M/1/K model, we have found the average packet service time after solving a set of nonlinear equations according to the two-dimensional Markov model representing the backoff procedure. We note that the packet service time depends on several system parameters and traffic characteristics. Since we compute the packet delay outage probability by applying the packet service time calculated in the proposed analytical framework in the M/M/1/K queueing model, the proposed model-based admission control method takes into account all system parameters and traffic characteristics. Even though the M/M/1/K model is an approximation it can help us find the average packet delay and the packet delay outage probability. We have verified the accuracy of the proposed model by performing ns-2 simulations. Another contribution of this part of the research is the theoretical estimation of the voice capacity without the need of numerous simulations. This would be considerably helpful when we apply new traffic specifications or different system parameters based on the new versions of the 802.11 networks. Applying the proposed model-based admission control technique, the access point can predict the packet outage probability before admitting a new call. Therefore it will deny or admit the new connection after solving the proposed model for MAC layer analysis and evaluating the voice delay outage probability metric. The admission control can be implemented on the access point to prevent network congestion and voice quality degradation. We have also investigated the impact of the traffic specifications and system parameters on the maximum number of voice connections in 802.11b and 802.11g networks.

In what follows, we briefly present the major contributions of each chapter. After providing a review of the previous work in the area of MAC layer performance analysis and enhancement, we proposed a novel analytical framework for MAC layer analysis in the infrastructure mode of WLANs. There are three assumptions in the proposed analysis:

1. The network is in the non-saturated mode which is a realistic assumption when each voice flow is characterized using the `on-off` model.
2. The collision probability of a packet transmitted by each wireless station in the uplink direction is different from the probability of collision for the packets transmitted from the access point in the downlink direction.
3. The channel is ideal.

Unlike previous work, we have provided a more detailed analysis by differentiating between backoff model for the uplink and downlink directions. Although each wireless station and the access point use the same backoff procedure to access the shared bandwidth, the packet transmission probability is different because they do not have the same arrival rate and the idle period in each wireless station is significantly larger in comparison with the idle period in the access point. This will also affect the backoff formulation because of the existence of the idle state and the transition probabilities that depend on the amount of traffic arrival rate. In contrast to previous studies, we apply two groups of equations, one group is defined for the wireless station and the other one for the access point. There are three nonlinear equations assigned to each tagged station. These equations represent the transmission probability, probability of collision and the probability of being in the busy state in terms of the number of wireless stations, the traffic arrival rate and system parameters such as contention window size, number of backoff stages and buffer size at the MAC layer. We have used the Poisson arrival approximation to find the probability that at least one packet is at the MAC layer buffer to transmit. This assumption is good since the constant arrival interval of the voice packets during the `on` period is in the order of milliseconds and the average slot time is in the order of microseconds. The Poisson arrival assumption is more realistic at the MAC layer buffer of the access point since there is the superposition of n traffic streams from n independent sources of voice. We computed several important parameters such as probability of collision, packet service time, packet delay and throughput after applying numerical methods for solving a set of nonlinear equations based on the proposed model. This model is capable of evaluating the asymmetric contention mechanism at the MAC layer of IEEE WLANs. Downlink traffic multiplexing at the AP does not speak in favor of uplink traffic from wireless stations. However, it causes larger collision probability and packet service time experienced by packets in the uplink direction. It also significantly degrades downlink throughput because of the buffer overflow.

Future work on MAC layer analysis, can involve the co-existence of different applications such as data, voice and video over WLANs. This can be done by modifying the proposed analytical framework to include a new set of equations assigned to the new traffic type as well as expressing the probability of being in the busy state according to the particular traffic modeling of the new application or the new codec standard. We also need to modify the backoff Markov model to take into account the errors due to the wireless link. By deploying more complex queueing models, we can extend the proposed analytical framework to theoretically find the maximum number of data, voice or video traffic flows that a 802.11 WLAN can support. Finally, we can find the optimal values of the size of the contention window, the maximum number of retransmissions, the `on` and `off` periods, etc to achieve the maximum throughput performance, while provisioning the QoS requirements for different applications.

References

[1] IEEE, *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications.* IEEE Computer Society/Local and Metropolitan Area Networks, 1999.

[2] IEEE 802.11e/D13.0, Part 11, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS).* draft supp. to IEEE 802.11 std., Jan 2005.

[3] Doc IEEE 802.11-06/0328r0, *802.11 TGs Simple Efficient Extensible Mesh (SEE-Mesh) Proposal.* IEEE 802.11s draft, Feb 2000.

[4] G. Lazarou and V. Frost, "Variance-time curve for packet streams generated by exponentially distributed on/off sources," *Communications Letters, IEEE*, vol. 11, no. 6, pp. 552–554, june 2007.

[5] H. S. Chhaya and S. Gupta, "Throughput and fairness properties of asynchronous data transfer methods in the IEEE 802.11 MAC protocol," in *Sixth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'95.*, vol. 2, Sep. 1995, pp. 613–617.

[6] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs," in *Seventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'96.*, vol. 2, Oct. 1996, pp. 392–396.

[7] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement," in *Proceedings Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. IEEEINFOCOM '98.*, vol. 1, Mar./Apr. 1998, pp. 142–149.

[8] T.-S. Ho and K.-C. Chen, "Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol," in *Seventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'96.*, vol. 2, Oct. 1996, pp. 407–411.

[9] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[10] J.-H. Yun, "Throughput analysis of IEEE 802.11 WLANs with automatic rate fallback in a lossy channel," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 689–693, Feb. 2009.

[11] 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 *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. INFOCOM '02.*, vol. 2, Jun. 2002, pp. 599–607.

[12] C. Foh and M. Zukerman, "Performance analysis of the IEEE 802.11 MAC protocol," in *Proc. Eur. Wireless*, 2002.

[13] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 protocol: design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Select. Areas Commun.*, vol. 18, no. 9, pp. 1774–1786, Sep. 2000.

[14] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, "On the linear behaviour of the throughput of IEEE 802.11 DCF in non-saturated conditions," *IEEE Commun. Lett.*, vol. 11, no. 11, pp. 856–858, Nov. 2007.

[15] I. N. Vukovic and N. Smavatkul, "Delay analysis of different backoff algorithms in IEEE 802.11," in *IEEE 60th Vehicular Technology Conference VTC '04.*, vol. 6, Sep. 2004, pp. 4553–4557.

[16] P. Raptis, V. Vitsas, K. Paparrizos, P. Chatzimisions, A. C. Boucouvalas, and P. Adamidis, "Packet delay modeling of IEEE 802.11 wireless LANs," in *Proc. Eur. Wireless*, 2005, pp. 184–190.

[17] H. Zhai and Y. Fang, "Performance of wireless LANs based on IEEE 802.11 MAC protocols," in *Proceedings 14th IEEE Personal, Indoor and Mobile Radio Communications, PIMRC '03.*, vol. 3, Sep. 2003, pp. 2586–2590.

[18] H. Zhai, X. Chen, and Y. Fang, "How well can the IEEE 802.11 wireless LAN support quality of service?" *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3084–3094, Nov. 2005.

[19] O. Tickoo and B. Sikdar, "A queueing model for finite load IEEE 802.11 random access MAC," in *IEEE International Conference on Communications*, vol. 1, Jun. 2004, pp. 175–179.

[20] D. Xu, T. Sakurai, and H. L. Vu, "An access delay model for IEEE 802.11e EDCA," *IEEE Transactions on Mobile Computing*, vol. 8, no. 2, pp. 261–275, Feb. 2009.

[21] A. Abdrabou and W. Zhuang, "Service time approximation in IEEE 802.11 single-hop Ad Hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 305–313, Jan. 2008.

[22] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Networking*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[23] I. Ada and C. Castelluccia, "Differentiation mechanisms for ieee 802.11," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, vol. 2, 2001, pp. 209–218.

[24] A. Veres, A. T. Campbell, M. Barry, and L.-H. Sun, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.

[25] J. L. Sobrinho and A. S. Krishnakumar, "Real-time traffic over the IEEE 802.11 medium access control layer," *Bell Labs Tech. J.*, vol. 1, pp. 172–187, 1996.

[26] S. Garg and M. Kappes, "Can I add a VoIP call?" in *IEEE International Conference on Communications, ICC '03.*, vol. 2, May 2003, pp. 779–783.

[27] K. Medepalli, P. Gopalakrishnan, D. Famolari, and T. Kodama, "Voice capacity of IEEE 802.11b, 802.11a and 802.11g wireless LANs," in *IEEE Global Telecommunications Conference GLOBECOM '04.*, vol. 3, Nov./Dec. 2004, pp. 1549–1553.

[28] D. Gu and J. Zhang, "A new measurement-based admission control method for IEEE 802.11 wireless local area networks," *Mitsubishi Elec. Research Lab. Tech. rep.*

[29] L. Zhang and S. Zeadally, "HARMONICA: enhanced QoS support with admission control for IEEE 802.11 contention-based access," in *Proceedings. 10th IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS '04.*, May 2004, pp. 64–71.

[30] D. Pong and T. Moors, "Call admission control for IEEE 802.11 contention access mechanism," in *IEEE Global Telecommunications Conference, GLOBECOM '03.*, vol. 1, Dec. 2003, pp. 174–178.

[31] D. Zhen-ning, K. amd Tsang and B. Bensaou, "Measurement-assisted model-based call admission control for IEEE 802.11e WLAN contention-based channel access," in *13th Workshop on local and metropolitan area networks*, Apr. 2004, pp. 55–60.

[32] D. Gao, J. Cai, and K. N. Ngan, "Admission control in IEEE 802.11e wireless LANs," *IEEE Network*, vol. 19, no. 4, pp. 6–13, Jul./Aug. 2005.

[33] H. Zhai, X. Chen, and Y. Fang, "A call admission and rate control scheme for multimedia support over IEEE 802.11 wireless LANs," *Springer*, 2006.

[34] C. Ortiz, J.-F. Frigon, B. Sansò, and A. Girard, "Effective bandwidth evaluation for VoIP applications in IEEE 802.11 networks," in *International Wireless Communications and Mobile Computing Conference, IWCMC '08.*, Aug. 2008, pp. 926–931.

[35] A. Gnassou, J.-F. Frigon, and B. Sansò, "Impact of wireless channel on VoIP QoS and admission regions in IEEE 802.11g WLANs," in *IEEE International Conference on Wireless and Mobile Computing, WIMOB '08.*, Oct. 2008, pp. 63–68.

[36] S. Mccanne, S. Floyd, and K. Fall, "ns2 (network simulator 2)," http://www-nrg.ee.lbl.gov/ns/. [Online]. Available: http://www-nrg.ee.lbl.gov/ns

[37] D. Bertsekas and R. Gallager, *Data networks.* Prentice-Hall, Inc., 1987.