

Context Aware Handover for WiFi and Its Extension to WiMAX



Abhijit Sarma



Context Aware Handover for WiFi and Its Extension to WiMAX

*Thesis submitted in partial fulfillment of the requirements
for the degree of*

Doctor of Philosophy

by

Abhijit Sarma

Under the supervision of
Prof. Sukumar Nandi



Department of Computer Science and Engineering
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Guwahati 781039, India
June, 2014



Declaration

I certify that

- a. The work contained in this thesis is original and has been done by myself and the general supervision of my supervisor.
- b. The work has not been submitted to any other Institute for any degree or diploma.
- c. Whenever I have used materials (data, theoretical analysis, results) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.
- d. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Place: IIT Guwahati

Abhijit Sarma

Date:

Research Scholar

Department of Computer Science and Engineering,

Indian Institute of Technology Guwahati,

Guwahati, INDIA 781039



Certificate

This is to certify that the thesis entitled “**Context Aware Handover for WiFi and Its Extension to WiMAX**” being submitted by **Abhijit Sarma** to the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, is a record of bona fide research work under my supervision and is worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

Place: IIT Guwahati

Prof. Sukumar Nandi

Date:

Department of Computer Science and Engineering,
Indian Institute of Technology Guwahati,
Guwahati, INDIA 781039



Acknowledgements

During this thesis work, I am supported by many people. I take the opportunity to express my gratitude to all of them.

First of all I am highly indebted to my supervisor Prof. Sukumar Nandi who provided great support during my PhD period, and generously helped clearing issues for completion of this research work. I thank him for all his help. I enjoyed spending time with him both at work and otherwise.

I am also highly grateful to my Doctoral Committee members, Prof. P Bhaduri, Prof. G. Sajith, Prof A. Mitra, and Prof. Partha Sarathi Mandal. Their comments and suggestions have helped me put my work in the right tract. I thank them all.

I express my sincere thanks to Prof. S. B. Nair, the present Head of the Department of Computer Science and Engineering, for the facilities and environment provided for research. I would also like to express my gratitude to the Director, the Deans and other managements of IIT Guwahati who contributed to make IIT Guwahati a highly regarded institute for research. I am thankful to all the faculties and the staffs of the Department of Computer science and Engineering for their cooperation and supports.

I thank the research scholars, M. Tech and the B. Tech students of this institute, with whom I have closely worked. I thank R. Gupta, S. Joshi and A. Choubey for their help. I am highly grateful to Dr. Sandip Chakraborty for investing his time hepling me to complete this work.

I thank my wife Mamoni and my daughters Loly and Lopa for their love and patience during my research work.

Place: IIT Guwahati

Date:

Abhijit Sarma

Research Scholar

Department of Computer Science and Engineering,
Indian Institute of Technology Guwahati,
Guwahati, INDIA 781039



Abstract

IEEE 802.11 or 'Wireless Fidelity' has become a popular wireless technology to offer high speed Internet access at public places called the 'Hot-Spots' as well as to support ubiquitous Internet connectivity through institute wide wireless local area networks (WLANs). However, existing researches has shown that due to wide-spread deployments of WiFi based network connectivity zones, more numbers of wireless access points (APs) are deployed than requirements, however, users tend to concentrate at few areas making traffic load imbalance across the network. The design philosophy of IEEE 802.11 connection establishment and handover from one AP to another is based on signal strength which is biased towards the distance between the AP and the client nodes. Severe performance and quality of service (QoS) degradation and capacity underutilization are observed due to this imbalance traffic distribution, which is the main concern of research in this thesis.

The first contribution of the thesis explores the inherent problems of IEEE 802.11 handover management policies, and proposes a context-aware handover mechanism to balance traffic load across the network. The proposed mechanism works in coordination of information exchange between the AP and the wireless client that experiences performance degradation due to traffic overload at its present point of attachment. This coordination helps the wireless client to perform a horizontal handover to another AP in the vicinity, that significantly improves the network capacity. The performance of the proposed context aware handover mechanism is analyzed using theoretical analysis as well as from practical testbed results. The second contribution of the thesis extends the context aware handover to incorporate multiple traffic classes, where different traffic classes require different amount of bandwidth to sustain for acceptable quality of experience (QoE) to the end users. Consequently, a class aware load balancing is designed to reserve traffic resources a priori when an impending handover is observed. The effect of class aware load balancing

ABSTRACT

and context aware handover over the QoS and QoE of multiple traffic classes has been analyzed using practical testbed results. To this end, the class aware load balancing and the context aware handover mechanism are further extended for vertical handover between WiFi and IEEE 802.16 or WiMAX networks, as the third contribution of this thesis. The performance of the proposed set of context aware vertical handover management schemes is analyzed using simulation results, and compared with other existing mechanisms.



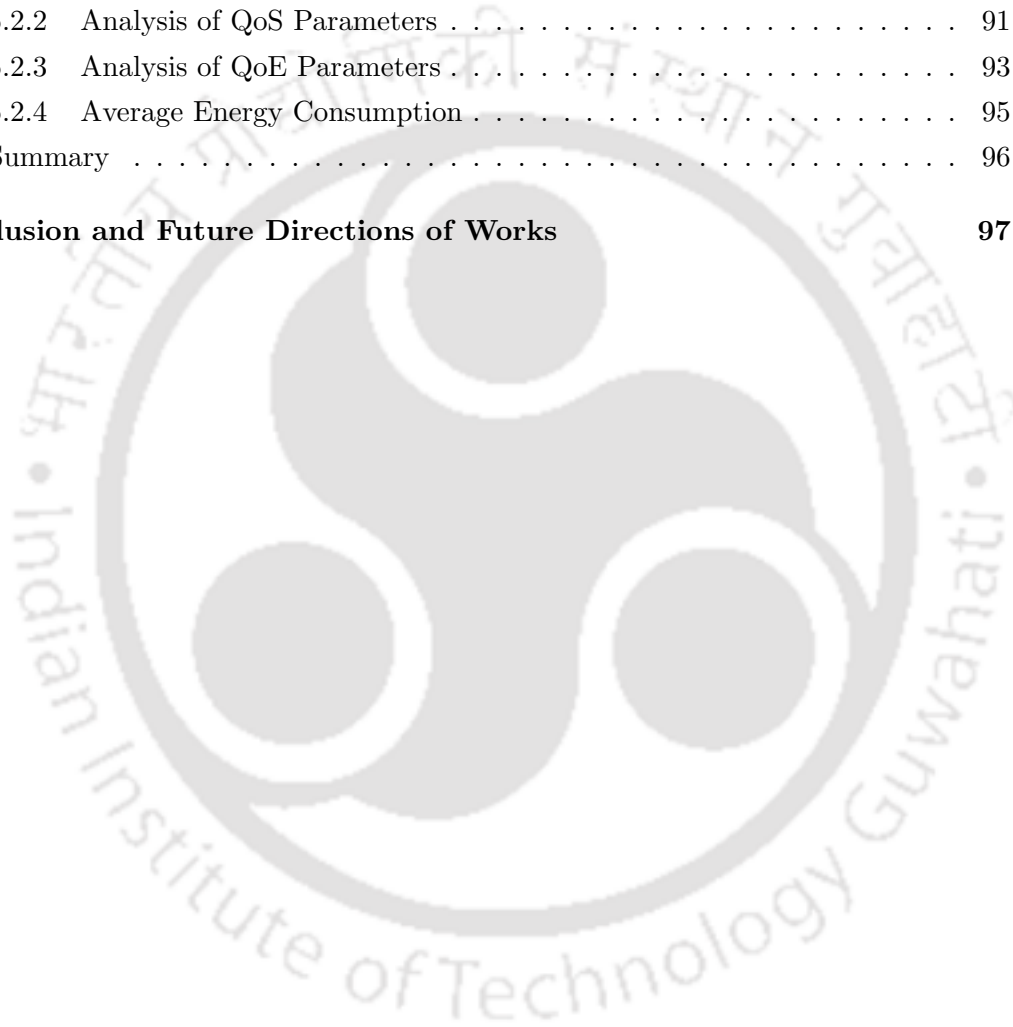
Contents

| | |
|--|-----------|
| List of Figures | xiv |
| List of Tables | xvi |
| Nomenclature | xviii |
| 1 Introduction | 1 |
| 1.1 Motivation of the Research | 3 |
| 1.2 Contribution of the Thesis | 6 |
| 1.3 Organization of the Thesis | 7 |
| 2 Background and Related Works | 9 |
| 2.1 Background: IEEE 802.11 WLAN Standardization | 9 |
| 2.2 IEEE 802.11e: QoS Associated Channel Access | 11 |
| 2.3 QoS Provisioning in IEEE 802.16 | 12 |
| 2.4 Inter-BSS Handoff in IEEE 802.11: A Brief Survey | 12 |
| 2.5 QoS Associated Handover Management in IEEE 802.11: The State-of-Art | 15 |
| 2.5.1 Call Admission Control over IEEE 802.11e | 16 |
| 2.5.2 Load Balancing in Inter-BSS Handover | 17 |
| 2.6 WiFi-WiMAX Integration and QoS Provisioning in the Literature | 17 |
| 2.7 Summary | 19 |
| 3 Context Aware Inter-BSS Handoff in IEEE 802.11 Networks: Efficient Resource Utilization and Performance Improvement | 21 |
| 3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off | 22 |
| 3.1.1 Modeling the Bandwidth Evolution within a BSS | 24 |
| 3.1.2 Numerical Results and Comparison with Simulation | 27 |

| | | |
|----------|--|-----------|
| 3.1.3 | Analyzing the Effect of Load-balancing over Inter-BSS Handoff through Numerical Results | 29 |
| 3.2 | Context-aware Inter-BSS Handoff Procedure | 33 |
| 3.2.1 | Interleaved Scanning | 34 |
| 3.2.2 | Detection of Performance Degradation | 36 |
| 3.2.3 | The Handoff Procedure | 37 |
| 3.3 | Performance Analysis | 39 |
| 3.3.1 | Testbed Analysis | 39 |
| 3.3.2 | Simulation Results and Comparison | 45 |
| 3.4 | Summary | 47 |
| 4 | Context Aware Handover Management: Sustaining QoS and QoE in a Public IEEE 802.11e Hotspot | 49 |
| 4.1 | Theoretical Foundation | 50 |
| 4.1.1 | Network Model | 51 |
| 4.1.2 | Analysis of Connection Blocking Probability, Connection Dropping Probability and Bandwidth Occupancy | 52 |
| 4.1.3 | Extension with a General Mobility Model | 54 |
| 4.1.4 | Numerical Results: Analysis of the Bandwidth Reservation Strategies | 56 |
| 4.2 | Context Aware Handover Management Sustaining QoS and QoE | 59 |
| 4.2.1 | Admission Control | 59 |
| 4.2.2 | Class Aware Load Balancing | 61 |
| 4.2.3 | Context Aware Handover | 62 |
| 4.2.4 | Implementation of the Context Aware Handover Management Module | 65 |
| 4.2.5 | Selection of Cutoff Bandwidth and Flexible Bandwidth | 66 |
| 4.2.6 | Discussion: MNs at the Hotspot Boundary | 66 |
| 4.3 | Experimental Evaluation and Comparison | 67 |
| 4.3.1 | Testbed Setup | 67 |
| 4.3.2 | Evaluation of QoS Parameters | 69 |
| 4.3.3 | Evaluation of QoE Parameters | 73 |
| 4.4 | Summary | 80 |
| 5 | Deciding Handover Points in a WiFi-WiMAX Heterogeneous Network Maintaining QoS and QoE | 81 |
| 5.1 | The Proposed Scheme: Load Balancing and Handover Policy | 82 |
| 5.1.1 | Bandwidth Reservation at AP and BS | 84 |

CONTENTS

| | | |
|----------|---|-----------|
| 5.1.2 | Admission Control | 85 |
| 5.1.3 | Class Aware Load Balancing and Context Aware Handover | 86 |
| 5.2 | Simulation Results, Performance Evaluation and Comparison | 89 |
| 5.2.1 | Network Scenario Setup | 89 |
| 5.2.2 | Analysis of QoS Parameters | 91 |
| 5.2.3 | Analysis of QoE Parameters | 93 |
| 5.2.4 | Average Energy Consumption | 95 |
| 5.3 | Summary | 96 |
| 6 | Conclusion and Future Directions of Works | 97 |





List of Figures

| | | |
|------|--|----|
| 2.1 | IEEE 802.11 Extended Service Set | 9 |
| 3.1 | BSS Traffic Model | 23 |
| 3.2 | A Birth-Death Process Model for Bandwidth Occupancy at BSS k | 25 |
| 3.3 | Network Scenario for Simulation and Numerical Analysis: Every circle represents an AP, and the dotted lines represent the overlapping in the communication coverage of two APs. All the APs are interconnected through distribution system | 27 |
| 3.4 | Connection Blocking Probability | 28 |
| 3.5 | Connection Dropping Probability | 29 |
| 3.6 | Connection Blocking Probability (CBP): Load-Balancing based Inter-BSS Handoff | 30 |
| 3.7 | Connection Blocking Probability (CBP): Standard Inter-BSS Handoff | 31 |
| 3.8 | Connection Dropping Probability (CDP): Load-Balancing based Inter-BSS Handoff | 31 |
| 3.9 | Connection Dropping Probability (CDP): Standard Inter-BSS Handoff | 32 |
| 3.10 | Interleave Scanning | 34 |
| 3.11 | Indoor Testbed Layout | 40 |
| 3.12 | AP Throughput (N_4) | 41 |
| 3.13 | Normalized MN Throughput | 41 |
| 3.14 | Packet Drop Rate | 42 |
| 3.15 | Fairness Index | 43 |
| 3.16 | Forwarding Delay | 44 |
| 3.17 | Average Flow Throughput | 44 |
| 3.18 | Average Forwarding Delay | 45 |
| 3.19 | Simulation Scenario | 46 |

LIST OF FIGURES

| | | |
|------|--|----|
| 3.20 | Average Flow Throughput | 46 |
| 3.21 | Average Forwarding Delay | 47 |
| 4.1 | Markov state diagram for bandwidth occupancy in a BSS, with two classes in eight channels | 52 |
| 4.2 | Connection Blocking Probability | 57 |
| 4.3 | Connection Dropping Probability | 58 |
| 4.4 | Normalized Bandwidth Occupancy | 58 |
| 4.5 | Bandwidth reservations for admission control management | 61 |
| 4.6 | Proposed management module in IEEE 802.11 protocol stack | 65 |
| 4.7 | IEEE 802.11 Indoor Wireless Testbed | 67 |
| 4.8 | Bandwidth Occupancy | 70 |
| 4.9 | Data Loss for QoS Traffics | 70 |
| 4.10 | Network Utilization | 71 |
| 4.11 | End-to-end one-way Forwarding Delay for Voice Traffics | 72 |
| 4.12 | Average Jitter for Voice Traffics | 72 |
| 4.13 | MOS for Voice Traffics | 74 |
| 4.14 | PSNR for Video Traffic | 74 |
| 4.15 | SSIM for Video Traffic | 75 |
| 4.16 | Throughput for AC_BK Traffics | 75 |
| 4.17 | Connection Losses (Block + Drop) | 76 |
| 4.18 | Video PSNR with respect to increasing AC_VO traffic | 77 |
| 4.19 | Background traffic throughput with respect to AC_VO traffic | 77 |
| 4.20 | Impact of Mobility over QoE Parameters: AC_VO MOS | 78 |
| 4.21 | Impact of Mobility over QoE Parameters: C_VI PSNR | 79 |
| 4.22 | Impact of Mobility over QoE Parameters: Background Traffic Throughput | 79 |
| 5.1 | Simulation Scenario | 90 |
| 5.2 | Average Throughput for QoS Traffics | 91 |
| 5.3 | Average Throughput for Best Effort Traffics | 92 |
| 5.4 | Average One-way Delay for Voice and Video Traffics | 93 |
| 5.5 | Average Jitter for Voice and Video Traffics | 93 |
| 5.6 | MOS for Voice Traffics | 94 |
| 5.7 | SSIM for Video Traffics | 94 |
| 5.8 | Average Per-User Energy Consumption | 95 |

List of Tables

| | | |
|-----|---|----|
| 2.1 | EDCA Access Classes | 11 |
| 2.2 | WiMAX: IEEE 802.16e QoS Classes | 12 |
| 4.1 | Parameters for numerical analysis | 56 |
| 4.2 | Testbed Parameters | 69 |
| 5.1 | Per User Traffic Demand | 91 |



Nomenclature

| | |
|------------------|--------------------------------------|
| AC | Access Categories |
| AIFS | Arbitration Inter-frame Spacing |
| AIFSN | Arbitration Inter-frame Space Number |
| ALM | Airtime Link Metric |
| BE | Best Effort |
| BEB | Binary Exponential Back-off |
| BG | Background |
| BSS | Basic Service Set |
| CA | Collision Avoidance |
| CDF | Cumulative Distribution Function |
| CSMA | Carrier Sense Multiple Access |
| CW | Contention Window |
| CW _{nd} | TCP Congestion Window |
| DCF | Distributed Coordination Function |
| EDCA | Enhanced Distributed Channel Access |
| EDR | Effective Data Rate |
| ESR | Estimated service rate |
| ETM | Expected Transmission Cost |
| ETX | Expected Transmission Count |
| FTP | File Transfer Protocol |
| FQA | Fair QoS Agent |
| HWMP | Hybrid Wireless Mesh Protocol |
| LKM | Loadable Kernel Module |
| LP | Linear Programming |
| LSP | Local Search Problem |
| MAC | Medium Access Control |

NOMENCLATURE

| | |
|--------|--|
| MAF | Mesh Access Fraction |
| MAFL | MAF Limit |
| MANET | Mobile Ad-Hoc Networks |
| MAP | Mesh Access Point |
| MARA | Metric Aware Rate Adaptation |
| MBSS | Mesh Basic Service Set |
| MBTT | Mesh Beacon Transmission Time |
| MCCA | Mesh Coordinated Channel Access |
| MCCAOP | MCCA Opportunities |
| MIMO | Multiple Input Multiple Output |
| MPM | Mesh Peer Management |
| OFDM | Orthogonal Frequency Division Multiplexing |
| PCF | Point Coordination Function |
| PPREQ | HWMP Proactive PREQ Message |
| PREP | HWMP Path Reply Message |
| PREQ | HWMP Path Request Message |
| QoS | Quality of Service |
| RTT | Round Trip Time |
| S-ALM | Schedule based Airtime Link Metric |
| SCV | Squared Coefficient of Variance |
| SNR | Signal to Noise Ratio |
| SOAR | Simple Opportunistic Adaptive Routing |
| STA | Mesh Station |
| TCP | Transmission Control Protocol |
| TDMA | Time Division Multiple Access |
| TFTP | Trivial File Transfer Protocol |
| UDP | User Datagram Protocol |
| VD | Video |
| VO | Voice |
| VoIP | Voice over IP |
| WiFi | Wireless Fidelity |
| WLAN | Wireless Local Area Network |
| WMN | Wireless Mesh Network |

Chapter 1

Introduction

IEEE 802.11 standardization [1–3] for wireless connectivity is gaining popularity day by day because of its low cost solutions and ease of deployment for providing ubiquitous end-user connectivity. In an IEEE 802.11 ‘Wireless Fidelity’ (WiFi) network, mobile users can connect to the Internet through wireless access points (APs) that form a backbone network through the wired distribution system. IEEE 802.11 WiFi technology is widely used to provide public wireless connectivity at airports, restaurants and other such public places, known as ‘WiFi HotSpots’ [4–6] as well as to support ubiquitous Internet connectivity through institute-wide wireless local area network (WLAN) architecture. Under such a WiFi HotSpot or WLAN coverage, users can enjoy the flexibility of migrating from one place to another under the wireless coverage area, while getting the advantage of seamless network connectivity.

Quality of Service (QoS) and Quality of Experience (QoE) are important issues that gain significant research interest now-a-days. For commercial and commodity network services, users generate traffic of different types that can be broadly classified into two groups - elastic traffic and inelastic traffic. Most of the data traffic from web browsers or file transfers uses Transmission Control Protocol (TCP) that incorporates the elastic traffic in the network. On the contrary, most of the real time traffic, such as voice and video are inelastic in nature. It is well known that, in the presence of elastic traffic, inelastic or real time traffic suffer to reserve required channel resources to ensure QoS and QoE [7, 8]. IEEE 802.11e [9] is the widely used standard for commercial deployment of WiFi HotSpots and institute-wide WLANs, that ensures QoS and QoE associated traffic delivery. The standard provides channel access specifications over the well established IEEE 802.11 media access protocols based on traffic class differentiation according to

1 Introduction

the QoS and QoE requirements. IEEE 802.11e supports Enhanced Distributed Channel Access (EDCA) that augments the basic IEEE 802.11 channel access protocol for QoS provisioning based on differentiated service architecture.

Though IEEE 802.11 based WLAN architecture has shown significant leap for public, commercial and commodity deployments, such as wireless HotSpots as well as institute-wide WLAN infrastructures, due to its cost-effective and ease-management advantages; the modern era of ‘wireless divide’ is gradually witnessing deployments of more advanced wireless technologies, such as IEEE 802.16 or Worldwide Interoperability for Microwave Access (WiMAX) [10]. WiMAX provides longer coverage area compared to WiFi technology through advanced modulation and coding schemes for signal transmissions. WiFi-WiMAX integration is an important and challenging research topic for next generation wireless Internet architecture and ‘Internet of Things’ (IoT) designs that attract significant attentions among the researchers [11–14]. The recent developments in WiFi-WiMAX integration technologies drive the developers to launch commercially available products, such as Intel WiMAX/WiFi Link 5350 and Intel WiMAX/WiFi Link 5150 [15]. However, WiFi provides wireless connectivity within short communication range, but with minimal infrastructure support, whereas WiMAX requires specific infrastructure deployments through it can be used for long distance point-to-point communications. Such a complex and integrated network is generally termed as heterogeneous network or *HetNet* that provides ample ability to the users to select a technology based on the underlying application and QoS/QoE demands.

The users in a wireless environments have flexibility to be mobile in nature, and therefore handover from the present point of attachment to another point of attachment is given significant research attentions from the beginning. In a wireless environment, handover is generally classified into two groups - *horizontal handover* [16], where both the present point of attachment and the next point of attachment use similar technology (both are either IEEE 802.11 or IEEE 802.16), and *vertical handover* [17], where the present and the next point of attachments are of different technologies (one is IEEE 802.11 while another is IEEE 802.16). The architecture and protocols for both horizontal and vertical handover are well studied in the existing literature, such as [18–26] and the references therein. These works have designed protocol elements and communication architectures, such as association, re-association, authentication and accounting, for efficient horizontal and vertical handover between either similar or different technologies, respectively. However, the existing works lacks to answer two important questions - first, how to balance traffic load across the network during horizontal and vertical handover,

and second; when a node should initiate a horizontal or a vertical handover to balance the traffic load while maintaining QoS and QoE demands of the users. This thesis deals with the user or MN distribution across a wireless network, either a simple WiFi HotSpot or a WiFi-WiMAX HetNet environment, to balance traffic load across the network while maintaining QoS and QoE of the end users. A balanced user distribution across the network significantly improves network capacity, that allows more users to be associated with the network, while the users' satisfaction towards Internet service providers is also ensured.

1.1 Motivation of the Research

According to the IEEE 802.11 standard [1], every AP along with the mobile nodes (MN) associated with it forms a '*Basic Service Set*' (BSS). In a community wireless network, such as wireless 'Hot-Spots' [27] or WLANs, multiple APs coexist with overlapping coverage area to provide wireless network connectivity in a broader space. In such scenarios, sufficient overlapping among the APs is assured so that the MNs can select the best AP in its vicinity [28]. An IEEE 802.11 WLAN with interconnected APs forms an '*Extended Service Set*' (ESS). In conventional approaches, APs from an ESS are selected based on either the received signal strength indication (RSSI) or the signal-to-interference and noise ratio (SINR) [29] measurements. A MN periodically measures the RSSI (or the SINR), and associates with the AP that provides maximum signal strength value based on the RSSI (or the SINR) measurement.

A number of works in the literature, such as [21, 30] and the references therein, have analyzed user behavior in a public wireless HotSpot or institute-wide WLAN. They have shown that in general, more than the required number of APs are deployed to ensure that every user can have the choice to select an AP that provides the best signal quality in terms of RSSI or SINR. However, RSSI or SINR is not the only factor that ensures good network connectivity and satisfactory application performance. In an ESS, multiple MNs can associate with a single AP, and being a shared and open wireless environment, they contend with each other to gain access to a common wireless channel. Therefore, traffic load (or number of users associated with an AP) is an important factor that decides the end-user performance under a wireless network.

The work by Balachandran *et al* [30] has shown that in a public wireless network, users have the tendency to cluster and congregate at some particular places. For example, in an airport, users normally get assembled near the boarding gates in operation. Consequently,

1.1 Motivation of the Research

the APs near that region get overloaded, as the SINR or RSSI, in general, has a direct proportional relationship with the distance between the transmitter and the receiver. Therefore, most of the users tend to connect with a single AP which is more closer to the user cluster, though the users are under coverage of multiple APs that can provide acceptable signal quality, in terms of SINR or RSSI. This results in significant traffic imbalance among the APs, and in turn degrade the application performance as well as capacity of the network.

The IEEE 802.11s support horizontal handover through inter-BSS migration where a MN can disassociate from an AP and re-associate to another AP. As discussed earlier, communication coverage areas of APs in wireless HotSpots and WLANs overlap sufficiently, and therefore a MN remains within the coverage of more than one APs. The capacity degradation in a wireless HotSpot or WLAN can be avoided by a proper distribution of the MNs among the overlapping (or neighboring) APs so that the APs do not get overloaded whereas the MN gets sufficient resources. In this context, if a MN experience performance degradation due to AP overloading, it can migrate to another AP in the vicinity that provides sufficient signal strength to sustain a communication. However, for QoS and QoE assurance, such a migration (or horizontal inter-BSS handover) requires a priory knowledge of resource demands and pre-allocation of resources based on the demands (or service level agreements) for the corresponding real-time traffic.

While load-balancing and pre-allocation of traffic resources can improve the capacity as well as service quality in a IEEE 802.11 network, they alone are not sufficient for QoS assurance in a HetNet environment. One of the important design aspects for WiFi-WiMAX integration is: when should a user perform handover from the WiFi interface to the WiMAX interface and vice versa? A lot of researches has been already carried out to answer how to handover between WiFi and WiMAX [22–26], however most of these works remain silent on when to perform the handover. IEEE 802.21 or media independent handover (MIH) framework [31] standardizes the vertical handover (handover between two different technologies) procedure, however it discusses about ‘how to perform the handover’, rather than ‘when to perform the handover’. In a typical WiFi-WiMAX HetNet environment, the WiMAX connectivity symbolizes an umbrella like coverage under which the WiFi access points (APs) or basic service sets (BSSs) formulate small WiFi connectivity areas. Because of its long communication ranges, it can be safely assumed that in such a HetNet environment, users have always access to the WiMAX connectivity, however access to WiFi is only available at few areas like the HotSpots.

However, users prefer WiFi over WiMAX because of its low-cost availability¹, and less power consumption², and the WiMAX is typically used as the back-haul to connect several WiFi HotSpots together, as well as to provide direct connectivity to the end-users. A user migrates from WiFi to WiMAX either because there is no WiFi connectivity, or the WiFi fails to provide the desired QoS requirement.

Mobility is not the only issue that triggers a vertical handover between two technologies in a HetNet environment. As discussed earlier, a quintessential problem of WiFi hotspots is unbalanced traffic load distribution among the APs or BSSs [21]. The traffic load at WiFi access points are highly dynamic and varies with time of the day and geographic locations [33]. This causes overload in some APs where as other APs in the network remains underutilized. QoE of users connected through the overloaded APs deteriorates [34] as the traffic load increases in the APs. Real time applications have strict bounds on network parameters such as end-to-end delay, jitter and packet loss rate. Violation of these bounds may cause disruptions in application performance. Users running voice over IP applications on their terminal may experience echo, noise or empty gaps in the communications if end-to-end delay exceeds 150 ms or packet loss exceeds 2%. In this scenario, vertical handover can be exploited where a MN migrate to the WiMAX network, when sufficient bandwidth is not available at the WiFi network to provide QoS guarantee.

This thesis explores the traffic and MN distribution properties of a WiFi network as well as of a HetNet environment, and devise effective handover management schemes for balancing traffic load across the network. The proposed management schemes in this thesis work as a sub-layer at the Media Access Control (MAC) of network protocol stack, and triggers the MNs to perform either a horizontal or a vertical handover based on the application performance and QoS/QoE assurance. The proposed schemes are a set of cooperative management schemes between APs and MNs for the horizontal handover, and among WiMAX base station (BS), WiFi APs and MNs in case of the vertical handover. The major contribution of this thesis in this context has been summarized in the next section.

¹Typically, WiMAX uses licensed bandwidth, whereas WiFi operates over unlicensed 2.4GHz channel. Therefore cost per bit is less in WiFi, compared to WiMAX.

²WiFi interface and AP consumes significantly less power compared to WiMAX interface and base stations [32]. This is because WiFi uses short-range communications and therefore consumes less power for transmit and receive.

1.2 Contribution of the Thesis

As discussed earlier, IEEE 802.11 has enabled users to access multimedia and other real time applications using wireless local area networks (WLAN). In IEEE 802.11 WLAN technology, associations between a mobile station and an access point (AP) is controlled by the mobile station, allowing the station to select an AP with the strongest signal in terms of either RSSI or SINR. In real time scenarios, the traffic patterns of mobile users are dynamic in nature. This leads to a situation where the traffic loads on the APs are unevenly distributed in the WLAN. Such imbalance in traffic load causes severe degradation in performance of the applications running on the mobile stations associated with the overloaded APs. In the first contribution of the thesis, we propose a scheme which dynamically improves the performance of the overloaded APs by handing off some of its associated stations to nearby APs. This handoff decision is taken by an AP in assistance with the mobile stations. The effectiveness of the load distribution through dynamic hand-over in a WLAN is analyzed through theoretical analysis. Simulation results show the overall improvements in terms of delay, throughput and number of stations that an AP can support. The performance improvement in the proposed scheme is also justified through the results obtained from a IEEE 802.11 WLAN testbed.

The second contribution of the thesis extends the scenario of context-aware handover for QoS and QoE associated MNs, where the MNs generate real time traffic that requires specific bandwidth reservation for good application performance. In this context, we propose a set of bandwidth management policies to achieve QoS and QoE handover management across an IEEE 802.11 network. The effectiveness of these policies are first analyzed theoretically, where the traffic requirements for a specific service class is determined based on a Markov chain analysis. According to the theoretical foundation, a context aware handover management scheme for proper load distribution in a public IEEE 802.11 network, supporting the class-aware bandwidth management policies, is designed. The performance of the proposed scheme is evaluated using an IEEE 802.11g+e wireless LAN testbed, and compared with other schemes proposed in the literature.

Finally, the thesis augments the proposed class-aware load balancing and context aware handover scheme for a HetNet environment where multiple WiFi HotSpots create coverage holes under a common WiMAX coverage area. The existing literatures have well studied the interoperability between these two technologies, and seamless vertical handover schemes are designed to support WiFi-WiMAX integration in a HetNet environment. However, effective utilization of these two technologies from the end users' point-of-

view is another important research area that has been explored in this contribution. This contribution of the thesis answers the following question: when a user should migrate between WiFi-WiMAX technologies in a HetNet environment to assure QoS and QoE? This thesis proposes a vertical handover decision mechanism in a WiFi-WiMAX integrated HetNet environment, that supports the ‘Quality of Service’ (QoS) and ‘Quality of Experience’ (QoE) requirements of the end users. As the QoS management policies are different for WiFi and WiMAX networks, in this contribution of the thesis, we have first designed an one-to-one mapping between WiFi service classes with WiMAX service classes. Based on this mapping, a traffic pre-reservation scheme is designed to support the class-aware load balancing and context-aware handover management. The effectiveness of the proposed scheme is analyzed using simulation results.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows:

Chapter 2 gives a brief background of the IEEE 802.11 and IEEE 802.16 network architecture, inter-BSS handover procedure as well as QoS provisioning during MAC layer channel access. This chapter also gives a thorough survey of the existing literatures that discuss about handover management in vertical and horizontal scenarios, as well as load balancing, admission control and QoS provisioning.

Chapter 3 discusses about the first contribution of the thesis where a context-aware handover management scheme is designed to balance traffic load across an IEEE 802.11 network. This chapter gives a theoretical foundation of the proposed scheme, along with the detailed description of the handover management procedure. Finally the chapter reports the results from simulation as well as a practical wireless testbed to analyze the performance of the proposed scheme.

Chapter 4 extends the first contribution of the thesis for QoS and QoE associated handover management in an IEEE 802.11 network. The effect of different service classes over bandwidth reservation is first analyzed using a Markov chain modeling. The theoretical foundation is then used to design a class-aware load balancing and context-aware handover management scheme. The performance of the proposed scheme is analyzed using simulation results.

Chapter 5 proposes a class-aware vertical handover management scheme in a HetNet environment. This chapter first maps the traffic classes of a WiFi network with that of the WiMAX network. Based on this mapping, the class-aware load balancing and the

1.3 Organization of the Thesis

context-aware handover schemes are designed to decide when a MN should migrate from a WiFi network to a WiMAX network and vice-versa. The performance of this scheme is analyzed using simulation results.

Finally, Chapter 6 concludes the thesis with the future works in this direction.



Chapter 2

Background and Related Works

This chapter gives the background of the IEEE 802.11 WLAN and IEEE 802.16 WiMAX network architecture with an emphasis to the handover management and QoS provisioning during MAC layer channel access. After that, the state-of-the-art works on vertical and horizontal handover management is discussed in detail. For this purpose, the discussion of state-of-the-art works has been classified into three subgroups - the works on IEEE 802.11 horizontal handover with admission control and load-balancing, a literature survey on QoS provisioning during horizontal IEEE 802.11 inter-BSS handover, and finally the works on QoS provisioning in WiFi-WiMAX vertical handover.

2.1 Background: IEEE 802.11 WLAN Standardization

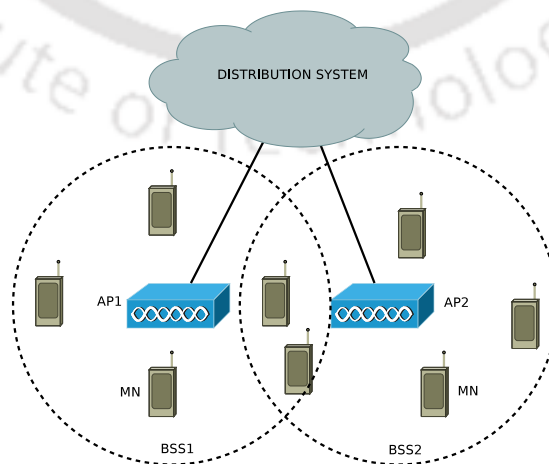


Figure 2.1: IEEE 802.11 Extended Service Set

2.1 Background: IEEE 802.11 WLAN Standardization

An IEEE 802.11 network has mainly two types of stations, namely, AP and MN. The basic service set (BSS) is a set of IEEE 802.11 nodes that can communicate with each other directly. There are two types of BSSs, namely, independent BSS and infrastructure BSS. An independent BSS (IBSS) is an ad-hoc network that contains no APs, and as such they can not connect to any other BSS. In an infrastructure BSS (iBSS) there is one AP and some MNs. All communications to and from the MNs must go through the AP. MNs in an iBSS can communicate with other MNs in its own BSS and other BSSs as well, through the distribution system. Two or more BSSs may be interconnected through connecting the APs of each BSS by a distribution system. An ESS is a set of one or more interconnected BSSs. An ESS increases both the coverage area and the bandwidth available within the area covered. The distribution system may be connected to other networks so that the stations in the ESS may communicate with other networks. Such an example setup is shown in Figure 2.1. It can be noted that, to avoid disruption of communication to MNs, different APs should have overlapped coverage area such that communication disruption can be avoided when a MN moves from one BSS to another.

There can be up to fourteen different channels in the 2.4GHz Industrial, Scientific and Medical (ISM) band in which an IEEE 802.11 WLAN operates. Due to the restrictions imposed, not all channels are available in all countries. Channels one to eleven are available in most of the countries. Each channel is of 22MHz bandwidth but spaced only 5MHz apart. As such there can be only three non overlapping channels among the eleven overlapping channels. Although the APs may operate on overlapping channels, it is not desirable to use the overlapping channels in adjacent APs with overlapping coverage area due to the interference and the resulting performance loss. To increase the available bandwidth in an ESS, the APs in an ESS should operate in different non-overlapping channels.

A MN should associate to the AP of a BSS to communicate. A MN associates to an AP by sending an association request message. The AP, if willing to allow association, replies with an association response message and the MN gets associated to the AP. At the time of associating an MN, the AP considers only its own status to decide whether the MN will be allowed association or not. Further association is based on signal strength only, which is the root cause behind load-imbalance problem in IEEE 802.11 ESS.

Table 2.1: EDCA Access Classes

| Access class | CW_{min} | CW_{max} | AIFS | TXOP |
|--------------|------------|------------|------|-------|
| AC_VO | 7 | 15 | 2 | 3.264 |
| AC_VI | 15 | 31 | 2 | 6.016 |
| AC_BK | 31 | 1023 | 7 | 0 |
| AC_BE | 31 | 1023 | 3 | 0 |

2.2 IEEE 802.11e: QoS Associated Channel Access

IEEE 802.11e EDCA [9] is an extension of the standard distributed coordination function (DCF) with four classes of services, also known as access classes (AC), - best effort (AC_BE), background (AC_BK), video (AC_VI) and voice (AC_VO). In DCF, channel access is controlled by four parameters - minimum contention window (CW) size (CW_{min}), maximum CW size (CW_{max}) and two inter-frame spacing (IFS) parameters, namely, short inter-frame spacing (SIFS) and DCF inter-frame spacing (DIFS). Whenever a station wants to transmit, it senses the channel for any ongoing transmissions, and if the medium is found to be idle for DIFS duration, the station starts transmitting data frames. If the channel is found busy, it performs back-off for a period. The back-off period is determined by a randomly generated number between 0 and CW_{min} . If the number generated is N , then it decrements N by one in each non busy slot, and whenever N reduces to zero, it tries to transmit the frame again after sensing the medium to be idle for DIFS slots. Every time it fails to transmit, it doubles up the CW , till the CW value reaches to CW_{max} . A station needs to wait for SIFS duration to transmit an unicast acknowledgement or response frame. The value of the SIFS duration is less than the DIFS duration. After the successful transmission of a data frame, the CW value is reset to its original value.

In EDCA, there are separate CW_{min} and CW_{max} for each access class. There is also an IFS, called the arbitration inter-frame spacing (AIFS) for each access class. In EDCA, stations use AIFS instead of DIFS to acquire channel. These different parameters provide different access probabilities to the service classes according to their priority. For example, shorter AIFS with smaller CW_{min} make the probability of acquiring the channel high. There is a transmission opportunity (TXOP) limit associated with each access class. If a channel is acquired by a station, it may transmit upto TXOP amount of time. This allows for transmitting a burst of traffic. The standard defined default parameter values for the different traffic classes are given in Table 2.1.

2.4 Inter-BSS Handoff in IEEE 802.11: A Brief Survey

Table 2.2: WiMAX: IEEE 802.16e QoS Classes

| Class | Description |
|-------|------------------------------------|
| UGS | Unsolicited grant service |
| ertPS | Extended Real time polling service |
| nrtPS | Non real time polling service |
| rtPS | Real time polling service |
| BE | Best effort |

2.3 QoS Provisioning in IEEE 802.16

IEEE 802.16 or WiMAX [10] provides a point-to-point wireless connection between the base station (BS) and the MNs. The standard uses scalable orthogonal frequency division multiple access (OFDMA) to carry data, supporting channel bandwidths of between 1.25 MHz and 20 MHz, with up to 2048 sub-carriers. The IEEE 802.16 MAC describes a number of Convergence Sub-layers which describe how wireline technologies such as Ethernet, Asynchronous Transfer Mode (ATM) and Internet Protocol (IP) are encapsulated on the air interface, as well as how data is classified. A key feature of IEEE 802.16 is that it is a connection-oriented technology. The MNs, also called the subscriber stations (SS) cannot transmit data until it has been allocated a channel by the BS. This allows WiMAX to provide strong support for quality of service (QoS) through integrated service architecture.

WiMAX introduces five traffic classes through IEEE 802.16e standard, as shown in Table 2.2. Unsolicited grant service (UGS) is intended for E1/T1 transport that provides fixed bandwidth for fixed packet sizes. Real time polling service (rtPS) is used for variable bit rate real time traffic classes. On the contrary, extended real time polling service (ertPS) is intended for constant bit rate traffics generated by real time traffic classes. Non-real time polling service (nrtPS) is used for non real time traffics that require minimum data rate assurance, and best effort (BE) traffic is for traffics that do not require any QoS guarantee.

The following sections describe the state-of-the art works in the context of vertical and horizontal handover with an emphasis on QoS provisioning.

2.4 Inter-BSS Handoff in IEEE 802.11: A Brief Survey

Most of the existing works for inter-BSS handoff in IEEE 802.11 WLAN networks target to improve the overall network performance in terms of handoff delay, throughput and

QoS. Several works in the literature target to perform load balancing operations during the association between a MN and an AP. Sawma *et al.* [35] have presented a scheme where the loads in the APs are adjusted dynamically, whenever a new MN tries to associate to an AP and the AP is already overloaded. In their scheme, several MNs are directed to change the APs to balance the traffic load. However, load-balancing is done only when a new MN is associated to the WLAN, but the scheme does not take care of the scenario when traffic load changes due to the increase in the traffic demand from the already associated MNs. Moreover, the decision to switch APs is made solely by the AP, based on the traffic load conditions of that APs only. In the scheme presented in [36], a periodic probing is used at each MN to determine the traffic conditions at the associated APs. If a performance degradation is observed due to the sudden increase in the traffic load, the MN decides to switch to an alternate AP in its vicinity. In their scheme, the probing interval is adjusted to 300 ms, that may disrupt the communication for real time traffic. Further, frequent probing significantly increases signaling overhead in the network.

In [37], the authors have proposed a load-balancing based handoff scheme depending on the traffic load and the QoS criteria at the APs. In the paper, the authors have proposed an integrated load balancing scheme incorporating the QoS-based fast handoff to support seamless handoff by eliminating both detection and scanning phases from the 802.11 WLAN handoff process. They have also designed a soft admission control mechanism to protect QoS of the existing connections when resources are low. For this purpose, they have designed an opportunistic handoff scheme, where the QoS parameters are measured and broadcast periodically by the APs, and the MNs select the best AP opportunistically by comparing their QoS requirement with the information obtained from the broadcast beacons. The disadvantage of this scheme is that it requires large signaling overhead due to beacon broadcast, and may result frequent AP switching at the MNs, which may degrade application performance severely.

In [38], the authors have considered RSSI and frame error rate (FER) for load balancing during inter-BSS handoff in IEEE 802.11 WLAN. They have designed a probabilistic metric based on the variation in RSSI and FER. The MNs take the handoff decision depending on the metric value. As discussed earlier, RSSI can not be a good indicator of load measurement, and therefore the dynamic load-balancing algorithm proposed in [38] results in poor performance when the variation in the traffic load is high [39].

Lee *et. al* [40] have proposed a static channel assignment and AP placement algorithms for load-balancing in an IEEE 802.11 ESS. They have considered the existing

2.4 Inter-BSS Handoff in IEEE 802.11: A Brief Survey

traffic patterns and the MN characteristics to design an optimal deployment of APs in a closed area. Further a static channel assignment algorithm is proposed so that the APs which overlap in the communication coverage area, use separate channel to avoid interference. Their scheme has not considered dynamic traffic load variation, and therefore has limited applications in terms of scalability. Several others load-balancing algorithms are proposed in literature, such as [18–21], where the load-balancing decision is taken centrally at the distribution system, and the MNs are directed to move all at once and re-associate with the APs with low traffic load. These schemes do not make any control over the re-association of the MNs, and therefore may further trigger load-imbalance problem if all the MNs move in the same direction.

In [41], the authors have designed an AP re-association mechanism to improve the performance for voice traffics in the network. The quality of voice applications greatly depends on the call blocking probability which is decided by all the potential voice stations. For this reason, they have defined a QoS-aware information element that is advertised from APs. On this basis, voice stations are able to select the AP which is associating with less number of voice stations so as to achieve a lower call blocking probability. This work is specific to the voice traffics, and does not consider general traffic loads of the network. Bejerano *et al.* [19] have considered the concept of cell breathing technique, used in cellular mobiles for load balancing. Their scheme requires the ability of dynamically changing the transmission power of the AP beacon messages. They have developed a set of algorithms that find the optimal beacon power settings, which minimize the load of the most congested APs, in polynomial time. They have also considered the problem of network-wide min-max load balancing. Another load-balancing algorithm based on transmission power control has been proposed in [42], where the APs reduce their transmission power to disassociate some of the MNs, when their traffic load increases beyond a threshold. The authors have also designed a connection management scheme for the stations that are expected to be disconnected after the power control of APs. However, dynamic power adjustment of beacons may introduce severe problem of uncontrolled interference at the network.

From the discussion of the state-of-the-art works on the load-imbalance problem in IEEE 802.11 inter-BSS handoff, it can be observed that there exist several design limitations which should be further addressed. The limitations of the existing schemes can be summarized as follows,

- Most of the schemes for the dynamic load balancing, like [35–38] devise mechanisms to find out optimal load distribution when an AP gets overloaded. The re-association decision is taken either by the APs or the MNs, where the AP broadcast current load

2.5 QoS Associated Handover Management in IEEE 802.11: The State-of-Art

information, and the MN select the best AP depending on the information. This perfect load balancing imposes extra signaling overhead due to the coordination among the APs and the MNs. Perfect load balancing is not always desirable because of this high signaling overhead. At the same time, these schemes consider only the number of MNs in association with an AP for decision making, and do not consider their individual traffic load, or the variation in actual traffic demand.

- Further, the existing schemes do not consider multi-channel effects during information broadcast. Different APs may use different channel for information broadcast, and therefore a suitable mechanism need to be designed at the MNs for efficient scanning at different channels to obtain the beacon information.
- The static AP placement and the channel assignment schemes, like [40] lack the scalability of the design, and does not consider on-demand traffic variations.
- The centralized schemes, like [18–21], also lack the scalability and the dynamicity of the wireless network.
- Load balancing based on dynamic transmission power control [19, 42] may result in uncontrolled interference in the network, which may severely degrade application performance.

2.5 QoS Associated Handover Management in IEEE 802.11: The State-of-Art

Although a number of works have been published in the literature over IEEE 802.11e channel access and its effectiveness for QoS provisioning, such as [43] and the references therein, very few works discuss about the impact of handover on the performance of IEEE 802.11e. Conversely, QoS based handover has been well studied in the literature, however most of the works, like [44] and the references therein, have considered only the vertical handover scenario. During vertical handover between two different wireless technologies, these schemes propose effective mechanism for bandwidth reservation for the higher priority traffic classes. Nevertheless as mentioned, IEEE 802.11e based public wireless hotspots suffer from performance degradation because of two main reasons - first, the load imbalance problem, and second, the uncontrolled traffic admissions during inter-BSS handover, which should be addressed simultaneously. To the best of our knowledge, no prior works address these two problems simultaneously in the context of IEEE 802.11e

2.5 QoS Associated Handover Management in IEEE 802.11: The State-of-Art

based wireless LAN. This section gives a brief overview of the works that have been proposed in the literature to address the problem from different viewpoints. For ease of presentations, the existing works in the literature has been classified into two groups - first, the works that propose admission control mechanism over IEEE 802.11e, and second, the works that discuss about load balancing during inter-BSS handover in a IEEE 802.11 wireless LAN.

2.5.1 Call Admission Control over IEEE 802.11e

As discussed earlier, through IEEE 802.11e supports differentiated QoS architecture for four classes of services, it does not provide any call admission control mechanism. Therefore in the standard, an AP does not have any control over the number of MNs associating with it, which may results in poor QoE even with the differentiated QoS architecture. A number of works have been proposed in the literature to design call admission control mechanism over the IEEE 802.11e EDCA protocol. In [45] and [46], the authors give a detailed survey of the works proposed in the literature for designing efficient admission control algorithms over IEEE 802.11e protocols. The admission control protocols proposed in the literature are broadly classified into two classes, discussed next.

A set of works [47,48] have proposed admission control mechanism over IEEE 802.11e EDCA through the measurement of network utilization characteristics, and reserving bandwidth for different classes of services according to the maximum network utilization. The connections are dropped whenever the network utilization gets reduced due to large number of lower class traffics. These works have two fundamental shortcomings,

1. the schemes are centralized in nature, as they require the measurement of global network utilization.
2. the admission control is done only after allowing connections and then measuring network utilization. This affects the performance of existing connections, as no differentiation is made between connections while selecting which one to drop due to poor network utilization.

The second set of works [49,50] propose model based admission control mechanism, where network connections and their utilizations are captured using mathematical models to design effective admission control rules. Though these schemes are distributed in nature, they have serious issues in scalability as the decision models show poor accuracy with large number of connections in the network.

2.5.2 Load Balancing in Inter-BSS Handover

A number of works have been proposed in the literature for load balancing in IEEE 802.11 networks. In [51] the authors have proposed a simple load balancing algorithm among overlapping wireless LAN cells, where every AP periodically broadcast its load information. If an AP gets overloaded, it forces some of the MNs to handover to the neighboring APs. Though this scheme balances load among the overlapping APs, the handover decision is solely on APs, that may degrade QoS and QoE at the MNs. Another load balancing algorithm has been proposed in [52], where the authors have explored the radio resource information obtained in IEEE 802.11k. Based on the information, some MNs disassociate from the currently associated AP, and switch over to another AP for better resource management. In [18], Bejerano *et al* have proposed a load balancing and association control mechanism to ensure fairness among the MNs. They provide a rigorous formulation of the association control problem that considers bandwidth constraints of both the wireless and backbone links. Their formulation indicates the strong correlation between fairness and load balancing, based on which they designed a load balancing technique for obtaining near optimal max-min fair bandwidth allocation. A group of works in the literature [19,53,54] proposed different techniques for load balancing in IEEE 802.11 wireless LAN networks. The cell breathing technique [19] controls the wireless LAN cell size (or the AP coverage area) to control number of MNs associated with an AP. However, this scheme is also AP controlled that may affect QoS and QoE of the MNs.

The above discussions show that the existing works have two major shortcomings that prevent the MNs to achieve desired QoS and QoE at an IEEE 802.11e wireless LAN. First, the admission control schemes do not consider the inter-BSS handover and load balancing which is essential in a hotspot scenario with significant numbers of overlapping wireless LAN cells. Second, most of the load balancing schemes proposed in the literature are network or AP controlled, that affects the QoS and QoE of the MNs. This shows the requirement of designing a new handover management scheme for IEEE 802.11e wireless LAN, while ensuring QoS and QoE to the end users.

2.6 WiFi-WiMAX Integration and QoS Provisioning in the Literature

As mentioned in Chapter 1, a number of works, such as [11–14] and the references therein, are published to integrate WiFi and WiMAX networks in a heterogeneous environment and

2.6 WiFi-WiMAX Integration and QoS Provisioning in the Literature

analyze the performance gain of such integration. The results and analysis of these works unveil eloquent promises to design next generation wireless network architecture. Though these works solve the lower layer technical issues, like how the handover is performed as the two technologies have different packet structures and lower layer modulation and coding schemes, they remain silent on users' expectations on performance issues for real time traffics, broadly the QoS and QoE. There are some, but still limited proposals published in the literature in order to improve the end users' performance in a WiFi-WiMAX heterogeneous network. In [26], the authors have proposed a scheme where MNs perform handover whenever the QoS of the underlying applications degrade. Similarly in [55] and [56], the authors have designed call admission control mechanisms during vertical handover in heterogeneous network environment. However, they do not consider the different traffic load scenarios and their sensitivity over the two different technologies.

In [25], the authors have proposed two different criteria for handover between WiFi and WiMAX. In the first case, they have used RSSI as the handover criteria, and in the second case, they have proposed to use the ratio of load between different networks as the handover criteria. However, they remain quiet about how to measure traffic loads at different networks, and use that information at the end users, as simultaneously measuring traffic loads at two different networks is a challenging issue. Several other schemes have been proposed in the literature, such as [57–59] and the references therein, to design QoS/QoE assisted vertical handover framework over the MIH architecture. However, these schemes are designed from the end users' perspective only, where an user performs a handover when it senses degradation in the QoS/QoE. Where this is one design aspect, the problem should also be looked from the network perspective, where the network should adjust itself to allow more number of users to associate, while maintaining the desired QoS/QoE profile.

In [24], Roy *et al.* have used packet loss and delay as criteria for performing handover from one network to another. Nevertheless, their results show 9 ms maximum delay which is not possible in WiMAX without altering scheduling parameters. In another work [23], Prasath *et al.* have suggested an integration model for WiFi and WiMAX network as well as a BS-assisted scheduling algorithm to provide QoS guarantee. In order to integrate WiFi and WiMAX at MAC layers, they have provided a mapping between access categories in WiFi and service classes in WiMAX. A scheduling algorithm is proposed in their scheme for providing QoS guarantee for real-time traffic in terms of delay, and non real-time traffic in terms of packet loss. In [60], the authors proposed a decision making process for mobile devices to select one network from among several available options in the

surroundings. Mobile node gathers the parameters, that reflects the network performance as well as its current location. This information is then used for taking handover decision. The drawback of this scheme is that it does not specify details of which parameters are measured, and how the parameters are collected at the end users.

In a recent work [44], Ma *et al.* have proposed a tight coupling based WiFi-WiMAX integrated HetNet architecture, and designed a QoS assisted handover procedure among the two technologies. They have designed an estimation mechanism, both at the WiFi and the WiMAX network, to find out available bandwidth and projected delay, and have used that information to instruct the users to switch between the two technologies. However, their architecture gives equal priority to the WiFi and WiMAX networks, which is not effective in terms of cost-per-bit and average power consumption, as discussed earlier. Further, they have not considered overlapping BSSs or WiFi hotspots architecture under the WiMAX coverage, which is more common in today's HetNet framework. As a consequence, their scheme does not address load balancing among different APs, as well as among different access technologies, which is essential for a complex HetNet as addressed in this paper.

2.7 Summary

From the discussions of the state-of-the-art works on vertical and horizontal handover as given in this chapter, we can conclude that though the existing literatures give a good understanding of the handover protocols, they remain silent about how to effectively use the handover procedure to support a robust load-balancing association scheme, so that traffic imbalance in a wireless network can be avoided. Further, the QoS and QoE assurance during vertical and horizontal handover is not clearly designed in the existing works. The next chapter provides the design and protocol elements of a context-aware handover scheme, where the inter-BSS handover is used effectively in an IEEE 802.11 network to avoid traffic load imbalance across the APs.



Chapter 3

Context Aware Inter-BSS Handoff in IEEE 802.11 Networks: Efficient Resource Utilization and Performance Improvement

As discussed in Chapter 2, several works in the existing literature [30, 33, 61–64] have shown that the traffic load in an ESS is often unevenly distributed across the network. Balachandran *et. al* [21, 30] have theoretically modeled the user behavior in a WLAN network. They have shown that MNs in a public WLAN network tend to concentrate in some specific area. For example, in an airport hot-spot, users get concentrated near the boarding gates. As a result traffic gets overloaded in the APs which are closer to the boarding gates, and provide maximum signal strength to the users. Though other APs may be available in that region that can provide similar performance, yet due to the signal strength based association mechanism, MNs do not associate with these APs. Thus there is a load variation among the APs that may affect overall network performance. Zhai *et al.* [34] have shown that WLAN attains maximum throughput and minimum delay when the nodes are unsaturated (attains traffic load less than the maximum capacity).

The load variation in an ESS of a WLAN network results in severe performance degradation due to the local congestion at some APs. As a result, traffic flows that pass through the overloaded APs get affected. This in turn also degrades the overall network throughput. Further, the traffic load of an AP may also increase suddenly due to the increase in application traffic demand. For this reason, admission control mechanisms [65]

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

can not fully solve the problem of overloading. The performance of the network can be improved if some MNs from the overloaded APs are transferred to other APs in the overlapping region, such that the performance of the transferred MNs do not get affected.

This chapter proposes an inter-BSS handoff procedure depending on the traffic load variation in the network. If the performance of an AP gets degraded due to the over-association of MNs, some of the MNs are re-associated to the neighbor APs, which are in the overlapping region. For this purpose, the MNs use an interleaved scanning mechanism to find out all the APs in the overlapping regions. The decision of hand-over is based on the cooperation of the AP and the MNs. The hand-off decision is designed based on a criteria that poses minimum scanning and signaling overheads. For this reason, a handoff metric is designed that considers both the MN's throughput and forwarding delay while measuring performance degradation. The MN decides about impending handoff based on the estimation of performance degradation, and with cooperation of the associated AP. The results from an indoor testbed is reported in this chapter to show the performance improvement of the proposed scheme in terms of throughput, delay and fairness. Further, the context-aware handoff scheme is compared with other related schemes proposed in the literature, through simulation results.

The rest of the chapter is organized as follows. An analytical model has been proposed in Section 3.1 to justify the requirement for load-balancing during inter-BSS handoff in IEEE 802.11 WLAN. Section 3.2 discusses about the design details of the proposed context-aware inter-BSS handoff procedure. The performance of the proposed scheme is evaluated through an indoor testbed, as well as using the results from simulation, as reported in Section 3.3. Finally, Section 3.4 concludes the chapter.

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

This section analyzes the effect of load-balancing over inter-BSS hand-off in a IEEE 802.11 ESS. According to the design criteria, it has been assumed that there are WLAN APs deployed in an area with considerable overlapping of coverage areas. Every AP forms a BSS. For the theoretical analysis, we have made the following assumptions: -

- The APs are uniformly distributed in the area.
- Every mobile node is always in the coverage of at least two APs. Each point in the coverage area of an AP is also covered by at least one other AP. It can be noted

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

that boundary MNs, which may be within the communication range of a single AP, are not considered for the theoretical modeling purpose, as they do not go through impending handoff due to load-balancing.

The basic idea used here is the conservation of traffic in a BSS. The total input and output traffic distribution in a BSS k is shown in Figure 3.1.

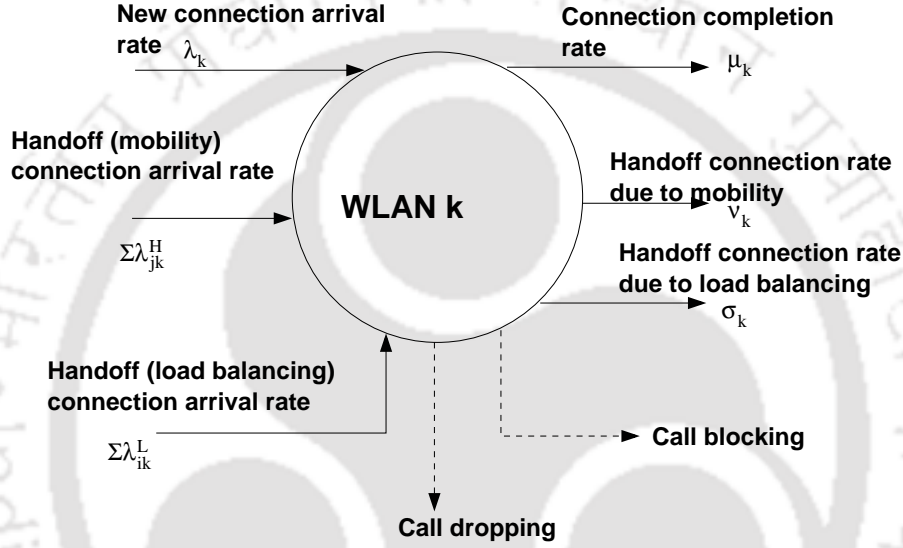


Figure 3.1: BSS Traffic Model

Suppose that there are N number of WLAN APs deployed uniformly in the arena. Let, W_k denotes the set of BSSs adjacent to the BSS k , and C_k denotes the set containing the overlapping BSSs with the BSS k . In this scenario, the total traffic arrival rate in BSS k has three components,

- (i) the arrival rate of newly generated connections (λ_k),
- (ii) total connection handoff rate due to mobility ($\sum_{\forall j \in W_k} \lambda_{jk}^H$), and
- (iii) total connection handoff rate due to the load balancing ($\sum_{\forall i \in C_k} \lambda_{ik}^L$).

Similarly, the total traffic departure from a BSS k has three components,

- (i) the connection completion rate (μ_k),
- (ii) connection handoff rate due to mobility (ν_k), and
- (iii) connection handoff rate due to load-balancing (σ_k).

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

3.1.1 Modeling the Bandwidth Evolution within a BSS

Suppose, the arrival rate of newly generated connections under BSS k be λ_k , and the new connection completion rate (the duration for which a connection remains active) under BSS k be μ_k . So the channel holding time of a connection in BSS k is exponentially distributed with mean $\frac{1}{\mu_k}$. At the end of the holding time, a connection in BSS k may terminate and leave the system with probability T_k , or move within the system and continue in an adjacent BSS (or an overlapping BSS) with probability $1 - T_k$.

In the proposed architecture, a handoff can occur because of two reasons - either due to mobility, or due to load balancing. Let X_{km}^H denotes the probability of attempting a handoff to an adjacent BSS m due to the mobility, and X_{kl}^L denotes the probability of attempting a handoff to another adjacent BSS l due to the load balancing. Therefore,

$$1 - T_k = \sum_{m \in W_k} X_{km}^H + \sum_{l \in C_k} X_{kl}^L \quad (3.1)$$

To prioritize handoff connections over newly generated connections, a new reservation parameter R_k is defined for BSS k . Let, B_k be the total capacity for BSS k (in terms of maximum number of connections that can be supported in an AP), and n_k be the number of active connections. When $n_k \leq R_k$, then BSS k accepts both handoff and newly generated connections. However, for $n_k > R_k$, it only accepts handoff requests, and the newly generated connections are blocked. Similarly, if $n_k = B_k$, then all the future handoff connections are dropped and newly generated connections are blocked. The bandwidth occupancy of the BSS k evolves according to a birth-death process, independent of other BSSs, as shown in Figure 3.2. Every stage of the birth-death process for BSS k represents number of active connections associated with the AP of BSS k .

Let $\lambda_k(i)$ denotes the total connection arrival rate at stage i of the BSS k . Therefore,

$$\lambda_k(i) = \begin{cases} \lambda_k + \sum_{m \in W_k} \lambda_{mk}^H + \sum_{l \in C_k} \lambda_{lk}^L & \text{if } 0 \leq i \leq R_k \\ \sum_{m \in W_k} \lambda_{mk}^H + \sum_{l \in C_k} \lambda_{lk}^L & \text{if } R_k < i \leq B_k \end{cases} \quad (3.2)$$

If a BSS can not support more connections due to traffic overload, then the newly generated connections may be blocked, and the existing connections may be dropped. As discussed, connection blocking (for new connections) is given priority over the connection dropping (for the existing connections) in the proposed system through the reservation parameter R_k . This in turn prioritizes existing connections over the new connections when

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

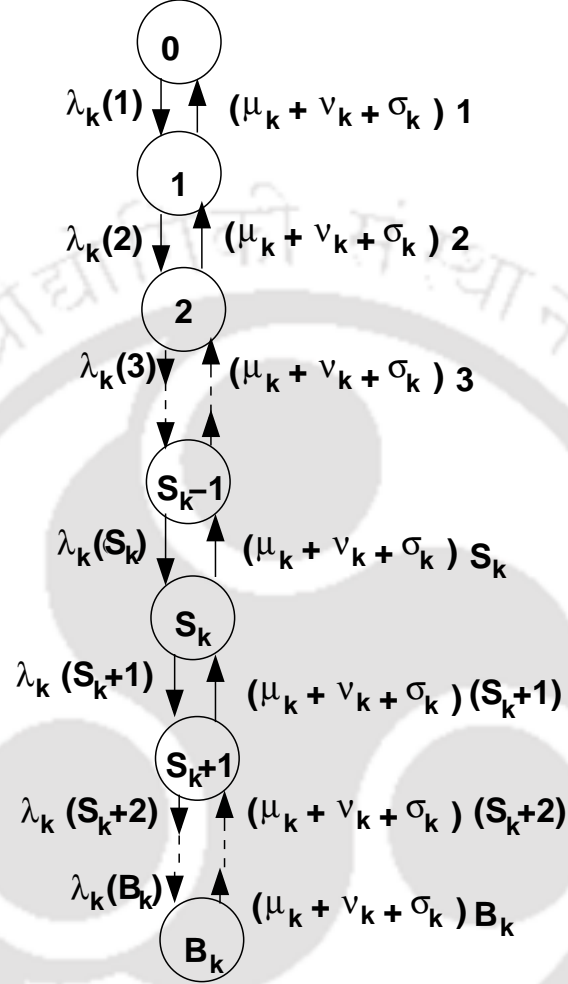


Figure 3.2: A Birth-Death Process Model for Bandwidth Occupancy at BSS k

the available bandwidth is not sufficient to accommodate all the connections. Suppose, b_k denotes the ‘connection blocking probability’ for the newly generated connections, and d_k denotes the ‘connection dropping probability’ for the existing connections. The ‘connection blocking probability’ is defined as the percentage of new connections blocked to the total number of new connections arrived. On the other hand, ‘connection dropping probability’ is defined as the percentage of existing connections dropped to the total number of existing connections. Accordingly, λ_{km}^H and λ_{kl}^L are expressed as follows;

$$\lambda_{km}^H = X_{km}^H \left(\lambda_k (1 - b_k) + \left(\sum_{u \in W_k} \lambda_{uk}^H + \sum_{v \in C_k} \lambda_{vk}^L \right) (1 - d_k) \right) \quad (3.3)$$

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

$$\lambda_{kl}^L = X_{kl}^L \left(\lambda_k (1 - b_k) \gamma_{kl} + \left(\sum_{u \in W_k} \lambda_{uk}^H + \sum_{v \in C_k} \lambda_{vk}^L \right) (1 - d_k) \gamma_{kl} \right) \quad (3.4)$$

Where γ_{kl} is the coverage factor between BSS k and BSS l , that is the ratio between the radio coverage area of the BSS k and the BSS l , with $0 < \gamma_{kl} < 1$.

Let $\pi_k(i)$ is the steady state probability that i active connections are established in BSS k at stage i . The balanced equations of the birth-death process, given in Figure 3.2, can be expressed as follows;

1. If $0 \leq i \leq R_k$:

$$\pi_k(i-1) \left(\lambda_k + \sum_{u \in W_k} \lambda_{uk}^H + \sum_{v \in C_k} \lambda_{vk}^L \right) = i \pi_k(i) (\mu_k + \nu_k + \sigma_k) \quad (3.5)$$

2. If $R_k < i \leq B_k$:

$$\pi_k(i-1) \left(\sum_{u \in W_k} \lambda_{uk}^H + \sum_{v \in C_k} \lambda_{vk}^L \right) = i \pi_k(i) (\mu_k + \nu_k + \sigma_k) \quad (3.6)$$

The newly arrived connections are blocked when $R_k < i \leq B_k$. Therefore,

$$b_k = \sum_{i=R_k+1}^{B_k} \pi_k(i) \quad (3.7)$$

Similarly, the handoff connections are dropped when $i = B_k$. Therefore,

$$d_k = \pi_k(B_k) \quad (3.8)$$

b_k and d_k can be calculated using fixed point iterations. These two parameters gives an estimation of the effect of load balancing on the handoff. Due to the load-balancing, the connection blocking and connection dropping should be less. Further the analysis introduces a reservation parameter R_k that reflects the scenario when on-demand load-balancing is performed, that is hand-over connections are dropped only when the remaining bandwidth is less than R_k . This shows that even on-demand load-balancing

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

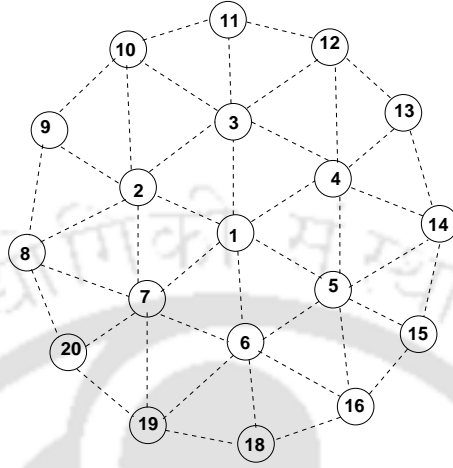


Figure 3.3: Network Scenario for Simulation and Numerical Analysis: Every circle represents an AP, and the dotted lines represent the overlapping in the communication coverage of two APs. All the APs are interconnected through distribution system

can significantly improve network performance rather than a perfect load balancing. Later during performance analysis, it is depicted that the proposed scheme can provide similar network performance compared to a perfect load-balancing scheme [37], with a much lower signaling overhead.

3.1.2 Numerical Results and Comparison with Simulation

This subsection compares numerical results obtained from the theoretical analysis with the simulation results, and analyzes the effect of load balancing through numerical results. The proposed load balancing based handoff scheme is implemented in Qualnet 5.0.1 network simulator [66] using discrete event simulation framework. The detailed scenario setup and the results are reported as follows.

Scenario Setup

A 20 AP network is considered for both the theoretical and the simulation purposes, as shown in Figure 3.3. The central AP (marked as AP 1) is considered to measure the output parameters (namely connection blocking probability and connection dropping probability). This type of scenario is considered to avoid the edge effect at the APs, such that sufficient overlapping is available at the central AP. There exist sufficient overlapping (around 20% overlapping connection coverage) among the neighbor APs. 40 mobile nodes are uniformly distributed in the network. A trace based simulation model is executed to

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

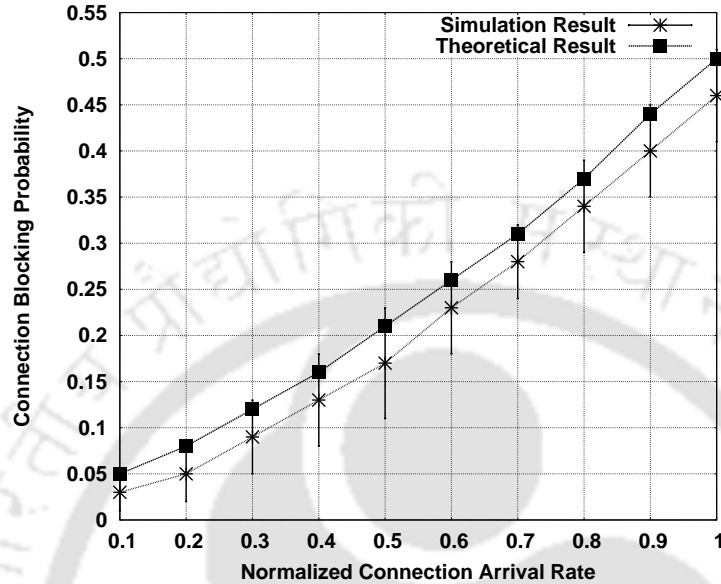


Figure 3.4: Connection Blocking Probability

match the association between the APs and the MNs for both the theoretical and the simulation analysis. In the trace based model, similar association scenario is generated as directed by the theoretical model, based on the load-balancing criteria. The mobility of the nodes are considered based on two-dimensional fixed direction mobility model, to make connection arrival rate predictable, which is necessary to compare the simulation output with the theoretical model. Later, the effect of different mobility model has been analyzed during performance analysis. 50 static nodes (different from the MNs) are uniformly deployed in the network to check the effect of handoff based on the load balancing. Every AP supports physical data rate of 6 Mbps. The static and the mobile nodes generate CBR traffic with a constant data rate of 64 kbps. However, the traffic start time is controlled based on the connection arrival rate. Every simulation setup is executed for 10 times with different seed values, and both the average result and the confidence intervals (the difference between the minimum and the maximum output) is shown in the graphs.

Comparison with Simulation

To compare the theoretical results with the simulation outputs, the results are plotted with respect to the normalized connection arrival rate. The normalized connection arrival rate is calculated as the ratio of the total incoming data rate to the total capacity of the APs. The vertical lines in the graphs shows the confidence intervals.

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

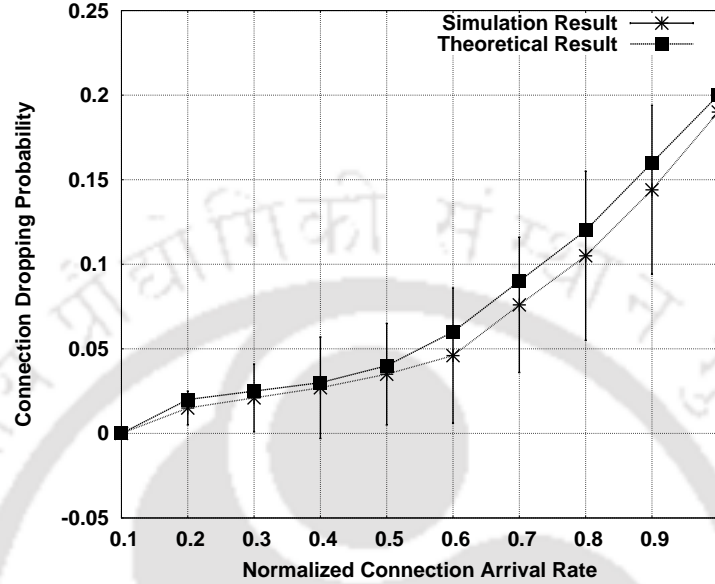


Figure 3.5: Connection Dropping Probability

Figure 3.4 shows the connection blocking probability and Figure 3.5 shows the connection dropping probability with respect to the normalized connection arrival rate. It can be seen from both the figures that the theoretical results are almost similar to the simulation results. The static nodes in the network participate in the handoff that occur due to load balancing among the APs, whereas the MNs participate in new connection establishment as well as handoff (due to both the mobility and the load balancing). For this analysis, R_k is taken as 70% of the total capacity. Therefore, if total bandwidth utilization goes beyond 70% of the total capacity, all the newly arrived connections are blocked. In the present scenario, this situation occurs with the normalized connection arrival rate at 0.5. Therefore beyond this point, connection dropping probability increases significantly, as shown in Figure 3.5.

3.1.3 Analyzing the Effect of Load-balancing over Inter-BSS Handoff through Numerical Results

This subsection analyzes the performance impact due to load-balancing over the inter-BSS handoff in IEEE 802.11 WLAN network. In the figures, 'CBP' and 'CDP' denote connection blocking probability and connection dropping probability, respectively. The impact of load-balancing is compared with the standard inter-BSS handoff. The numerical results for standard inter-BSS handoff is obtained by considering no handoff due to traffic

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

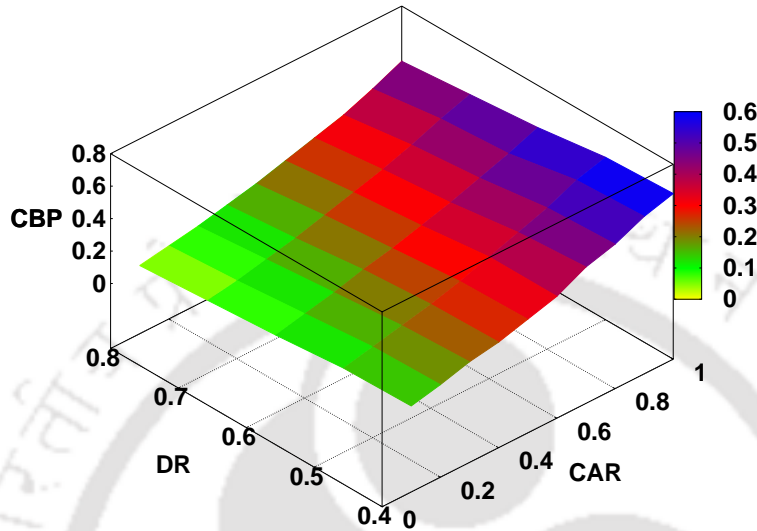


Figure 3.6: Connection Blocking Probability (CBP): Load-Balancing based Inter-BSS Handoff

overload, while assuming $X_{lk}^L = 0$. In this scenario, when $X_{lk}^L = 0$, R_k does not make any impact over the connection blocking and the connection dropping probabilities. Similar network scenario, as discussed in previous subsection, is also considered for this analysis.

Effect on Connection Blocking Probability

Figure 3.6 shows the impact of load-balancing over connection blocking probability during inter-BSS handoff. As discussed earlier, new connections are blocked when an AP can not accept new connections due to capacity limitation. In the figures, ‘DR’ represents the reservation parameter R_k , and ‘CAR’ represents *connection arrival rate*. Figure 3.7 shows the connection blocking probability for the standard inter-BSS handoff. A comparison between Figure 3.6 and Figure 3.7 show that connection blocking probability is considerably less for the load-balanced network. The reservation parameter R_k further reduces the connection blocking probability for a network that supports load-balancing based handoff. The R_k value indicates how much bandwidth should be reserved for the existing handoff connections, which in turn prioritizes the existing connections over the new connections. As a result, the connection blocking probability increases as the value of R_k is decreased.

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

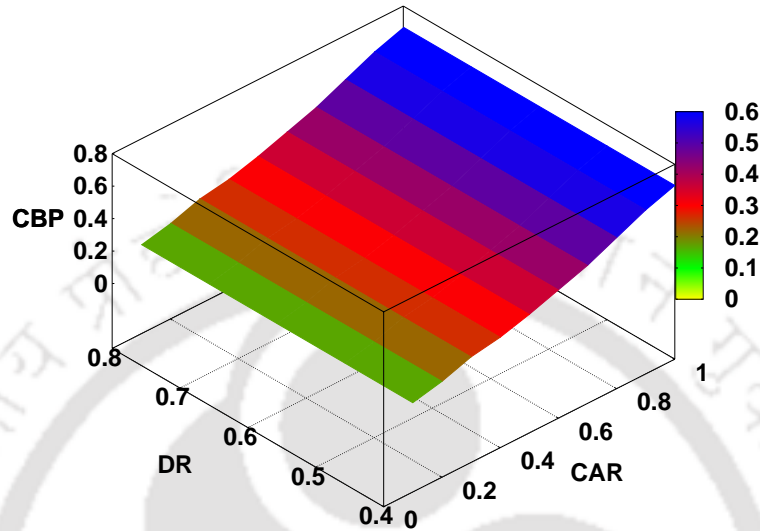


Figure 3.7: Connection Blocking Probability (CBP): Standard Inter-BSS Handoff

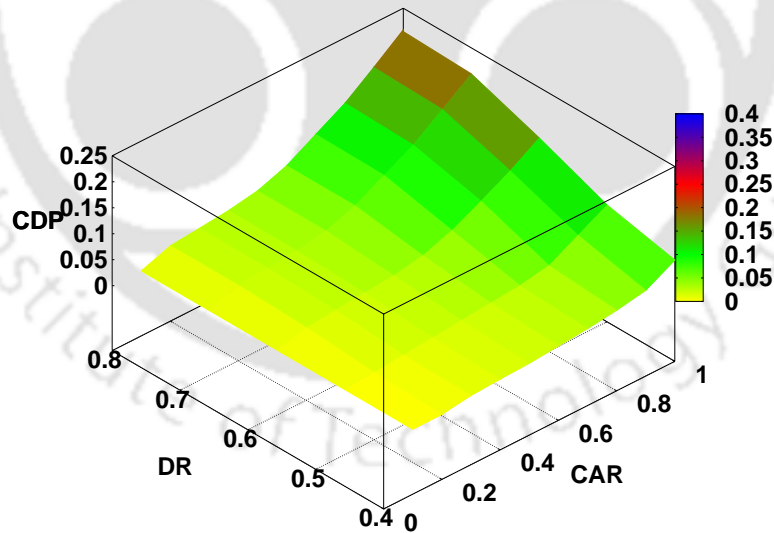


Figure 3.8: Connection Dropping Probability (CDP): Load-Balancing based Inter-BSS Handoff

Effect on Connection Dropping probability

Figure 3.8 shows the call dropping probability for the load-balancing based inter-BSS handoff, and Figure 3.9 shows the same for the standard inter-BSS handoff. As earlier,

3.1 Analytical Model: Effect of Load Balancing over Inter-BSS Hand-off

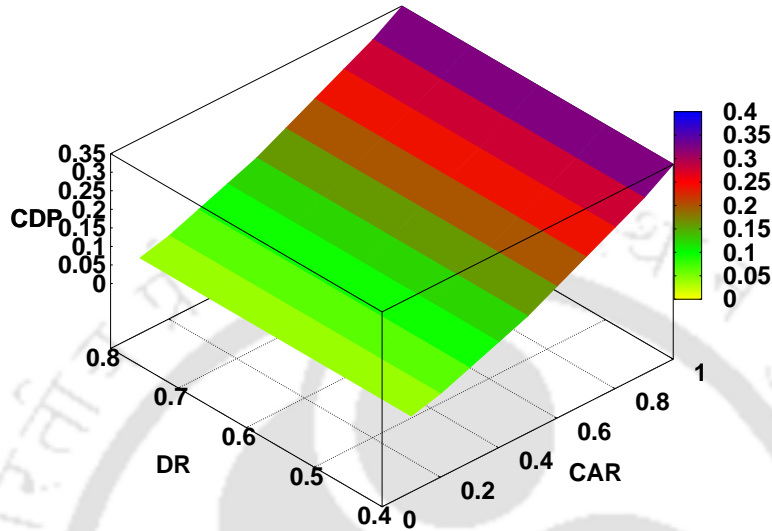


Figure 3.9: Connection Dropping Probability (CDP): Standard Inter-BSS Handoff

the load-balancing procedure can significantly reduce connection dropping probability for the existing connections. One interesting observation can be made by comparing Figure 3.6 with Figure 3.8. As the value of the reservation parameter R_k decreases, the call blocking probability increases, however, the call dropping probability decreases. As discussed earlier, the R_k value indicates how much bandwidth should be reserved for the existing connections, which in turn prioritizes the existing connections over new connections. Therefore, a lower value of R_k indicates that existing connections are more prioritized compared to new connections by reserving more amount of bandwidth for them. As a result, call dropping probability decreases with the decrease of R_k value. However, the call blocking probability gets increased. A suitable value of R_k is required to be decided for the proper distribution of the available bandwidth between the new connections and the existing connections.

The analytical model presented in this chapter shows that load-balancing among overlapping BSSs can significantly improve network performance by reducing call dropping and call blocking probabilities. Due to load-balancing, more number of new connections can be accommodated by shifting some of the existing connections to the neighboring APs with lower traffic load. Based on this observation, a context aware inter-BSS handoff scheme is designed, where the existing connections are shifted to the neighboring APs on detection of the performance degradation due to load-imbalance. The AP detects

the performance degradation due to load-imbalance, and the handoff decision is taken cooperatively by the AP and the MNs, as discussed in the following section.

3.2 Context-aware Inter-BSS Handoff Procedure

As discussed earlier, the existing load balancing schemes in IEEE 802.11 infrastructure networks are found to be inadequate because of these schemes, while trying for a perfect balancing of load, incur large signaling overhead. Also, these schemes do not have a proper scanning mechanism for multi-channel network. In this chapter we propose a context-aware inter-BSS handoff scheme, where load in an ESS having different APs with overlapped coverage area, are adjusted by changing the association of MNs that experience degradation of performance. The adjustment is aimed at improving the throughput and delay of the suffering MNs and in turn the overall performance of the ESS. The design of the proposed context aware handoff procedure is based on following assumptions.

- All APs are connected through a distribution system, and form an ESS.
- The APs have overlapping coverage area in such a way that a MN gets sufficient switchover time if a performance degradation is observed due to traffic overload at the associated AP.
- The overlapping BSSs operate in non-overlapping channels. According to IEEE 802.11b/g physical layer standard, there can exist a maximum of three non-overlapping channels, and therefore every point in the target region is maximum three connectivity covered, if IEEE 802.11b/g is used as the physical communication standard. IEEE 802.11n supports more number of non-overlapping channels, and accordingly every point of the target region can be covered by more number of BSSs, as supported by the standard.

All these assumptions are practically implementable in public HotSpots, where every target region is covered by multiple APs. Based on these assumptions, the handoff scheme is designed, as follows.

The proposed scheme works in three phases,

- (i) Every MN use a scanning mechanism to find out the list of APs with which it can be associated. The proposed scanning mechanism, termed as *interleaved scanning* in this chapter poses very low signaling overhead at actual handover time reducing

3.2 Context-aware Inter-BSS Handoff Procedure

handoff latency. The list of APs found through the interleaved scanning is stored in a list structure, called *ScanAPList*.

- (ii) Every MN periodically measures the degradation in performance in terms of throughput and delay.
- (iii) Once a severe performance degradation is observed, a MN consults with the associated AP to find out another AP from the *ScanAPList* with which it can re-associate. Depending on the coordination between the MN and the associated AP, the handoff procedure occurs.

These three phases are discussed next in details.

3.2.1 Interleaved Scanning

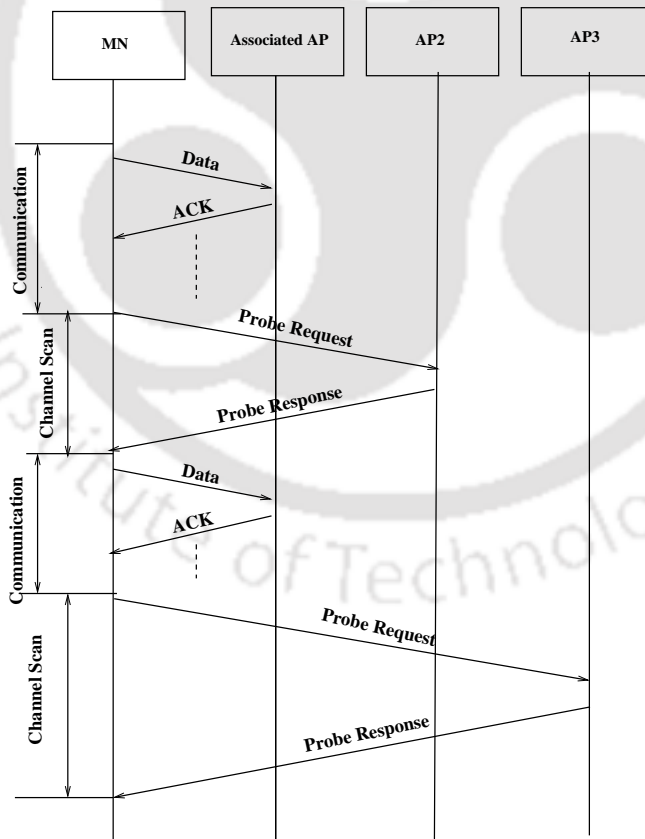


Figure 3.10: Interleave Scanning

As shown in Figure 3.10, the proposed scheme uses a scanning mechanism to find

3.2 Context-aware Inter-BSS Handoff Procedure

out the alternate APs, where the scanning and the data communication are interleaved in time. The MN keeps the information of different APs found during the scanning in a list called *ScanAPList*. In the interleaved scanning mechanism, the MN prepares a circular list of channel numbers to scan. These are the channels that contains all the IEEE 802.11 supported channels (based on physical layer standard), except the current channel and the channels that overlaps with the current channel. It initializes a variable *ProbeChannel* to the channel number of the first channel in the list. The *ScanAPList* is initialized to an empty set (denoted as \emptyset). It also starts a timer, referred as *ScanRepeatTimer*. Let the time out value of this timer be T . On the expiry of time T , the timer event is triggered and the MN executes the following steps in order.

- Step 1: The MN completes the ongoing packet transmissions (if any) with the associated AP, and subsequently stops transmitting further packets.
- Step 2: It stores the current channel number into a temporary variable *CommChannel*, and deletes all entries from the *ScanAPList* corresponding to the channel *ProbeChannel*.
- Step 3: The MN then switches the operating channel to *ProbeChannel*.
- Step 4: It sends a probe request at that channel, and waits for a time period of t_1 .
- Step 5: If the channel is sensed busy during t_1 time period, it extends the waiting time by another time period of t_2 .
- Step 6: If any probe response is received within the waiting time, the AP information obtained with the probe response is stored into the *ScanAPList*, corresponding to the entry *ProbeChannel*.
- Step 7: It updates *ProbeChannel* to the entry which is cyclically next in the channel list.
- Step 8: It starts *ScanRepeatTimer*.
- Step 9: Finally, the MN switches the operating channel to *CommChannel*, and restarts the normal communication procedure for remaining packet transmissions.

The interleaved scanning mechanism proposed in this chapter is done simultaneously with the communication, in an interleaved way. Therefore, communication disruption due to scanning is very less, which significantly improves overall network performance. In the traditional approaches, as discussed in Chapter 2, MNs scan all the channels in a single run,

3.2 Context-aware Inter-BSS Handoff Procedure

incurring communication disruption for a long duration, which significantly affect network performance. Once the AP information is gathered through the interleaved scanning mechanism, the handoff decision is taken once a MN detects performance degradation due to load-imbalance at the current associated AP, as discussed in the next subsection.

3.2.2 Detection of Performance Degradation

In the proposed context-aware inter-BSS handoff mechanism, the degradation of performance is sensed from two different parameters. As the first parameter, the packet drop rate is used to measure the degradation of performance. A moving average of the number of packet dropped within a time window is used as the measurement metric. In this method, whenever the average packet drop increases above some threshold, the MN tries to find an alternate AP for association. It is observed that, when the traffic in the ESS increases gradually, packets gets delayed considerably even when the throughput does not get effected. Therefore, changing the association of MNs based on the measurement in packet drop can result in the performance improvement in terms of throughput. However, there may be performance degradation in terms of delay, as packet drop rate does not account forwarding delay.

In the second parameter, the degradation of performance is sensed by observing the interface queue length at the associated AP, when a MN detects performance degradation in terms of average packet drops. A MN can obtain the interface queue length of the associated AP on request. The AP forwards the information through piggybacking with data packets transmitted to the MN, and therefore, it does not incur much overhead. To smooth out the variations in the interface queue length, the *exponentially weighted moving average* (EWMA) [67] of the queue length is calculated after every t units of time. Let E_{t-1} be the EWMA of the queue length, and Y_{t-1} be the actual queue length at time $t - 1$. Then we calculate E_t as,

$$E_t = \alpha \times Y_{t-1} + (1 - \alpha) \times E_{t-1}, 0 < \alpha \leq 1 \quad (3.9)$$

The value of α decides how fast the MN discards the old values of Y . Whenever E_t crosses a threshold, the MN tries to find an alternate AP for association.

In the proposed scheme, the MN uses a combination of the packet drop rate and the EWMA of the queue length as the performance measurement parameter. The combined parameter, termed as *expected channel quality degradation* (ECQD), can measure the performance in terms of both throughput (through packet drop rate) and delay (through

EWMA of the queue length), and defined as,

$$ECQD = \delta \times \frac{PDR}{PDR_{MAX}} + (1 - \delta) \times \frac{EQL}{Q_{MAX}} \quad (3.10)$$

where PDR denotes the measured packet drop rate, EQL denotes the measured EQMA of the queue length (equation (3.9)), PDR_{MAX} denotes the maximum packet drop rate supported by the network, and Q_{MAX} denotes the maximum queue length at the AP. In equation (3.10), δ is a smoothing parameter. $\delta = 0.5$ gives equal weight to the throughput and the delay as the performance metric.

Once a MN detects the performance degradation through the measurement in the ECQD metric, it initiates a handoff procedure with cooperation of the associated AP. The details of the handoff procedure is explained in the next subsection.

3.2.3 The Handoff Procedure

Whenever a MN senses degradation of performance, it scans the different channels using the interleaved scanning mechanism to find suitable APs to which it can perform handoff. It can be noted that though interleaved scanning can be performed periodically, the proposed scheme starts scanning only after the sensing of performance degradation, to limit the signaling overhead due to scanning. Once all the channels are scanned, further scanning is disabled till the handover procedure is completed. Through the interleaved scanning mechanism, a MN collects a list of APs with which it can re-associate if needed. As mentioned earlier, the list of such APs are stored in the variable *ScanAPList*. The MN informs its associated AP, say AP_a , about its desire to disassociate from AP_a and re-associate with another AP where a better performance may be expected. For this purpose, the MN sends a *MoveRequest* message, containing the *ScanAPList* and its desired ECQD value, to AP_a . The MN keeps updating the ECQD metric, and sends the *MoveRequest* message repeatedly at a fixed interval of t_{repeat} , for upto n_{repeat} times or until it gets a response from AP_a .

Whenever AP_a receives a *MoveRequest* message from a MN, it sends a *LoadRequest* message to all the APs in the *ScanAPList* through the distribution system. The APs receiving the *LoadRequest* message, replies with a *LoadResponse* message through the distribution system containing its current load status in terms of EWMA of the queue length. It can be noted that packet drop rate is measured at the MN, and therefore only EWMA of the queue length for the APs is considered at this stage. AP_a receives the *LoadResponse* message, and based on the replies it receives and the *ScanAPList*, it

3.2 Context-aware Inter-BSS Handoff Procedure

selects one or more APs to which the MN can handoff. AP_a keeps the list of such APs in *HandoffCandidateList*, which is constructed as follows.

Let AP_i be the i^{th} AP in the ESS. Let L_a be the EWMA of the queue length for AP_a . Let L_i be the EWMA of the queue length for the AP_i . Then *HandoffCandidateList* is constructed as follows,

$$\text{HandoffCandidateList} = \{AP_i | L_i - L_a < \phi\} \quad (3.11)$$

Where ϕ is a small fraction of the maximum throughput decided a priori by the service provider. If the value of ϕ is too low, then the MN that performed handoff to get better performance may tend to handoff again, and therefore too many handoffs may occur unnecessarily. If the value of ϕ is too large, potential opportunity to increase performance may be lost.

The AP sends the *HandoffCandidateList* in its reply message, which is termed as *HandoffTargetMessage*, to the MN. The MN, based on the RSSI, chooses one of the APs in this list and performs handoff to that AP (AP_N) from which it receives the strongest RSSI.

It is likely that more than one MN sends *MoveRequest* messages to one AP. If all such MNs are allowed to handoff to another AP, a severe imbalance in load may occur, and therefore may result in performance degradation. To prevent it, whenever an AP processes a *MoveRequest* message, it drops all further *MoveRequest* messages from other MNs for some fix period of time T_{ignore} .

After selecting an AP (AP_N) from the *HandoffCandidateList*, a MN sends a pre-association message to AP_N . It can be noted that the pre-association messages differs from the normal association messages in the sense that it reflects an existing connection to be re-associated. Every AP sets a bandwidth limit, R_k as discussed in the theoretical analysis, beyond which it only accepts pre-association messages, and drops any new association messages. This way the existing connections are prioritized over the new connections, so that minimum QoS can be assured for them. Further, the associated AP, AP_a , forwards the impending handoff information to AP_N through the distribution system, which minimizes the handoff latency. Scanning is re-enabled after handover and actual scanning may start if performance degradation is detected and the whole process repeats again.

Though this chapter does not discuss about security aspects during handoff, it can be implemented along with the re-association mechanism. AP_a generates a random nonce and forwards it to the MN as well as to AP_N . As earlier, AP_a communicates with AP_N through the distribution system. For the nonce distribution, a *channel handshake*

authentication protocol (CHAP) [68] protocol is used for secured communication. Then AP_N can authenticate the MN through the random nonce received from AP_a .

The next section analyzes the performance of the proposed scheme through simulation results and the data obtained from an indoor wireless testbed.

3.3 Performance Analysis

The performance of the proposed context-aware inter-BSS handoff scheme is evaluated through a practical testbed as well as from simulation results. The results from the testbed is used to show the performance improvement of the proposed scheme in a real network scenario. On the other hand, the scalability of the proposed scheme is analyzed using simulation results. Further the simulation results are used to compare the performance of the proposed scheme with similar approaches proposed in the literature, such as [37] and [42].

3.3.1 Testbed Analysis

The proposed scheme is implemented and evaluated using a 6-node 802.11g indoor mesh testbed deployed over the IIT Guwahati computer science department research labs, as shown in Figure 3.11. Each AP in the testbed is a Skiva Easyconnect RT001 N300 WiFi router with RaLink RT-3352 chipset [69]. The Ralink RT-3352 router on chip combines 802.11b/g/n draft compliant 2T2R MAC/BBP/PA/RF, a high performance 400MHz MIPS24KEc CPU core, a Gigabit Ethernet MAC, 5-ports integrated 10/100 Ethernet Switch/PHY, 64MB of SDRAM and 32MB of Flash. The routers are configured to operate in IEEE 802.11g mode. This mode of operation can support up to 54 Mbps data rate with maximum transmission power of 200 mW, that corresponds to average 30m communication range and 45m interference range, with a deviation of $\pm 5m$ based on the external noise factor.

Testbed Setup

The APs are interconnected through a distribution system that is connected to the Internet through a proxy gateway. In the testbed setup, the gateway node works as the corresponding node (CN) and the MNs communicate with the CN through the APs. 50 static nodes are distributed in the network, which constantly communicate with the CN, and increases the traffic load at the APs. The static nodes are desktop computers with wireless interface, which are placed in the research labs, at the nearby locations of

3.3 Performance Analysis

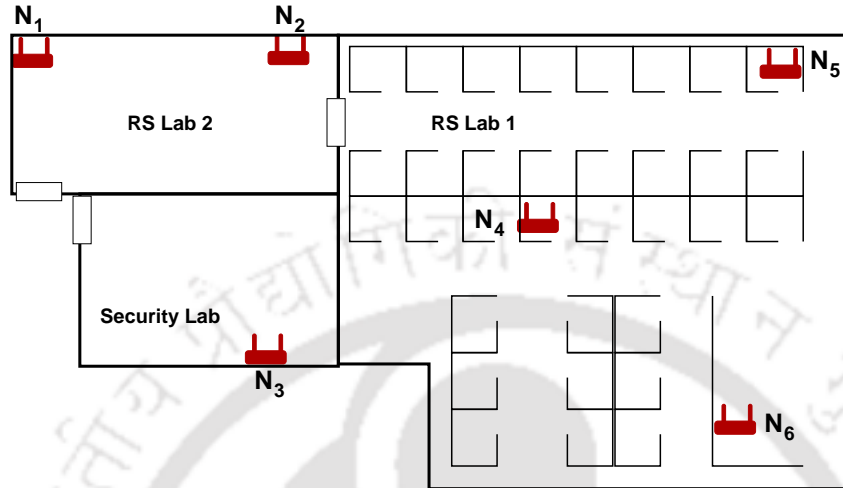


Figure 3.11: Indoor Testbed Layout

the APs. The performance is measured at 5 MNs which are distributed in the network. The MNs use a random mobility, and tends to concentrate near the AP N_4 . In the traditional approaches, those MNs associate with the AP N_4 always. However in the proposed scheme, they may re-associate with the neighbor APs (N_2 or N_5) that shares non-overlapping channels with AP N_4 . It can be noted that a static channel assignment algorithm is used based on the IEEE 802.11g channel model.

Trivial File Transfer Protocol (TFTP) is used as the application layer traffic, that uses UDP at the network layer. The traffic generation is based on a Poisson distribution with mean traffic generation rate as 10 Mbps per client (static or mobile nodes). The variation in traffic generation rate is around 5 Mbps. Therefore traffic generation rate at every client varies in between 5 Mbps to 15 Mbps. As indicated, the AP operates at 54 Mbps data rates. In the implementation of the context-aware inter-BSS handoff procedure, R_k , α and δ are fixed at 75% of the total capacity, 0.6, and 0.5 respectively. The value of ϕ is taken as the 20% of the interface queue length of AP_a .

Performance Measurement during Load-imbalance Scenario

This set of experiments shows the performance of the proposed scheme at high traffic load (average traffic generation rate is 10 Mbps, with a variation of 5 Mbps) when the problem of load-imbalance is a serious issue, and compares the performance with the RSSI based load balancing scheme.

Figure 3.12 shows the average throughput for AP N_4 with respect to time, where

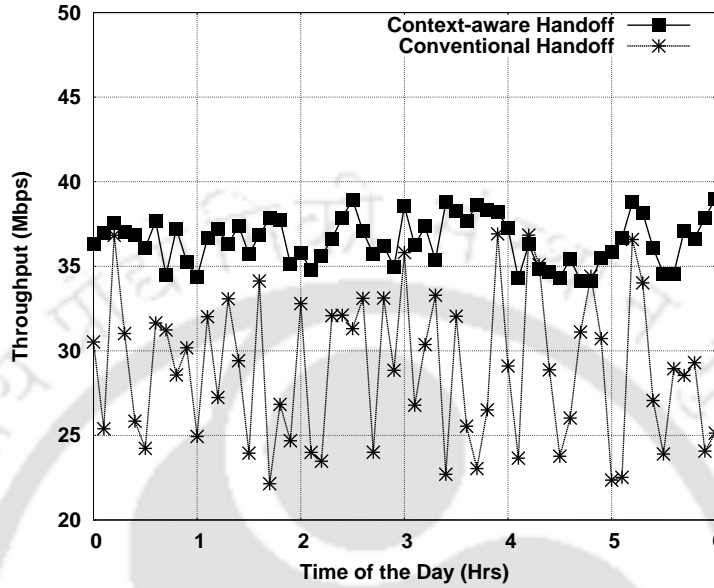


Figure 3.12: AP Throughput (N_4)

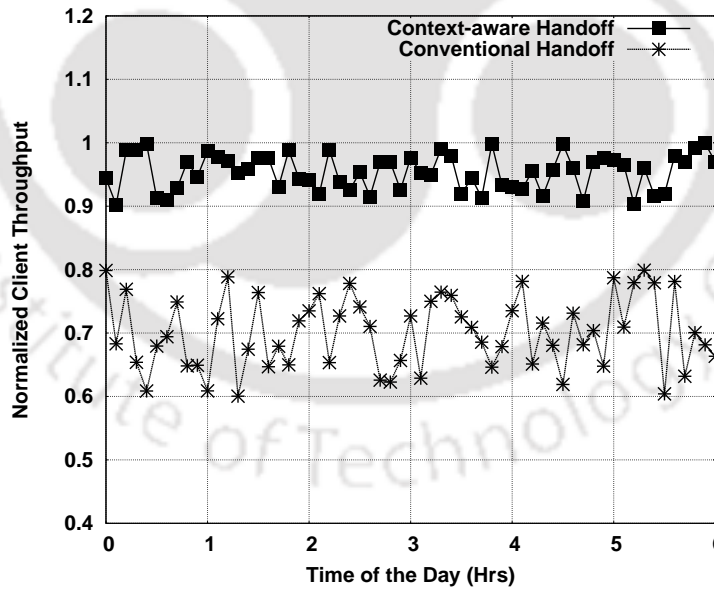


Figure 3.13: Normalized MN Throughput

the problem of load-imbalance is serious due to MN concentration. The AP throughput is calculated as the amount of data packet transmitted successfully through this AP. The figure shows that the throughput for the conventional RSSI based scheme (termed as *conventional handoff* in the figure) varies significantly with respect to time. Sometime,

3.3 Performance Analysis

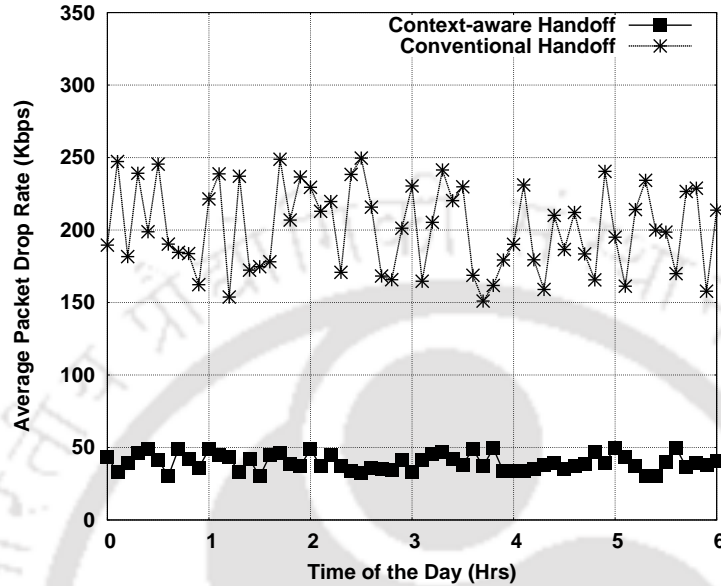


Figure 3.14: Packet Drop Rate

the throughput even drops as low as 15 Mbps. It has been observed from the data trace that at those times, all the clients (static and mobile nodes) starts transmitting traffic near to the maximum rate, that is around 15 Mbps. The figure reveals the the throughput improves significantly for the proposed context-aware scheme, both in aggregate and in variation. The data trace for the proposed scheme has confirmed that when the traffic load becomes higher, some of the MN de-associate from N_4 , and re-associate to either N_2 or N_5 , which share non-overlapping channels with N_4 .

Figure 3.13 shows the normalized client throughput for the MNs in the network. The normalized client throughput is calculated as the ratio of data received successfully to the data sent. The figure shows that the MN performance improves significantly in the proposed scheme, in terms of normalized throughput. It has been observed from the data trace that when a MN detects a degradation in the performance, it switches to another AP, and the performance gets improved.

Figure 3.14 compares the proposed scheme with the conventional scheme in terms of packet drop rate. Packet drop rate is significantly lower for the proposed scheme, as revealed by the figure.

Figure 3.15 compares the two schemes in terms of fairness. *Jain Fairness Index* is used to measure the fairness, which is defined as follows. Let λ_i denotes the throughput for flow i , and there are n number of flows in the network. Then Jain fairness index, J_F

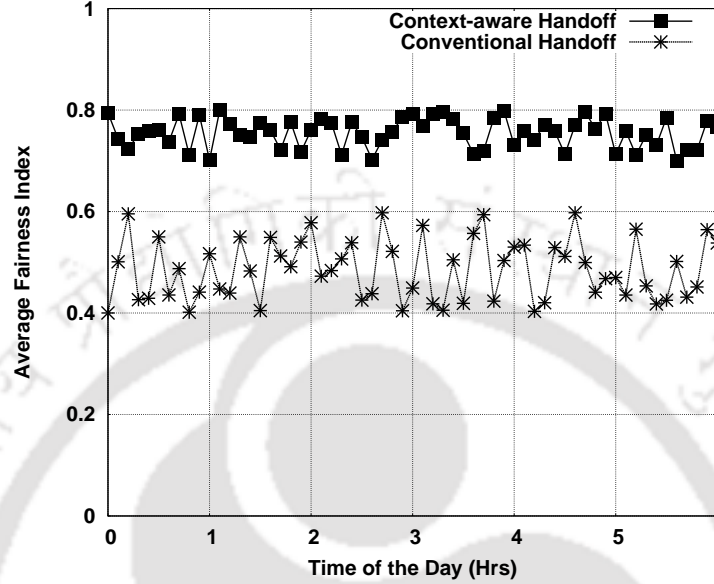


Figure 3.15: Fairness Index

is calculated as,

$$J_F = \frac{(\sum_{\forall i} \lambda_i)^2}{\sum_{\forall i} \lambda_i^2} \quad (3.12)$$

The value of the fairness index remains between zero and one, where a index value of 1 denotes perfect fairness. The figure shows that the proposed scheme improves fairness significantly. In the conventional scheme, the fairness index value degrades when the MN attains low throughput due to traffic overload at the associated APs. This problem is solved in the proposed context-aware handoff scheme which in turn improves fairness among the clients.

Figure 3.16 shows the per packet forwarding delay for the two schemes. The forwarding delay is computed for a MN, attached with AP N_4 . It can be seen from the figure that per packet delay varies significantly for the conventional RSSI based approach. Further, few packets experience very large forwarding delay. The forwarding delay is minimized in the proposed scheme by re-associating some of the MNs to other APs, when they experience a degradation in performance.

Performance Analysis by Varying Load

In this set of experiment, the performance of the proposed scheme is observed by varying data generation rate at the clients. The trace is collected for 3 hours, and the average is

3.3 Performance Analysis

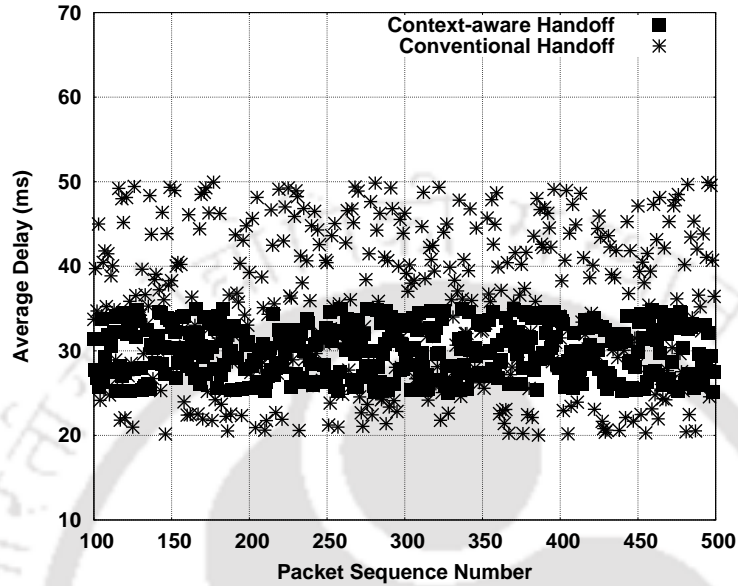


Figure 3.16: Forwarding Delay

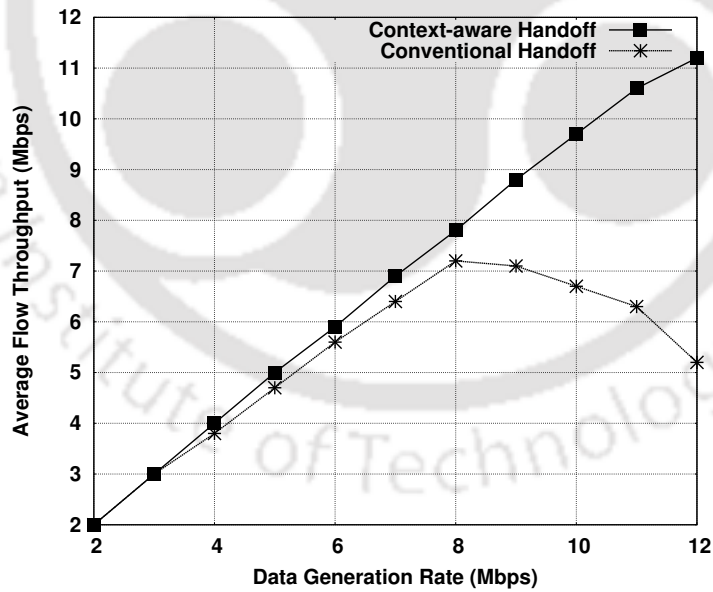


Figure 3.17: Average Flow Throughput

taken for plotting the graphs.

Figure 3.17 compares the two schemes in terms of average flow throughput from the clients (both static and mobile), and Figure 3.18 shows the average forwarding delay. As the traffic generation rate at the clients increases, the average flow throughput tends to

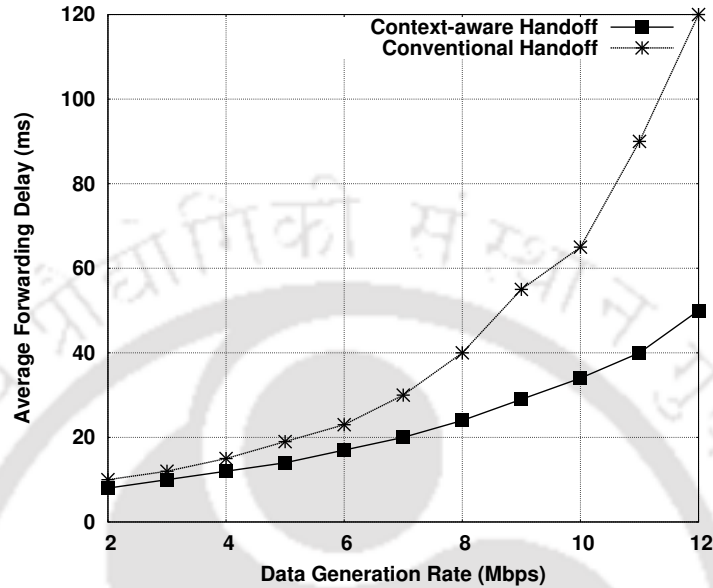


Figure 3.18: Average Forwarding Delay

decrease once network saturation is reached, that is, the total traffic generation rate at the interfering flows reach the maximum network capacity. While the AP gets saturated at 7.2 Mbps average per flow throughput, for the conventional approach, the proposed scheme can support more than 11 Mbps average per flow throughput. It has been observed from the trace data that some MNs gets re-associated to other APs, that increases the performance of the flows. As a result, the AP can support more traffic from the existing flows. Similar trend is observed in terms of forwarding delay. In the conventional approach, the forwarding delay increases exponentially once the AP gets saturated. The forwarding delay for the proposed scheme is significantly lower than the conventional approach.

3.3.2 Simulation Results and Comparison

The proposed scheme is implemented in Qualnet-5.0.1 [66] simulator framework, and the performance is compared with two other schemes, namely the perfect load balancing scheme proposed in [37], and the load balancing based on transmission power adjustment, as proposed in [42]. A 30 AP network scenario is considered for simulation, where the APs forms a cellular hexagonal structure, as shown in Figure 3.19. 100 MNs are uniformly deployed in the arena. The traffic type is similar to earlier that uses TFTP as the application layer protocol. Every simulation is executed 10 times, and the average is taken to plot the graphs.

3.3 Performance Analysis

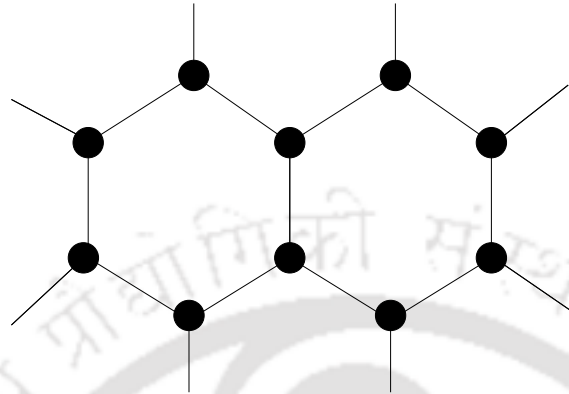


Figure 3.19: Simulation Scenario

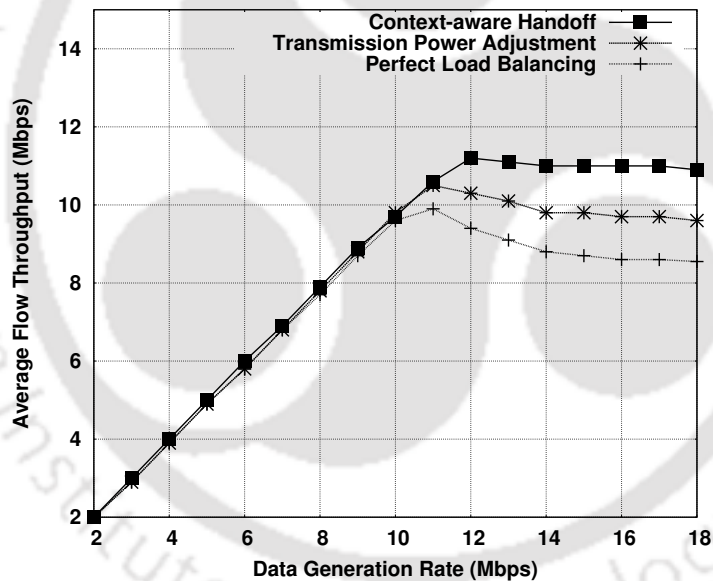


Figure 3.20: Average Flow Throughput

Figure 3.20 compares the three schemes in terms of average flow throughput. The perfect load balancing scheme, proposed by [37], incurs maximum signaling overhead, as a result of which the throughput at saturation is minimum among the three. The load balancing scheme based on transmission power adjustment [42] attains lesser throughput compared to the proposed scheme in this chapter because of uncontrolled interference effect due to power adjustment. Similar trend is observed in case of average packet forwarding delay, as shown in Figure 3.21. The proposed scheme results minimum delay among the three schemes. The interleaved scanning mechanism proposed in this chapter have

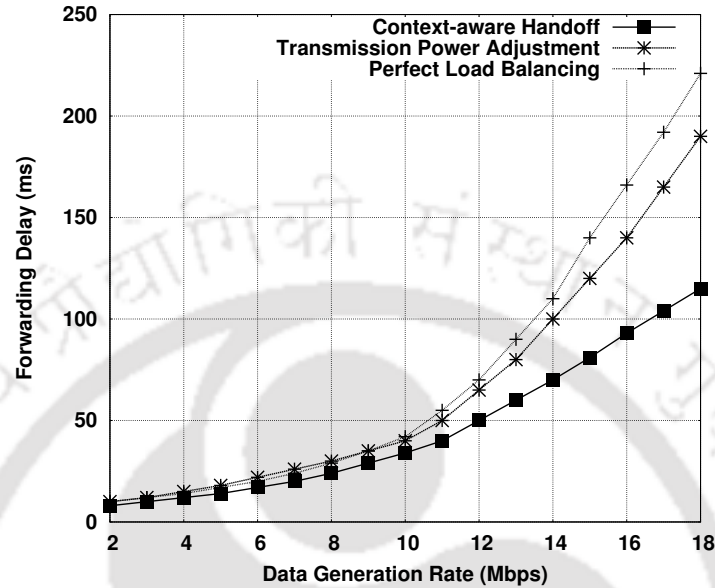


Figure 3.21: Average Forwarding Delay

signaling overhead at par with the standard handoff scheme, yet the signaling overhead during handoff is reduced thus improving handoff speed. At the same time, the re-association of the MNs are based on cooperative decisions with the APs. For this, the APs communicate with each other through the distribution system, which does not incur any extra overhead to the wireless interface. The metric used for handoff considers both delay and throughput during decision making. As a combined effect, the context-aware inter-BSS handoff scheme shows better performance compared to others, in terms of average flow throughput, and average forwarding delay.

3.4 Summary

This chapter proposes a context-aware inter-BSS handoff scheme for IEEE 802.11 WLAN, where the MNs switches between overlapping APs when a performance degradation is detected due to load-imbalance. A theoretical model is used to show the requirement of load-balancing based handoff in an IEEE 802.11 WLAN network. The theoretical model shows that on-demand load balancing is sufficient for performance improvement. An interleaved scanning mechanism is designed for this purpose, to find out the alternate APs in the vicinity of a MN. When a MN detects a degradation in performance, depending on a metric that combines both throughput and delay, it takes a handoff decision cooperatively

3.4 Summary

with the associated AP. The performance of the proposed scheme is analyzed using results from an indoor testbed. Further, simulation results are used to compare the performance of the proposed scheme with other similar schemes discussed in the literature.

The proposed scheme in this chapter considers a single-class traffic only. The next chapter extends the context-aware handover management procedure for QoS and QoE provisioning for multiple traffic classes.



Chapter 4

Context Aware Handover Management: Sustaining QoS and QoE in a Public IEEE 802.11e Hotspot

As discussed earlier, QoS and QoE is an end-to-end problem and therefore difficult to manage in a distributed wireless environment. The recent studies of the traffic distribution in a public wireless HotSpot, such as [70–73] and the references therein, reveal two important observations. First, the number of APs deployed is more than the actual requirement, and second, the users visit only a few APs. According to the standard inter-BSS handover mechanism, the users associate with the AP that provides maximum signal strength, and therefore most of the traffic loads get concentrated over a few APs. A number of schemes [74, 75] have been proposed in the literature for load balancing based AP selection during the inter-BSS handover procedure. However, none of these works had considered the effect of handover on the QoS of the network and QoE experienced by end users.

This chapter proposes a simple and cost-effective solution for context aware inter-BSS handover management along with QoS and QoE assurance for different classes of services. For proper bandwidth management during admission of new connections as well as during inter-BSS handover, a set of bandwidth reservation policies are designed. The effectiveness of these policies are analyzed using theoretical modeling. Based on the foundation of the theoretical modeling, the proposed scheme takes into account both the current service

4.1 Theoretical Foundation

load at the APs and the differentiated bandwidth occupancy according to the service level agreement (SLA) for different IEEE 802.11e service classes. The existing connections are handed over based on the service load of the APs as well as the bandwidth occupancy for that specific service class. Consequently, new connections are admitted only if the SLA for that service class can be assured. The proposed scheme is implemented in a real IEEE 802.11g+e testbed, and the performance is evaluated using extensive experiments for different QoS and QoE parameters, and compared with the standard and other similar approaches proposed in the literature.

The rest of the chapter is organized as follows. The foundations of the bandwidth management policies used in the proposed scheme is discussed in Section 4.1 using mathematical modeling. Based on the theoretical foundations, the context aware handover management scheme for QoS and QoE assurance in an IEEE 802.11 wireless LAN is introduced and discussed in Section 4.2. The proposed scheme is evaluated through the results from an IEEE 802.11g+e testbed, as reported in Section 4.3. Finally Section 4.4 concludes the chapter.

4.1 Theoretical Foundation: Context Aware Handover with Multi-class Traffic Services

The proposed handover management scheme to solve the QoS and QoE degradation problem at public wireless LAN is based on three strategic policies,

1. bandwidth reservation at every AP for different classes of services, according to IEEE 802.11e service differentiation (called the cutoff bandwidth strategy in this chapter),
2. differentiation in bandwidth reservation for new connections and handover connections where handover connections are given priority over the new connections (called the bandwidth flexibility strategy),
3. inter-BSS handover for load balancing, where MNs with lower traffic priority are forced to re-associate with neighboring APs to accommodate MNs with higher priority traffic (called the class aware load balancing and context aware handover).

The proposed scheme assumes that there exists sufficient overlapping among the BSSs, such that every MN has the option to connect with more than one AP. This assumption is based on existing analysis of WiFi HotSpots and urban wireless deployments [72, 73], which shows that APs are deployed in a highly dense manner to

ensure sufficient coverage and to provide multiple connection options to the end users. However, this assumption may be relaxed at the boundary areas of a HotSpot, where a MN may not have multiple connection opportunities. Nevertheless, those areas are beyond the control of the proposed management scheme, and therefore lower priority connections are dropped if they can not be accommodated to assure SLA. This section theoretically models the performance of an IEEE 802.11 ESS consisting of several overlapped BSSs, to show that a network management scheme with above three policies significantly improves the QoS and QoE of the MNs. Based on these policies, the detailed management scheme has been discussed in the next section, along with implementation details.

4.1.1 Network Model

In this section, we consider a network with M BSSs, where each BSS is numbered from $m = 0, 1, \dots, M - 1$. MNs arriving at each BSS are partitioned into r separate classes based on their SLA. To discretize the bandwidth analysis, we assume that the total bandwidth is divided into small sub-channels, and the bandwidth reservation is expressed in terms of numbers of sub-channels occupied by a MN. In a BSS, there can be two types of MN arrivals - new connection arrival, and handover connection arrival. In the proposed management scheme, these two types of arrivals are treated differently according to the bandwidth flexibility strategy. Let c_i denote a new connection, and c_i^+ denote a handover connection of traffic class i . As these two connections are treated separately with different bandwidth reservation policies (cutoff bandwidth and flexible bandwidth), there are in effect $2r$ number of traffic classes in a system. Each class $i = 0, 1, \dots, 2r - 1$ of subscriber requires a minimum bandwidth of b_i , in terms of numbers of sub-channels.

The bandwidth occupancy in a BSS for a multi-class traffic system can be modeled as a Markov state diagram, as shown in Figure 4.1 for a two class, eight channels system (the total bandwidth is divided into eight discrete channels). Every state is represented as (i, b_i) where i denotes the traffic class, and b_i denotes the total bandwidth occupancy in terms of channels for all the MNs with traffic class $i \in [0, 2r - 1]$. In the Markov state diagram, the lower states ($i \in [0, 1]$) denote the states of a new connection, and the upper states ($i \in [2, 3]$) denote the state of a handover connection. As indicated, there are two bandwidth thresholds, the cutoff bandwidth R_i where $i \in [0, r - 1]$ (for a new connection), and the flexible bandwidth U_i where $i \in [r, 2r - 1]$ (for an existing connection). For a new connection, if the bandwidth occupancy for class i exceeds R_i , the connections are blocked. For an existing handover connection with traffic class i , if the bandwidth occupancy exceeds U_i , the connections are dropped. The actual implementation of R_i and

4.1 Theoretical Foundation

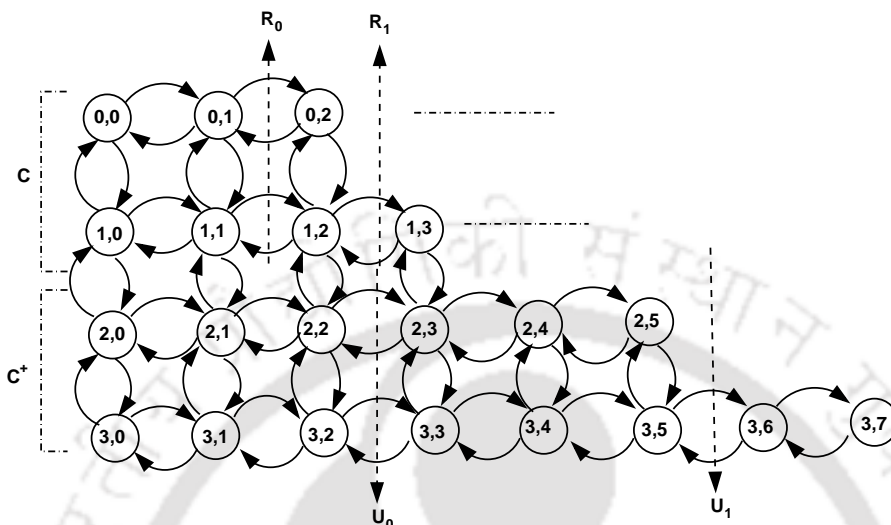


Figure 4.1: Markov state diagram for bandwidth occupancy in a BSS, with two classes in eight channels

U_i in the context of IEEE 802.11e service differentiation is discussed later in Section 4.2.

Let us assume that the new connection arrival rate for each class i in a BSS m is Poisson [72] with rate λ_{im}^N . The rate of handover from BSS j to BSS k for class i traffic is denoted as α_i^{jk} . The rate of handover has two components, the handover due to mobility and the handover due to load balancing. Let \mathbb{N}_m denote the set of BSSs that overlap with BSS m , and λ_{im} denote the total connection arrival rate for class i at BSS m . Then,

$$\lambda_{im} = \lambda_{im}^N + \sum_{j \in \mathbb{N}_m} \alpha_i^{jm} \quad (4.1)$$

It is conceivable that different classes of connections have different holding times (time to use the channels), each class of traffic has an exponential connection holding time distribution [72, 76] with mean $1/\mu_i$. Let the time a subscriber of class i spends at BSS m is exponentially distributed with mean $1/\sigma_{im}$. Then,

$$\sigma_{im} = \mu_i \sum_{j \in \mathbb{N}_m} \alpha_i^{jm} \quad (4.2)$$

4.1.2 Analysis of Connection Blocking Probability, Connection Dropping Probability and Bandwidth Occupancy

Let the steady state vector for the number of subscribers n_i for each service class i in a cell be $\boldsymbol{\pi}(\mathbf{n})$, where $\mathbf{n} = (n_0, n_1, \dots, n_{2r-1})$ where $\{0 \dots r-1\}$ denotes the new connections,

and $\{r \dots 2r - 1\}$ denotes the handover connections, that is $c_i^+ = c_{r+i}$. Let BW be the total bandwidth. Then, $\sum_{i=0}^{2r-1} n_i b_i \leq BW$. For the ease of analysis, initially we assume that the handover occurs only due to the load balancing. In the next subsection, this assumption is relaxed with the incorporation of a mobility model, where the handover occurs due to load balancing as well as general mobility of the MNs. Let ρ_{im} denote the utilization in terms of offered load at BSS m for service class i , then,

$$\rho_{im} = \frac{\lambda_{im}}{\sigma_{im}} \quad (4.3)$$

In the subsequent analysis, the parameters are calculated for a target BSS m , and without the loss in generality, we drop the subscript m for making the notation simple. Henceforth, ρ_{im} is represented as ρ_i , and similar convention is followed for other notations. Depending on the threshold for bandwidth reservation (based on cutoff bandwidth and flexible bandwidth), we define the policy functions for the utilization in terms of offered load at BSS m for service class i . Let us assume that whenever a MN reaches to a state above the cutoff bandwidth threshold R_i , for $i \in [0, r - 1]$, the new connections of service class $i \in [0, r - 1]$ are blocked with a probability $(1 - w_i^R)$. Similarly, whenever a MN reaches to a state above the flexible bandwidth threshold U_i , for service class $i \in [r, 2r - 1]$, the existing connections of service class $i \in [r, 2r - 1]$ are dropped with probability $(1 - w_i^U)$. Accordingly, the policy function ρ_i^+ is defined as follows.

$$\rho_i^+ = \begin{cases} w_i^R \rho_i & \text{if } n_i b_i > R_i \text{ and } i \in [0, r - 1] \\ w_i^U \rho_i & \text{if } n_i b_i > U_i \text{ and } i \in [r, 2r - 1] \\ \rho_i & \text{otherwise} \end{cases} \quad (4.4)$$

Using Equation (4.4), the steady state probability mass function for the Markov state diagram can be calculated using standard formula as given in [77] as,

$$\pi(\mathbf{n}) = \frac{1}{Z} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i}}{n_i!} \quad (4.5)$$

where the normalization constant Z is written as,

$$Z = \sum_{\forall \mathbf{n}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i}}{n_i!}$$

In order to derive connection blocking probabilities for service class i , we define θ_i^{R+} as the set of population vectors \mathbf{n} such that,

4.1 Theoretical Foundation

$$\theta_i^{R+} = \{\mathbf{n} | R_i \leq n_i b_i < BW; i \in [0, r-1]\} \quad (4.6)$$

The connection blocking probability, denoted as P_{B_i} is expressed as,

$$P_{B_i} = \frac{1}{Z} \sum_{\mathbf{n} \in \theta_i^{R+}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i}}{n_i!} + \frac{1}{Z} \sum_{\mathbf{n} \in \theta_i^{R+}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i} (1 - w_i^R)}{n_i!} \quad (4.7)$$

Existing connections of class i are dropped when the state of a MN reaches above the flexible bandwidth threshold U_i . In order to derive existing connection dropping probabilities for service class i , we define θ_i^{U+} as the set of population vectors \mathbf{n} such that,

$$\theta_i^{U+} = \{\mathbf{n} | U_i \leq n_i b_i < BW; i \in [r, 2r-1]\} \quad (4.8)$$

The connection dropping probability for the existing connections of service class i , denoted as P_{D_i} is expressed as,

$$P_{D_i} = \frac{1}{Z} \sum_{\mathbf{n} \in \theta_i^{U+}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i}}{n_i!} + \frac{1}{Z} \sum_{\mathbf{n} \in \theta_i^{U+}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i} (1 - w_i^U)}{n_i!} \quad (4.9)$$

Similarly, the bandwidth occupancy \bar{B} is expressed as,

$$\bar{B} = \frac{1}{Z} \sum_{\forall \mathbf{n}} \prod_{i=0}^{2r-1} \frac{\rho_i^{+n_i}}{n_i!} n_i b_i \quad (4.10)$$

4.1.3 Extension with a General Mobility Model

As discussed earlier, the total handover traffic has two components, the handover due to mobility and the handover due to load balancing. The probability of handover due to load balancing is equal to the probability of a connection of class i to successfully complete a handoff from BSS j to BSS k (denoted as $[Pr_i]_{jk}$) without any mobility constraint, given as,

$$[Pr_i]_{jk} = \frac{\alpha_i^{jk} (1 - P_{B_i})}{\sigma_{ij}} \quad (4.11)$$

To incorporate the mobility model, we use Erlang fixed point approximation with multiple classes. The mobility probability matrices for each class i , denoted as $[\Phi_i]_{jk}$ from BSS j to BSS k , is defined as the probability of a MN to move from BSS j to BSS k , according to a specific mobility model. Any well-defined mobility model, like fixed

way-point mobility model, highway mobility model, fixed direction random acceleration mobility model and so on, can be used in this context.

Essentially, the total handover probability is equivalent to the summation of the probability due to load balancing and the mobility probability. We also define another matrix $G_i(x)$ for traffic class i at iteration x , where the diagonal elements are defines as follows;

$$[G_i(x)]_{mm} = 1 - P_{D_i}(x) \quad (4.12)$$

where $P_{D_i}(x)$ is the connection dropping probability calculated using Equation (4.9) at iteration x . The probabilities are calculated for class i at every BSS $m = 0, 1, \dots, M-1$. As discussed earlier, for every class i , the connection arrival rate vector has two components, the new connections, denoted as Λ_{Ni} , and the handover connections arrival rate denoted as Λ_{Hi} . Λ_{Ni} is expressed as,

$$\Lambda_{Ni} = [\lambda_{i0}^N, \lambda_{i1}^N, \dots, \lambda_{i(M-1)}^N] \quad (4.13)$$

Λ_{Hi} is calculated iteratively. At iteration x , $\Lambda_{Hi}(x)$ is expressed as,

$$\Lambda_{Hi}(x) = \Lambda_{Ni} \Lambda_{Hi}(x-1) [G_i(x-1) (\Phi_i + Pr_i)] \quad (4.14)$$

such that,

$$\Lambda_{Hi}(0) = [0, 0, \dots, 0]$$

Therefore the starting point of the Erlang fixed point approximation is the assumptions of new connections only and the rate of handover connections into each cell is zero. At every iteration x , the matrix $G_i(x)$ is calculated based on the Λ_{Ni} and $\Lambda_{Hi}(x)$ obtained by the fixed point approximation. The λ_{im} values obtained using this approximation is used to calculate the steady state probability mass function given is Equation (4.5). The fixed point approximation is terminated based on the bound on the normalization constant Z at iteration x , denoted as $Z(x)$, as follows,

$$\left| \frac{Z(x)}{Z(x-1)} - 1.0 \right| < \epsilon \quad (4.15)$$

where ϵ is a small predefined threshold value.

4.1 Theoretical Foundation

Table 4.1: Parameters for numerical analysis

| Parameter | Value | Parameter | Value |
|-----------|----------|-----------|-----------------|
| M | 20 | r | 4 |
| BW | 48 Mbps | Mobility | Random Waypoint |
| b_0 | 64 Kbps | b_1 | 128 Kbps |
| b_2 | 256 Kbps | b_3 | 512 Kbps |
| n_0 | 50 | n_1 | 50 |
| n_2 | 100 | n_3 | 100 |
| R_0 | 4 Mbps | R_1 | 16 Mbps |
| R_2 | 8 Mbps | R_3 | 0 |
| U_0 | 6 Mbps | U_1 | 20 Mbps |
| U_2 | 12 Mbps | U_3 | 0 |

4.1.4 Numerical Results: Analysis of the Bandwidth Reservation Strategies

In this subsection, we analyze the effectiveness of the three policies, cutoff bandwidth, bandwidth flexibility and context aware handover along with class aware load balancing, through the numerical results obtained from the theoretical model. The QoS of the network can be effectively captured through the average bandwidth occupancy, while the QoE to the end users can be analyzed through the connection blocking and connection dropping probabilities. The numerical values used in different parameters are shown in Table 4.1. These values are selected keeping in mind the IEEE 802.11e classes of services. The mobility model is considered as random waypoint mobility [78], and the probability matrix $[\Phi_i]_{jk}$ is calculated based on the stochastic model described in [78].

The numerical results are compared for two different scenarios - the first scenario is the proposed management scheme with the three policies (cutoff bandwidth, bandwidth flexibility and context aware handover along with class aware load balancing), and the second scenario is the standard handover without these three policies. The numerical results for the standard handoff is calculated through R_i and U_i set to zero for all classes i , and assuming that handover occurs only due to the mobility, that is $Pr_i = 0$. In the standard handover, a new connection of class i is blocked, or an existing connection of class i is dropped, if it does not receive the minimum bandwidth b_i , according to the SLA. The numerical results are calculated with respect to the mobility acceleration seed value that varies between 0 and 1. As the mobility acceleration seed increases, the MN becomes

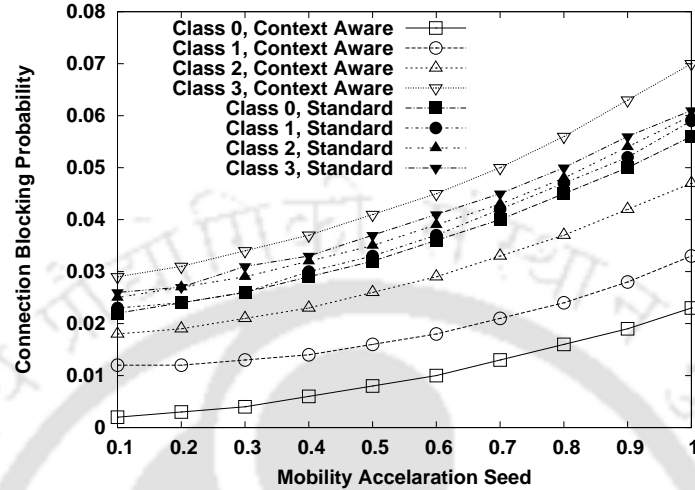


Figure 4.2: Connection Blocking Probability

more mobile in a random waypoint mobility [78].

The numerical results are shown in Figure 4.2 to Figure 4.4. In the figures, the term ‘standard’ indicates the standard handover procedure and the term ‘context aware’ indicates the handover procedure with the proposed bandwidth reservation policy. Figure 4.2 shows the numerical results in terms of connection blocking probability. As the mobility increases, connection blocking probability gets increased, because the number of handover due to mobility gets increased, that blocks new connections. However, with the bandwidth reservation policies as discussed in this section, the differentiation among different classes of services gets increased. According to the parameter setups as depicted in Table 4.1, class 0 has the maximum priority service in terms of per connection bandwidth requirement and corresponding bandwidth reservation, followed by class 1, class 2 and class 3. Class 3 does not have any bandwidth reservation, while it has maximum traffic demand. Therefore the connection blocking probability for class 3 is the maximum. However, the standard does not provide such differentiations during admission control and handover, and therefore connection blocking probability is almost similar for all the four traffic classes.

Similar result is observed for connection dropping probabilities, as shown in Figure 4.3. The differentiation among different classes of services is prominent with the bandwidth reservation policies, whereas the standard procedure shows almost similar results for all the four service classes. Further as the mobility increases, the connection dropping probability increases for the standard, whereas it increases slowly with the

4.1 Theoretical Foundation

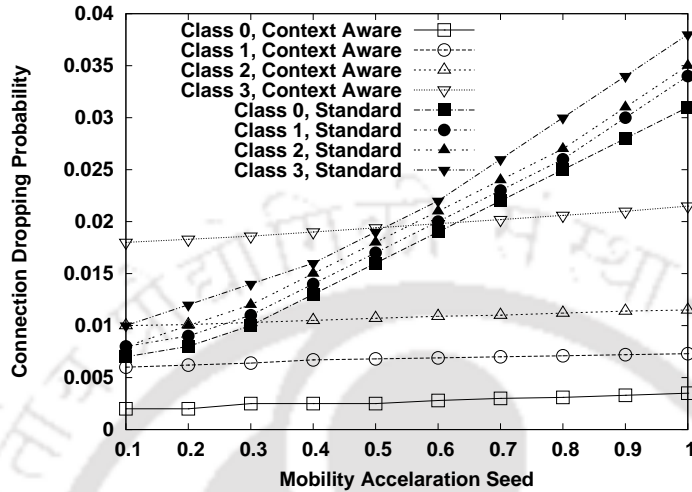


Figure 4.3: Connection Dropping Probability

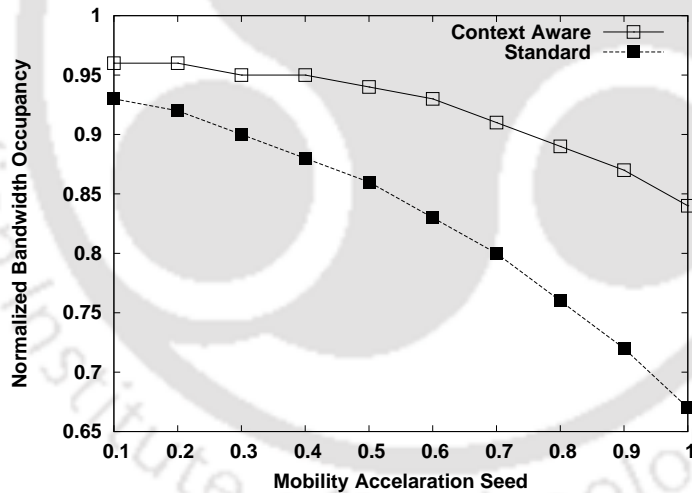


Figure 4.4: Normalized Bandwidth Occupancy

bandwidth reservation policies. As mobility affects the existing connections severely in the standard procedure without any flexible bandwidth reservation, the drops of existing connections get increased irrespective of service classes. This figure clearly reveals the advantage gained with the flexible bandwidth strategy. Figure 4.4 shows the normalized bandwidth occupancy, calculated as \bar{B}/BW . As the bandwidth reservation policies reduce average number of connection blocking and connection dropping and distribute the connections among the BSSs through load balancing, the normalized bandwidth occupancy

gets improved.

The theoretical analysis clearly reveals the advantages gained with the incorporation of bandwidth reservation policies along with the load balancing strategy during inter-BSS handover management. An improvement over the standard IEEE 802.11e and the inter-BSS handover management is proposed in this chapter with the incorporation of the cutoff bandwidth, bandwidth flexibility and the context aware handover along with the class aware load balancing, as discussed in the next section.

4.2 Context Aware Handover Management Sustaining QoS and QoE: The Proposed Scheme

In the proposed scheme, an inter-BSS handover management mechanism is designed that improves the performance of an IEEE 802.11e ESS by balancing load across different APs in the presence of a mix of different EDCA traffic classes. The management scheme has three components: (i) admission control based on cutoff bandwidth and bandwidth flexibility strategy, (ii) class aware load balancing, and (iii) context aware handover.

4.2.1 Admission Control

Admission control is required to limit the traffic load at APs within the available bandwidth, and also to keep provision for the different traffic classes according to the SLA. Admission control is necessary for two scenarios. The first is to manage the QoS of the network and QoE of the existing connections, when a new MN first associates with an AP in the ESS. Let us call this as initial admission control (IAC). Admission control is also needed when a MN performs handover, to assure the QoE of that MN as well as other MNs which are already associated with the AP to which the MN performs handover. Let us call this as handover admission control (HAC).

Initial Admission Control

In IAC, a fixed percentage of the maximum bandwidth is reserved at the APs for traffic classes AC_VO, AC_VI and AC_BK, according to the cutoff bandwidth strategy. Let the cutoff bandwidth for AC_VO, AC_VI and AC_BK be R_{vo} , R_{vi} , and R_{bk} respectively. The cutoff bandwidth is determined by the service providers based on the average number of subscribed users that want to avail services for a specific service class. If the total

4.2 Context Aware Handover Management Sustaining QoS and QoE

bandwidth is BW then it is required that,

$$R_{vo} + R_{vi} + R_{bk} < BW. \quad (4.16)$$

The amount of spare bandwidth S , calculated using Equation (4.17), is required to keep provision for occasional traffic bursts of the connections in the three classes.

$$S = BW - R_{vo} - R_{vi} - R_{bk} \quad (4.17)$$

A new connection of class c is admitted into an AP only if the total bandwidth occupied by all the traffics for that specific service class does not exceed the cutoff bandwidth for that class. Let a new connection of class c having a bandwidth requirement of b_c needs to be admitted at AP_k . It is admitted only if it satisfies Equation (4.18),

$$b_c + O_c^k \leq R_c \quad (4.18)$$

where R_c is the cutoff bandwidth for class c , and O_c^k is the bandwidth already occupied by all the traffics of class c associated with AP_k .

It can be noted that no reservation is made for the AC_BE class. An AC_BE class traffic is admitted if there is any bandwidth available irrespective of the traffic classes. A connection of traffic class AC_BE is admitted to AP_k , if it satisfies Equation (4.19).

$$BW \geq O_{vo}^k + O_{vi}^k + O_{bk}^k + O_{be}^k \quad (4.19)$$

Where O_{vo}^k , O_{vi}^k , O_{bk}^k and O_{be}^k are the bandwidth occupied by connections of class AC_VO, AC_VI, AC_BK and AC_BE, respectively, at AP_k . One or more connections of AC_BE class are dropped any time if a new connection of some non-BE class needs to be admitted, and the new connection satisfies Equation (4.18), but Equation (4.19) is not satisfied. In such case, AC_BE connections are dropped until Equation (4.19) gets satisfied.

Handover Admission Control

In case of HAC, to let ongoing communications to continue, subject to certain conditions that follows, a connection may be admitted even if it violates Equation (4.18). We define upper limit of reservations according to the bandwidth flexibility strategy, called the flexible bandwidth. Let the flexible bandwidths for the classes AC_VO, AC_VI and AC_BK be U_{vo} , U_{vi} and U_{bk} respectively. For a class c having flexible bandwidth U_c , U_c must be greater than the cutoff bandwidth R_c . For a connection of class c with a bandwidth requirement of b_c , is admitted in AP_k , if Equation (4.20) is satisfied.

$$b_c + O_c^k \leq U_c \quad (4.20)$$

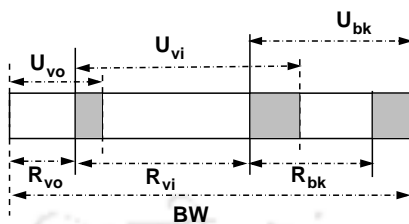


Figure 4.5: Bandwidth reservations for admission control management

To understand the relation between U_c and R_c , let us consider Figure 4.5. The cutoff bandwidth R_c for a traffic class c shows the amount of bandwidth that can be utilized for new connections as well as handover connections. However, to keep extra bandwidth provisioning for handover connections, the flexible bandwidth U_c is used. The figure shows a shaded region for each traffic class that indicates the difference between the cutoff bandwidth and the flexible bandwidth ($U_c - R_c$), that is used as the extra bandwidth provisioning during handover among neighboring APs.

It may so happen that there is no available bandwidth in the AP to admit the handing over MN. In that case, the AP tries to shift some load to adjacent APs by handing over some MNs, as described later in subsection 4.2.2. If it is not possible to adjust the load, the AP starts disassociating MNs in the following way, till Equation (4.19) gets satisfied. Let the incoming MN have connections of traffic class c . The disassociation is based on following rules.

1. If c is of type AC_VO, then first AC_BE, then AC_BK and finally AC_VI traffics are dropped.
2. If c is of type AC_VI, then first AC_BE and then AC_BK traffics are dropped.
3. If c is of type AC_BK then only AC_BE traffics are dropped.

If it is not possible to satisfy Equation (4.19) by dropping connections of classes lower than c , then the incoming MN is not admitted. A MN with AC_BE class traffic is admitted only if there is enough bandwidth available for admitting the MN.

4.2.2 Class Aware Load Balancing

An AP may get overloaded if it has to accommodate some MNs that are performing handover from neighboring APs. Such a MN might have been forced to handover due to poor RSSI (or SINR) in the old AP. As the MN is already admitted into the ESS, to

4.2 Context Aware Handover Management Sustaining QoS and QoE

be fair¹, it is required to accommodate the MN in the target AP, AP_{target} , to which it has requested a re-association due to handover, even when there is no spare bandwidth available in AP_{target} . In the proposed scheme, such a MN is accommodated by shifting some traffic load of AP_{target} to other co-located APs (those have common or overlapping coverage area), so that enough bandwidth becomes available to accommodate the MN. Following steps are performed for class aware load balancing in the proposed scheme,

1. AP_{target} finds the MN with lowest RSSI (or SINR), say MN_{target} and sends *handoverInitiate* message to it.
2. MN_{target} scans for available APs for handover, and builds a list of APs found into a list called *ScannedAPList*. MN_{target} replies to AP_{target} using a message *FindScannedList* containing *SacnnedAPList* to AP_{target} .
3. AP_{target} sends *GetLoad* message to each AP in *SacnnedAPList*, through the distribution system.
4. APs receiving *GetLoad* message reply back with a *LoadReply* message, containing spare capacity information for each traffic class in it.
5. AP_{target} collects replies for some fixed period, and filters out those APs where sufficient spare capacity is not available to accommodate MN_{target} , according to its service class.
6. AP_{target} then sends the resulting list to MN_{target} .
7. MN_{target} chooses the AP with the best RSSI (or SINR) and performs an inter-BSS handover.

The class aware load balancing is executed for every classes except for the AC_BE traffic classes. A proper load metric is required to be designed for the proposed class-aware load balancing mechanism. This has been discussed in following subsection.

4.2.3 Context Aware Handover

In general, inter-BSS handover is triggered whenever a MN can no longer communicate through it's associated AP due to poor signal condition. This may be detected by several

¹The priority is given to the existing connections over the new connections. The existing connections are dropped only when no spare bandwidth is available among the neighboring APs due to admission of MNs with higher class traffics.

ways. Whenever RSSI falls below a certain threshold, or a few beacons are lost, or the packet drop becomes high, MN triggers a handover to an alternate AP having common coverage area. The traditional handover triggering according to the IEEE 802.11 standard has two drawbacks. First, disruption in communication occurs before the actual handover is triggered, and second, the MN continues to associate to an AP even when there is an overload in that AP, and thus suffer from excessive packet loss and delay. In the proposed handover management scheme in this chapter, handover is triggered by excess load as detected by either the AP or the MN, considering the bandwidth availability for the specific service class.

For this purpose, traffic load needs to be measured to trigger handover as well as to perform load balancing. Load can be measured in several ways. Conventionally, the amount of time the wireless channel remains busy reflects the traffic load in a BSS. Derivatives of this conventional approach is used as load metric in several existing works, such as [79] and [20]. In [20] the authors have shown that available bandwidth is reflected by channel busy status that can be obtained from the network allocation vector (NAV) in EDCA. However, the existing load metrics do not consider the service class differentiations, and therefore a new metric is required to be designed.

Load Measurement and Overload Detection

In the proposed scheme, number of bytes sent and received by a MN for a specific service class type over its wireless interface in a unit time is taken as the load metric for that service class. It directly reflects the amount of traffic for a service class flowing through the wireless interfaces, and is simple to measure. The traffic load at MN_m for class c , denoted as $Load_m^c$, at time t_i is calculated as,

$$Load_m^c = \frac{BytesTxRx_c}{(t_i - t_{i-1})} \quad (4.21)$$

where

$$BytesTxRx_c = \sum_{j=0}^{n_c} PacketSize(j_c) \quad (4.22)$$

where, n_c is the number of packets of traffic class c sent and received over the wireless medium by MN_m during the interval t_{i-1} to t_i , and $PacketSize(j_c)$ is the size of the j^{th} packet of traffic class c sent or received after time t_{i-1} , in bytes.

Each node measures the load continuously at a fixed time interval $t_{measure}$. An AP is called overloaded for a traffic class c , if the total traffic load for traffic class c from its

4.2 Context Aware Handover Management Sustaining QoS and QoE

Algorithm 4.1 Interleaved Scanning

- 1: $associatedChannel \leftarrow$ the current channel of association;
 - 2: Stop communicating in $associatedChannel$;
 - 3: Delete all entries from $ScannedAPList$ where the channel number of the AP equals $NextChannel$;
 - 4: Change channel to $NextChannel$;
 - 5: Send a probe request;
 - 6: $tillTime \leftarrow$ current time + t_1 ; { t_1 is a small timeout value}
 - 7: **while** current time < $tillTime$ **do**
 - 8: Add AP informations if any obtained from probe response or beacons received into $ScannedAPList$;
 - 9: **end while**
 - 10: Update $NextChannel$ to the cyclically next to $NextChannel$ entry in the channel list;
 - 11: Change channel to the $associatedChannel$;
 - 12: Restart $ScanRepeatTimer$
-

associated MNs reaches closer to U_c . During load balancing, the handover is initiated after detecting traffic overload for that specific service class.

Another aspect of the context aware handover procedure proposed in this chapter is to minimize the scanning time during the search for the alternate APs having common coverage area. This is discussed in the following subsection.

Interleaved Scanning

In the proposed scheme, the scanning and the data communication are interleaved in time. The scanning technique used in this chapter is a customized version of the scanning technique proposed in the Chapter 3. The MN keeps the information of different APs found during scanning in a list, called the $ScannedAPList$. When the MN senses degradation of performance, it prepares a circular list of channel numbers to scan. These are the channels that contain all the IEEE 802.11 channels except the current channel and the channels that overlap with the current channel. It initializes a variable $NextChannel$ to the channel number of the first channel in the list. It empties $ScannedAPList$. It also starts a timer, called $ScanRepeatTimer$. Let the time out value of this timer be T . At the expiry of time T (when the timer event is triggered), a MN executes Algorithm 4.1.

As the $ScanRepeatTimer$ is restarted, it will be triggered again and Algorithm 4.1 will be executed again. Till that time normal communication continues. The timer is

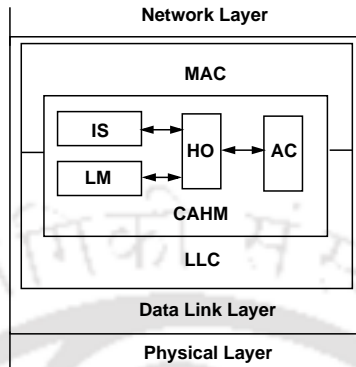


Figure 4.6: Proposed management module in IEEE 802.11 protocol stack

canceled when all the channels in the circular list of channels are scanned. This scanning technique avoids long scanning delays experienced while using the conventional scanning technique.

4.2.4 Implementation of the Context Aware Handover Management Module

The context aware handover management (CAHM) module works cooperatively along with the data link layer in the IEEE 802.11 protocol stack. This management module has four submodules - (i) the interleaved scanning (IS) submodule, (ii) the load measurement (LM) submodule, (iii) the handover (HO) submodule, and (iv) the admission control (AC) submodule. The CAHM module of a MN has all the four submodules, however the CAHM module at an AP consists three submodules except the IS. The inter-working of these submodules is shown in Figure 4.6. The complete working principle of the CAHM is summarized as follows,

- The IS submodule at a MN periodically scans the channels using the interleaved scanning mechanism, as discussed in Subsection 4.2.3, to construct the set of APs within the vicinity.
- The LM submodule measures the traffic load metric at the APs and the MNs, using the procedure discussed in Subsection 4.2.3.
- During the initial connection establishments as well as handover, the AC submodules (discussed in Subsection 4.2.1) of the AP and the MN execute the admission control strategies cooperatively, as discussed earlier.

4.2 Context Aware Handover Management Sustaining QoS and QoE

- The HO submodule takes care of the handover decisions, as discussed in Subsection 4.2.2 and Subsection 4.2.3. Whenever the LM submodule of an AP detects the overloaded condition, the HO submodules at the AP and the MN execute the context aware handover procedure with the help of the neighboring AP information obtained through the IS submodule of the MN.

4.2.5 Selection of Cutoff Bandwidth and Flexible Bandwidth

In the proposed context aware handover management scheme, admission control and bandwidth reservation are controlled by the cutoff bandwidth and the flexible bandwidth, as discussed earlier. A higher cutoff bandwidth, R_c for traffic class c indicates that more number of connections of traffic class c can be supported in the network without QoS and QoE degradation. Similarly, a higher flexible bandwidth U_c for traffic class c indicates that the connections of traffic class c are more sustainable during mobility and handover to maintain minimum QoS and QoE guarantee. However, as the total available bandwidth is constant for a network, selection of R_c and U_c depends on the choice of how much connections of specific traffic classes can be supported by a service provider, and therefore, it is completely policy indicative. Individual service providers can select cutoff bandwidth and flexible bandwidth for their networks based on the pricing and connection support policies. A number of such policy related issues has been discussed in the literature, such as [80] and the references therein, and similar policy can be adopted for the selection of cutoff and flexible bandwidths.

4.2.6 Discussion: MNs at the Hotspot Boundary

An assumption of the proposed context aware handover scheme is that there exist sufficient overlapping among the BSSs to ensure that the MNs have multiple connection opportunities. Though this assumption is based on the existing analysis of the wireless HotSpots characteristics and the urban wireless deployment fundamentals [72, 73], the assumption may get violated at the boundary of a HotSpot, where a MN may not get multiple connection opportunities. In the proposed scheme, the admission control and the handover modules work in a cooperative manner. Further, the handover decision is a coordinating decision between the AP and a MN, and the AP instructs for an impending handover to a MN only when it has multiple feasible connection opportunities. During the admission control, an incoming connection is blocked only when neither sufficient bandwidth is available, nor required bandwidth can be accumulated through context aware

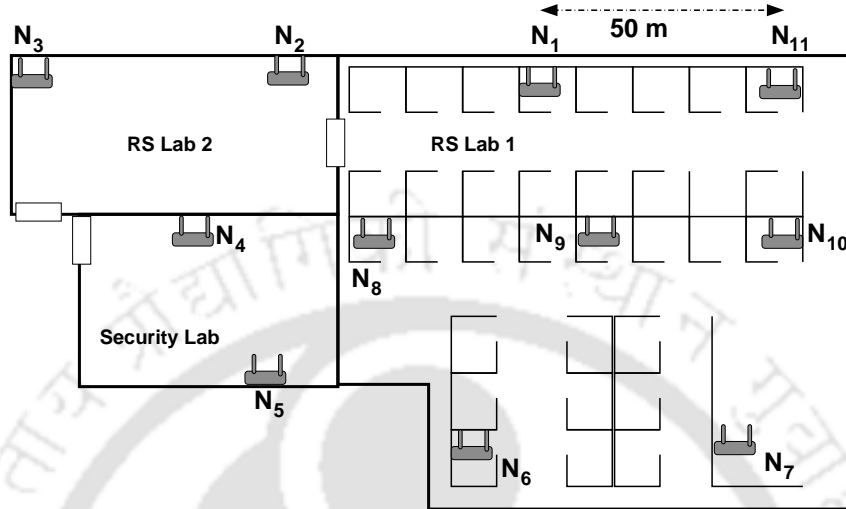


Figure 4.7: IEEE 802.11 Indoor Wireless Testbed

handover based on load balancing. This way, the proposed scheme takes care of the MNs that are at the HotSpot boundary.

4.3 Experimental Evaluation and Comparison

The performance of the proposed context aware handover mechanism in IEEE 802.11e ESS is evaluated using an 11-node IEEE 802.11g+e indoor wireless LAN testbed deployed over the IIT Guwahati computer science department research labs, as shown in Figure 4.7. Each wireless router is a Skiva Easyconnect RT001 N300 WiFi router with RaLink RT-3352 chipset [69]. The Ralink RT-3352 router on chip combines 802.11b/g/n draft compliant 2T2R MAC along with BBP/PA/RF MIMO, a high performance 400MHz MIPS24KEc CPU core, a Gigabit Ethernet MAC, 5-ports integrated 10/100 Ethernet Switch/PHY, 64MB of SDRAM and 32MB of Flash. This chip can support up to 54 Mbps data rate with maximum transmission power of $16dBm$, using IEEE 802.11g physical layer technology. The RT-3352 wireless chipset is equipped with Linux Kernel 2.6.14 along with IEEE 802.11e protocol stack. The proposed CAHM module is implemented as a loadable kernel module (LKM) in the IEEE 802.11e protocol stack.

4.3.1 Testbed Setup

Every wireless router is connected with the wired distribution system through a Gigabit Ethernet LAN, and we assure that the wired network has sufficient bandwidth available,

4.3 Experimental Evaluation and Comparison

so that it does not affect the wireless performance. MNs are randomly distributed in the testbed area. The distribution of MNs are specified later based on the experimental setups. Different classes of traffics are generated using Seagull Multi-protocol Traffic Generator [81]. Voice over IP (VoIP) traffic is used for AC_VO, video streaming is used for AC_VI, trivial file transfer protocol (TFTP) with UDP traffic is used for AC_BK and FTP with TCP traffic is used for AC_BE. Table 4.2 shows the testbed parameters for different classes of services. N_c and Q_c denote the mean number of connections per AP² and minimum traffic demand according to SLA, respectively, for traffic class c . It can be noted that there is no specific traffic demand for best effort traffic connections, as those are elastic in nature. The mean number of connections given in the table is only indicative, and applies only when there is no specific mentions. The cutoff bandwidth and the flexible bandwidth are also indicative, that supports a maximum of 64, 32 and 64 numbers of voice, video and background traffics, respectively, in the network with minimum bandwidth guarantee according to the SLA. In the testbed evaluation, a MN may generate connections of more than one types of service classes. For example, an user may watch video streaming while uploading a file through FTP transfer. In this case the users generates connections of service class types AC_VI and AC_BE simultaneously. In this scenario, a MN is characterized by the connection with the traffic class of the highest priority that it generates. The connections with lower priority traffics are admitted only if sufficient bandwidth is available at the concerned AP for those traffic classes. As in this mechanism, a MN is characterized by a connection of a specific traffic class, the terms MN, traffic and connection are used interchangeably in this section. In the testbed emulation, we have considered random waypoint mobility with average speed 2 m/s. 30% of the users are mobile, where rest others are static users.

The QoS of the network is evaluated through the bandwidth occupancy for every service classes, the network utilization, the packet loss rates for AC_VO and AC_VI, the end-to-end forwarding delay and the jitter for AC_VO traffics. The QoE of a voice connection is measured in terms of ‘Mean Opinion Score’ (MOS) [82]. The value of MOS varies between 1 to 5, where 1 denotes the worst quality, and 5 denotes the best quality. On the other hand, the QoE for video streaming is measured in terms of ‘Peak Signal to Noise Ratio’ (PSNR) and ‘Structural Similarity Index’ (SSIM) [83]. These parameters

²Mean number of connections per AP does not indicate actual number of connections associated with an AP. We have clustered the MNs based on their proximity to an AP. The mean number of connections per AP denotes the average number of connections generated from the APs (which are closer to a the target AP). In the standard association procedure, most of these connections associate with that AP, though this is not the case in the proposed approach.

4.3 Experimental Evaluation and Comparison

Table 4.2: Testbed Parameters

| Parameter | Value | Parameter | Value |
|-----------|---------|-----------|----------|
| R_{vo} | 4 Mbps | U_{vo} | 6 Mbps |
| R_{vi} | 16 Mbps | U_{vi} | 20 Mbps |
| R_{bk} | 8 Mbps | U_{bk} | 10 Mbps |
| R_{be} | 0 Mbps | U_{be} | 0 Mbps |
| N_{vo} | 20 | Q_{vo} | 64 Kbps |
| N_{vd} | 20 | Q_{vd} | 512 Kbps |
| N_{bk} | 25 | Q_{bk} | 128 Kbps |
| N_{be} | 20 | Q_{be} | Elastic |

measure QoE for video streaming using frame by frame analysis, and higher values of PSNR and SSIM denote better quality of video streaming. The QoE for AC_BK traffic is measured using end-to-end throughput.

4.3.2 Evaluation of QoS Parameters

For these set of experiments number of connections of traffic class AC_BE is increased gradually in the network to evaluate the performance. Because of the elastic nature of the AC_BE connections with TCP traffics, the QoS associated traffic classes, like AC_VO and AC_VI, are affected most. The experiments are conducted for six hours, and the average is taken to plot the graphs.

Figure 4.8 shows the average bandwidth occupancy per AP with respect to the mean number of AC_BE connections per AP. In this experiment, as well subsequent experiments, the deviation in number of AC_BE connections is kept quite high (on average 20), so that the traffic concentration scenario, similar to the public wireless HotSpots, can be emulated. In the standard procedure, most of the connections associate with a single AP. Therefore, as the number of AC_BE connections increases in the proximity of an AP, the QoS associated traffic classes get affected. The figure reveals that as the number of AC_BE connections increases in the network, the bandwidth occupancies for the AC_VI, AC_VO and AC_BK get decreased. It has been observed from the emulation traces that when the number of AC_BE connections increases beyond 30, the minimum traffic demands for the QoS associated classes get violated. On the contrary, the proposed scheme reserves bandwidth (cutoff and flexible) for the QoS associated classes. Further the load balancing procedure distributes traffic loads among the overlapping APs. Therefore, increasing the

4.3 Experimental Evaluation and Comparison

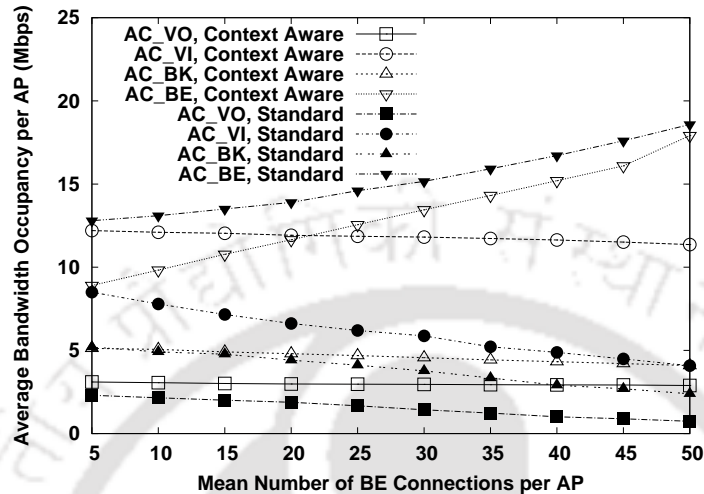


Figure 4.8: Bandwidth Occupancy

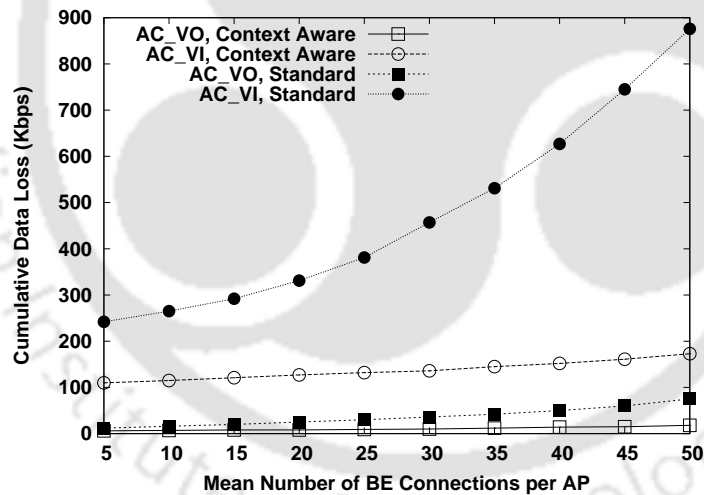


Figure 4.9: Data Loss for QoS Traffic

number of AC_BE traffics does not affect the performance of other traffic classes with higher priority. As a consequence, the bandwidth occupancies for the AC_VO, AC_VI and AC_BK traffics remain almost constant.

One of the important QoS parameters for AC_VO and AC_VI traffics is the data loss. Voice communications as well as video streaming can not sustain if packet (or data) loss is very high. Figure 4.9 compares the proposed context aware handoff scheme and the standard procedure with respect to the cumulative data loss (the total data loss of all

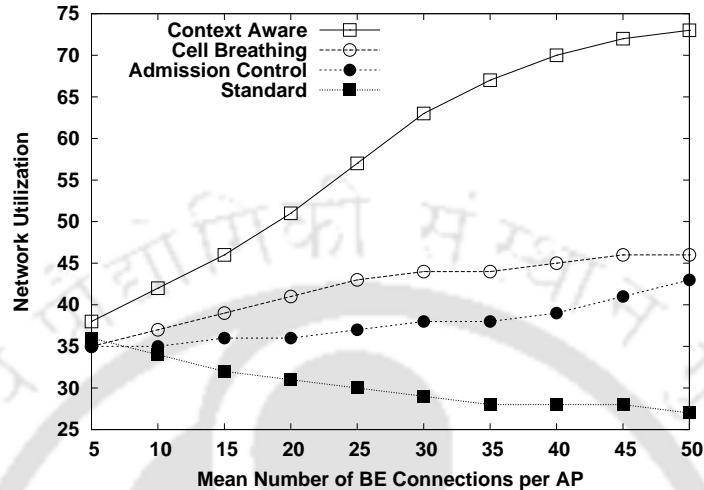


Figure 4.10: Network Utilization

the connections for a specific service class). The standard shows exponential increase in the packet loss as the number of elastic AC_BE connections increases in the network. On the contrary, the proposed scheme bounds the packet loss within tolerable limits (5 Kbps per connection for video traffics, and 20 bps per connection for voice traffics). From the testbed emulation traces, it has been revealed that the bandwidth reservation strategy significantly reduces packet losses. Again with large number of connections, the load balancing strategy further reduces packet losses by decreasing network congestion through the proper distribution of connections among the overlapping APs.

Figure 4.10 shows the network utilization with respect to the mean number of AC_BE connections per AP. We have compared network utilization as well as subsequent QoS parameters with two other schemes proposed in the literature - the cell breathing technique [19] for load balancing, and the admission control mechanism proposed in [84] for IEEE 802.11e. The network utilization is defined as the percentage of available bandwidth used by the existing connections. The figure indicates that, as the mean number of AC_BE traffics increases, the network utilization gets reduced for the standard procedure. However, the network utilization significantly increases for the proposed context aware handover procedure with the increase in AC_BE traffics. In case of the standard procedure, most of the connections get associated with the AP that provides maximum signal strength (normally the AP that is more closer). This affects the network utilization, as most of the APs in the network remain underutilized. The network utilization is low for the cell breathing mechanism as it does not distribute the load explicitly in the network, rather

4.3 Experimental Evaluation and Comparison

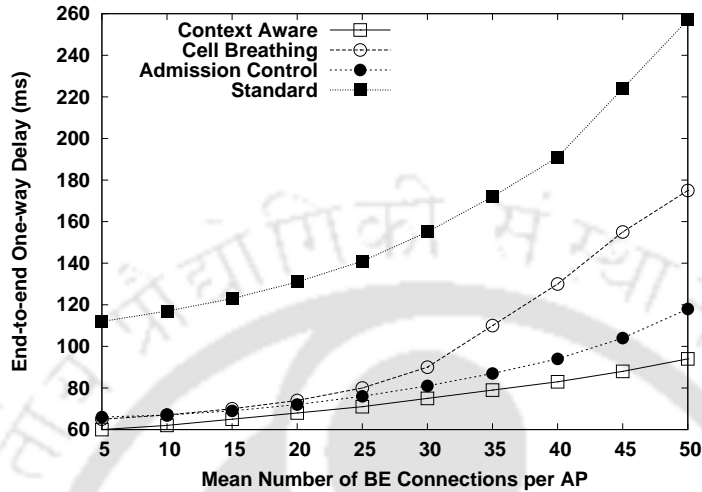


Figure 4.11: End-to-end one-way Forwarding Delay for Voice Traffics

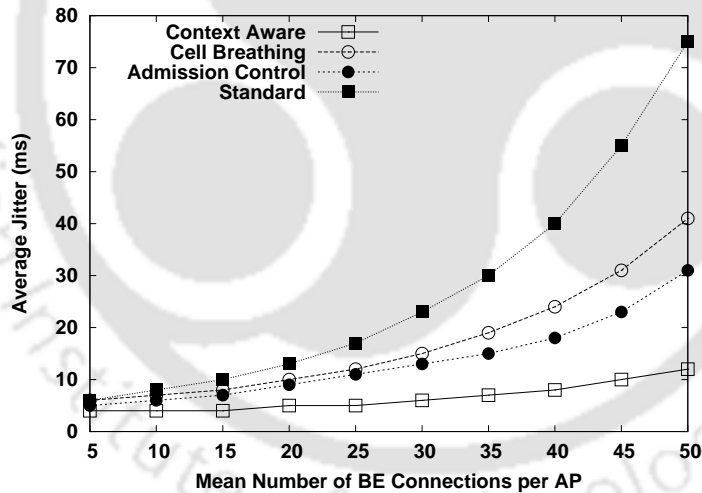


Figure 4.12: Average Jitter for Voice Traffics

every AP only reduces its own load by reducing the BSS coverage area. Further, we have observed that the cell breathing technique even causes coverage hole in the network, when two neighboring APs, both overloaded, reduces their coverage area. The admission control mechanism adjusts traffic load at individual APs, whereas the bandwidth at underloaded APs remains underutilized, resulting in poor network utilization. On the contrary, the proposed context aware handover procedure along with the load balancing distribute the connections among the overlapping APs, that significantly improves network utilization.

End-to-end one-way forwarding delay is an important QoS parameter for voice traffics. Voice traffic can generally sustain for 150 ms one-way end-to-end forwarding delay with 30 ms of jitter. Figure 4.11 compares the four schemes with respect to the one-way forwarding delay, and Figure 4.12 compares them with respect to the average jitter. The forwarding delay and the average jitter for the standard procedure is significantly higher as most of the connections concentrate near a few APs, and associate with those APs which get overloaded, resulting in higher delay in traffic forwarding, as well as introduces high jitter. The cell breathing technique reduces delay when the traffic load is low, however delay gets increased with higher traffic load as the APs control only their own loads without considering neighboring APs. Similar situation is observed for the admission control mechanism. Though the one-way delay and jitter is less for the cell breathing and the admission control techniques compared to the standard, the parameters are quite high in comparison with the proposed scheme. The proposed context aware handover scheme reduces delay and jitter by means of two strategies. First, every AP explicitly reserves bandwidth for the higher priority traffic classes, and second, when overloaded, the traffic load is distributed among the overlapping APs with cooperation of both the APs as well as MNs. Further, the flexible bandwidth strategy allows smooth handover for the higher priority traffics. In this way, the proposed scheme improves QoS parameters for higher priority voice and video traffics.

4.3.3 Evaluation of QoE Parameters

In this subsection, we evaluate the performance of the proposed scheme in the viewpoints of the end users. As mentioned earlier, MOS is used to evaluate the users' satisfiability for voice traffics, whereas PSNR and SSIM are used to measure quality of the video traffics.

Figure 4.13 shows the MOS for the voice traffics, and compares the four schemes with respect to the MOS. The MOS for the standard scheme is very low with high traffic load, as it provides minimum bandwidth guarantee and does not respect minimum delay and jitter requirements for voice traffics. Though the cell breathing and the admission control strategies improve MOS compared to the standard, still they results in lower performance compared to the proposed scheme, because of the poor network utilization that affects the minimum bandwidth guarantee. The proposed scheme significantly improves the MOS for the voice traffics through the effective bandwidth reservation policies during initial admission control as well as handover, and through context aware load balancing to distribute the traffic load while maintaining QoS requirements.

Figure 4.14 and Figure 4.15 show the PSNR and SSIM, respectively, for video

4.3 Experimental Evaluation and Comparison

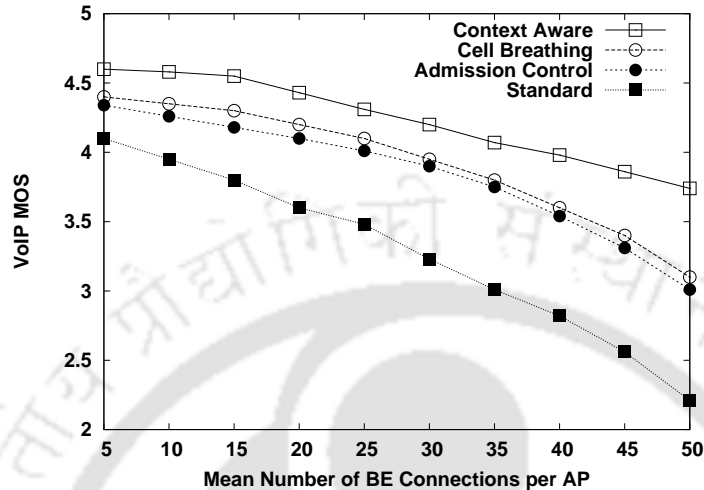


Figure 4.13: MOS for Voice Traffics

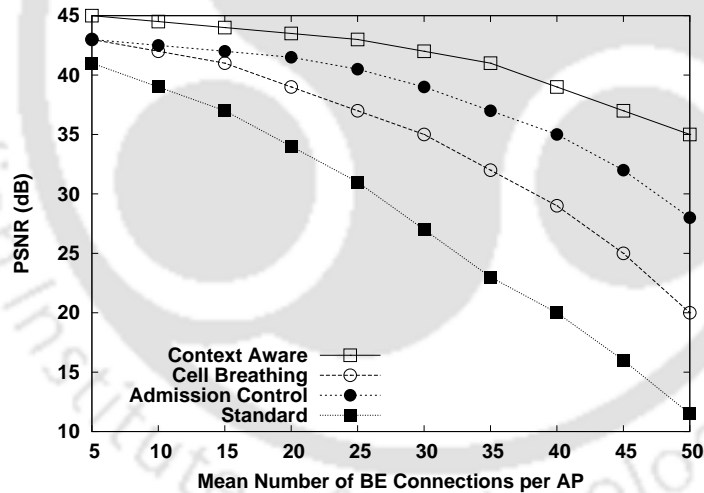


Figure 4.14: PSNR for Video Traffic

traffic with increasing load in the network. The standard scheme results in poor video performance because of lower bandwidth provisioning and higher data loss, as indicated in earlier results. It can be observed that for video traffics, the admission control mechanism provides better performance than the cell breathing. The reason is that though the cell breathing technique controls total traffic load of an AP, it fails to explicitly reserve bandwidth for video traffics. Therefore, the best effort traffics affect the performance of the video traffics. The proposed context aware handover management mechanism provides

4.3 Experimental Evaluation and Comparison

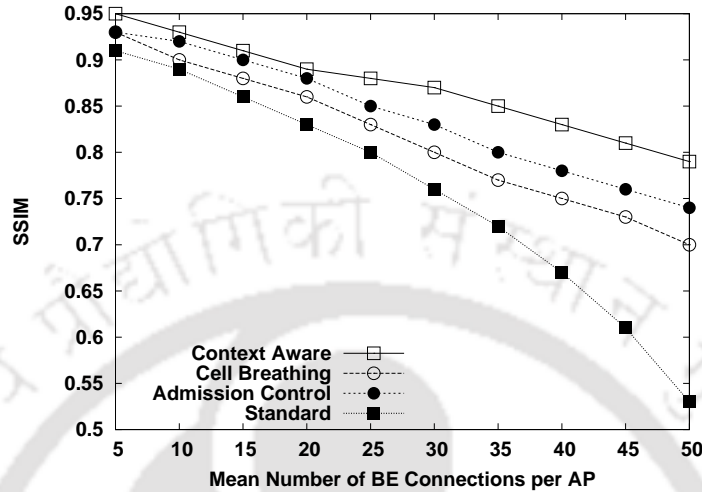


Figure 4.15: SSIM for Video Traffic

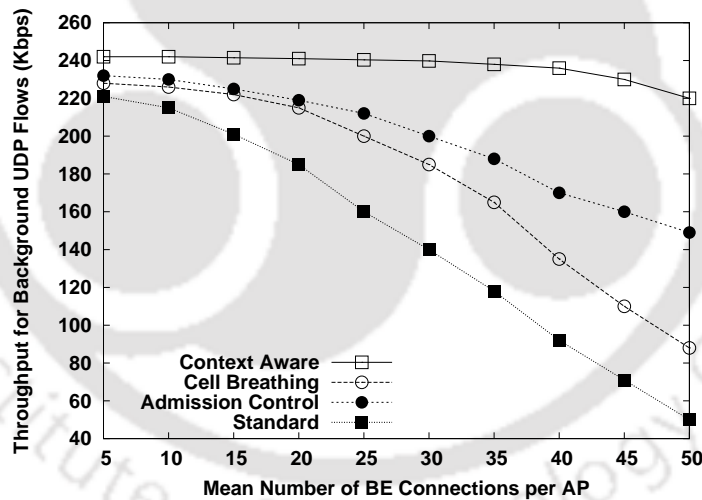


Figure 4.16: Throughput for AC_BK Traffic

sufficient bandwidth to the video traffics, while reduces data losses by distributing traffics among the neighboring APs (reduces packet drops due to congestion). Therefore significant improvement is observed for the performance of video traffics with the proposed scheme, for both the QoE metrics.

To analyze the performance of AC_BK traffics, Figure 4.16 compares the background UDP flow throughputs (per flow throughput) for the four methods. In the standard procedure, the throughput for the background connections drops significantly with the

4.3 Experimental Evaluation and Comparison

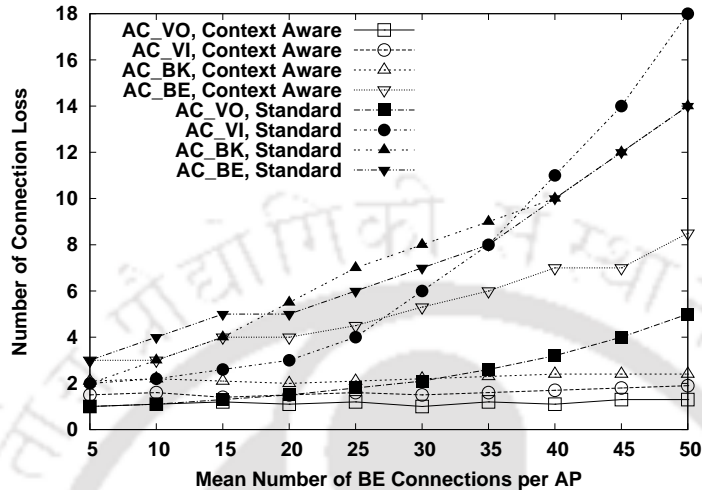


Figure 4.17: Connection Losses (Block + Drop)

increase in the elastic AC_BE traffics. It can be observed that the throughput even drops below 128 Kbps, the minimum traffic demand in our testbed setup. Though the cell breathing and the admission control mechanisms improve per flow throughput, the results are less than the proposed scheme. For background traffics, the cell breathing mechanism performs poorly as it does not have control over the number of AC_BE connections associated with an AP (the load balancing is not class aware). The admission control mechanism only ensures the minimum traffic guarantee (128 Kbps). However the proposed context aware scheme improves the performance even more than the minimum traffic guarantee through load balancing. While the per class bandwidth reservation ensures minimum traffic guarantee, the context aware handover along with class aware load balancing allow the traffics to utilize the available bandwidth effectively.

Figure 4.17 shows average number of connection losses (blocked + dropped) with respect to the increasing number of best effort connections. The results for cell breathing and admission control schemes are not shown, as the cell breathing technique does not explicitly address class aware load balancing, and the results for the admission control scheme follows almost similar to the proposed scheme, through the numeric values are bit higher³. The figure shows that with the increase in the AC_BE connections, the connections in the standard procedure are either blocked or dropped, irrespective of class types. Further, the data loss rate for the video connections become considerably

³In the absence of load balancing, more number of best effort connections are dropped if only admission control is implemented.

4.3 Experimental Evaluation and Comparison

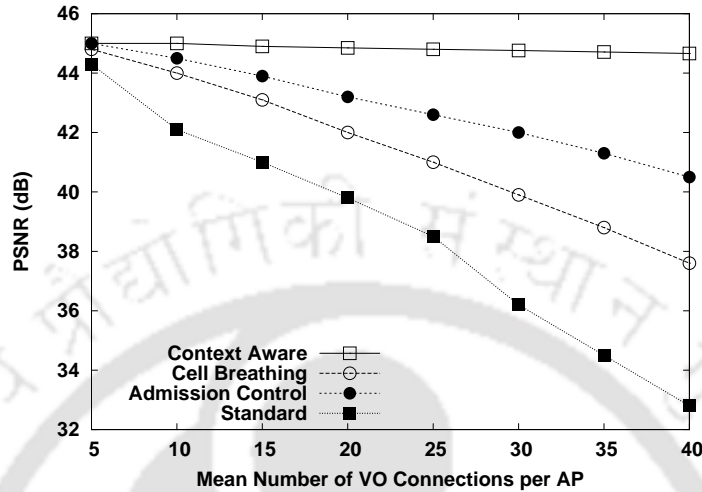


Figure 4.18: Video PSNR with respect to increasing AC_VO traffic

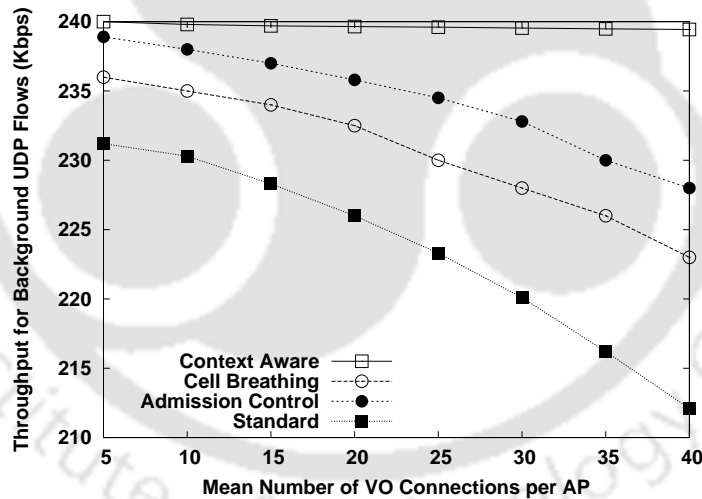


Figure 4.19: Background traffic throughput with respect to AC_VO traffic

higher in the standard procedure, as it demands maximum traffic. On the contrary, the increase in AC_BE connections does not affect the higher priority traffic classes. The figure reveals that only the AC_BE connections are lost in the proposed scheme, as the number of AC_BE connections are increased. We have observed from the testbed emulation traces that connections are lost in the proposed scheme only when the APs are not able to accommodate them even after load balancing. Therefore, the proposed scheme significantly reduces connection losses in the network.

4.3 Experimental Evaluation and Comparison

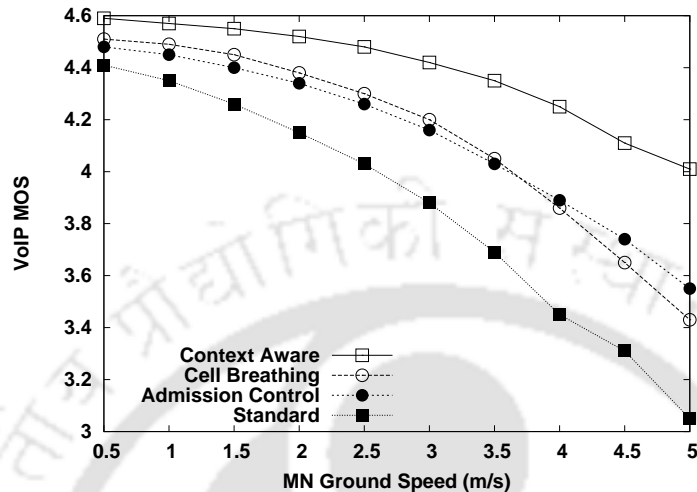


Figure 4.20: Impact of Mobility over QoE Parameters: AC_VO MOS

In the next set of experiments, we have investigated the effect of increasing higher priority traffic over the lower priority traffic classes. For this purpose, initially the network consists of 5 AC_VO, 5 AC_VI, 5 AC_BK and 20 AC_BE traffic. It can be noted that more number of best effort traffics are introduced to keep the network loaded, as the QoS and QoE degradation is mainly observed when the network is sufficiently loaded. The performance of video and background traffic is measured with increasing number of voice traffic in the network. Figure 4.18 shows the PSNR of video traffic with increasing number of AC_VO traffic in the network. The PSNR of the proposed context aware scheme remains almost constant, whereas the PSNR for other three schemes degrades as the number of AC_VO traffic increases in the network. The standard prioritizes the AC_VO traffic through CW, AIFS and TXOP, that impacts the performance of other lower priority traffic classes. Similar observation is founded for the performance of AC_BK traffic, as shown in Figure 4.19. It can be seen from these experiments that the context aware scheme not only maintains QoS and QoE for higher priority traffic, but also preserves the performance of low priority traffic classes those require minimum performance guarantee for user satisfaction.

Next we investigate the impact of mobility over the QoE parameters. For this set of experiments, 10 AC_VO, 10 AC_VI, 15 AC_BK and 20 AC_BE traffic are distributed in the network. Out of these different traffic sources, 40% traffic is mobile, and rest are static. The distribution of static users follows a similar pattern as described earlier. The effect of mobility is analyzed with respect to the ground speed of the mobile users, that is measured

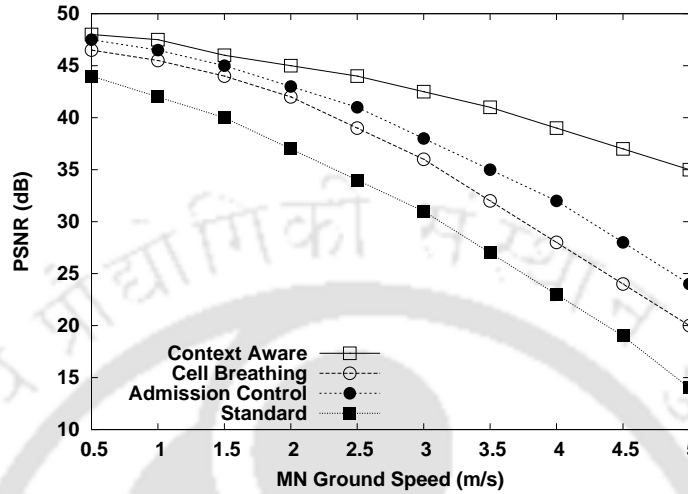


Figure 4.21: Impact of Mobility over QoE Parameters: C_VI PSNR

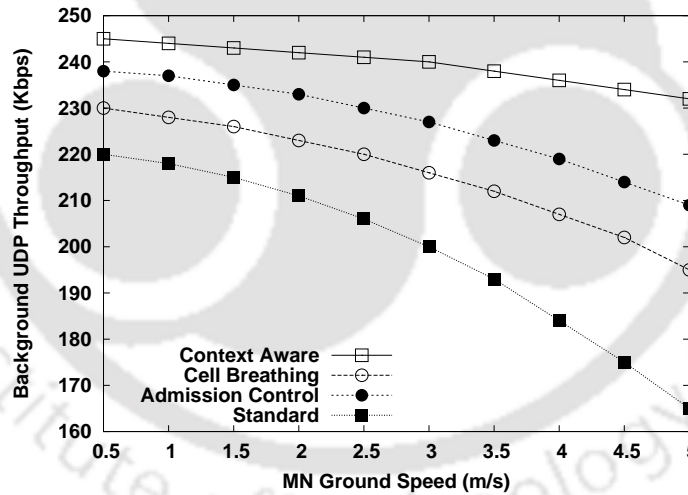


Figure 4.22: Impact of Mobility over QoE Parameters: Background Traffic Throughput

from the geographical positioning sensors equipped with the MNs. Figure 4.20 shows the impact of mobility over voice traffic. It can be noted that high mobility impacts the end-to-end delay as well as jitter in the network, whereas the bandwidth occupancy is controlled by the traffic distribution at the APs. For voice traffic, cell breathing performs better than admission control at low mobility, whereas admission control performs better than cell breathing at high mobility. At low mobility, cell breathing can control traffic concentration at the APs, however at high mobility, cell breathing fails to control traffic concentration

4.4 Summary

as nodes frequently enter and leave the coverage of an AP. Nevertheless, the proposed scheme always shows consistent MOS, higher than the other schemes, by controlling traffic concentration and bandwidth pre-reservation for the mobile nodes. Figure 4.21 shows the PSNR for video traffic and Figure 4.22 compares the four schemes in terms of background traffic throughput. The context aware scheme significantly improves performance at high mobility through bandwidth pre-reservation and admission control. It has been observed from the traces, that the proposed scheme drops best effort connections when traffic concentration becomes high, whereas the QoS and QoE for higher priority traffic are always maintained.

4.4 Summary

In this chapter, we presented a context aware handover management scheme for public IEEE 802.11e wireless LANs, considering the QoS and QoE to the end users. The proposed scheme improves the performance of overloaded APs by handing over traffic to nearby APs, while maintaining the QoS and QoE of the connections. We have developed a bandwidth management scheme for different IEEE 802.11e traffic classes, and designed an admission control mechanism over the standard procedure. A handover management scheme is also developed based on the class aware traffic load information, collected locally at the MNs and the APs. The performance of the proposed scheme is evaluated using a practical IEEE 802.11g+e testbed, and compared with other related schemes proposed in the literature.

Till now, the handover management scheme is discussed for horizontal handover in an IEEE 802.11 network. The next chapter extends the proposed scheme for vertical handover, where multiple overlapping IEEE 802.11 BSSes form communication holes under the common coverage of an IEEE 802.16 WiMAX network.

Chapter 5

Deciding Handover Points in a WiFi-WiMAX Heterogeneous Network Maintaining QoS and QoE

As discussed earlier in Chapter 1, mobility is not the only issue that triggers vertical handover between two technologies in a HetNet environment. This chapter considers a scenario where users migrate from WiFi to WiMAX when QoS/QoE of the running application drops due to traffic overload at the WiFi network. It can be noted that QoS provisioning in WiFi is based on differentiated service architecture (DiffServ) where service provisioning is provided, however strict service guarantee is not maintained. On the contrary, WiMAX is based on integrated service architecture (IntServ), where strict service guarantee is maintained. However, as discussed earlier, cost of per bit data transmission as well as power consumption are significantly lower in WiFi compared to WiMAX, and therefore this chapter designs a handover policy in a WiFi-WiMAX HetNet environment, such that both the QoE of the end users as well as cost-to-pay per bit and average power consumption are minimized (by means of allowing more users to communicate via WiFi, whenever available, while maintaining QoS and QoE). The proposed handover policy can coexist and work on the top of any vertical and horizontal handover (handover between two similar technologies) framework, that takes care of lower layer complexities, such as message decoding, differences in lower layer (MAC and physical) frame formats, signal decoding issues etc. The major contributions of this chapter are as follows.

5.1 The Proposed Scheme: Load Balancing and Handover Policy

- This chapter proposes a bandwidth management and admission control scheme for proper distribution of total network traffic over the WiFi-WiMAX integrated environment. The load-imbalance problem in WiFi BSSs is mitigated through distributing the traffic load among the overlapping APs in a WiFi hotspot. The WiMAX network is used as the back-haul to distribute traffic load among the underlying WiFi hotspots, as it has access to all the users' information.
- Based on the bandwidth management scheme, a handover policy is designed that instructs the users when to do a handover between WiFi and WiMAX interfaces, other than normal handover performed due to mobility, to maintain QoS and QoE of the end-users while preferring WiFi interface for communication.

The performance of the proposed scheme is analyzed using simulation results. The performance is also compared with two other recently proposed schemes in the literature - one that proposes a WiFi-WiMAX integrated environment based on media independent handover strategy [22], and another one that considers QoS assurance during vertical handover between WiFi and WiMAX [44].

5.1 The Proposed Scheme: Load Balancing and Handover Policy

This chapter proposes a load balancing and vertical handover decision policy that instructs a user to switch between the two wireless technologies. In a HetNet environment, the users may have mobility, and therefore we use the term mobile node (MN) throughout the rest of this chapter. We assume a scenario where there are several APs interconnected through a back-haul network or distribution system (DS). There are WiMAX BSs that provide larger coverage area, and WiFi BSSs creates hotspots under the WiMAX coverage. The WiMAX BS is also connected with the back-haul. The HetNet architecture considered in this chapter has following properties.

1. A group of overlapping APs creates a WiFi hotspot under the WiMAX coverage area. There can be more than one WiFi hotspots under the coverage of a single WiMAX BS. However, it is not necessary that WiFi connectivity exists at all points under the WiMAX coverage.
2. Every MN has at least two interfaces, one WiFi and the other WiMAX.

5.1 The Proposed Scheme: Load Balancing and Handover Policy

3. If a MN generates traffics from more than one service classes, the the higher priority traffic class is considered for handover. This assumption is necessary to decide whether the MN meets the specific QoS requirements. We give more weightage to the higher priority traffics to ensure QoE of a MN (or user).
4. The bandwidth reserved for the WiMAX UGS traffics is not used to allocate bandwidth to the MNs. UGS represents a traffic of fixed sized data packet at fixed intervals, intended mainly for leased lines like E1 and T1 traffic. It is difficult to provide such a service in WiFi in presence of other traffics in the same BSS. Such a service shall be retained in the WiMAX network only.
5. The proposed load balancing scheme is designed for a scenario, where excess bandwidth is available at the network to allow MNs to associate either with an AP or with a BS, while maintaining QoS and QoE. It can be noted that traditionally WiFi network does not have any admission control mechanism, and therefore if more MNs associate with an AP than its maximum capacity, QoS/QoE can be affected. Therefore, an admission control mechanism is designed in this chapter to tackle this situation.

During load balancing, traffic shall be transferred between WiFi and WiMAX. Conventionally, WiFi and WiMAX have different group of traffic class. However to maintain interoperability among these two technologies, mapping from WiFi service classes to the WiMAX service classes is necessary. The mapping between the different traffic classes is done as follows.

- WiMAX ertPS traffic maps to WiFi AC_VO traffic; as both represents constant bit rate real time traffic like VoIP with silence suppression.
- WiMAX rtPS traffic maps to WiFi AC_VI traffic; as both represents variable bit rate real time traffics, like video streaming.
- WiMAX nrtPS traffic maps to WiFi AC_BK traffic; as both represents non-real time traffic with some minimum bandwidth guarantee.
- WiMAX BE traffic maps to WiFi AC_BE traffic; as both represents best effort services like elastic traffics.

In the proposed scheme, we present a technique that improves the QoS and QoE of the end users, while minimizing the cost-per-bit and average energy consumption, by

5.1 The Proposed Scheme: Load Balancing and Handover Policy

balancing traffic load across WiFi APs and WiMAX BS in a WiFi-WiMAX HetNet, in the presence of a mix of traffics from different service classes. It takes advantage of the presence of WiMAX BS to smooth out the handover related glitches, such as increase in handover latency and occasional transient overload in APs. The scheme has three components: (i) bandwidth reservation at AP and BS, (ii) admission control at AP and BS, (iii) class aware load balancing and context aware handover.

5.1.1 Bandwidth Reservation at AP and BS

A fixed percentage of the maximum bandwidth is reserved for traffic class AC_VO, AC_VI and AC_BK, so that a fixed number of MNs with traffics of the different classes may be served by an WiFi AP at any time. Let n_{vi} , n_{vo} and n_{bk} be the number of flows of class AC_VI, AC_VO and AC_BK respectively which may be served by an AP. Let average bandwidth requirement for each of class AC_VO, AC_VI and AC_BK be B_{vo} , B_{vi} and B_{bk} respectively. Amount of bandwidth reserved per flow should be of decreasing order for AC_VI, AC_VO and AC_BK traffic respectively (video traffic requires more bandwidth compared to voice traffic, whereas voice traffic requires strict delay and jitter requirements). This is given by Equation (5.1):

$$\frac{B_{vi}}{n_{vi}} > \frac{B_{vo}}{n_{vo}} > \frac{B_{bk}}{n_{bk}} \quad (5.1)$$

We keep some variation of required bit rate provisioning for each traffic class to ensure strict delay and jitter requirements for voice traffics whereas to meet bandwidth requirements for video traffic. Because of this variation, the reserved bandwidth of each traffic class is given an upper limit and a lower limit. Let the lower limit of reservation for AC_VO, AC_VI and AC_BK be L_{vo} , L_{vi} , and L_{bk} respectively and the upper limit of reservation for AC_VO, AC_VI and AC_BK be U_{vo} , U_{vi} , and U_{bk} respectively. Let the total bandwidth is BW . The relation between the allowed variation of AC_VO, AC_VI and AC_BK traffic classes are given by;

$$\frac{U_{vo} - L_{vo}}{L_{vo}} < \frac{U_{vi} - L_{vi}}{L_{vi}} < \frac{U_{bk} - L_{bk}}{L_{bk}} \quad (5.2)$$

It is also required that AC_VO and C_VI always get the required bandwidth within its bit rate variation limit and AC_BK gets the minimum reserved bandwidth. This requires that Equation (5.3) gets satisfied.

$$U_{vo} + U_{vi} + L_{bk} < BW \quad (5.3)$$

The amount of spare bandwidth at WiFi AP S_{ap} , given by Equation (5.4), is required to be greater than zero to keep provision for accommodating a MN performing handover and

5.1 The Proposed Scheme: Load Balancing and Handover Policy

an occasional traffic burst of some flows. It also provides some head room for the AC_BE class flows.

$$S_{ap} = BW - (U_{vo} + U_{vi} + L_{bk}) \quad (5.4)$$

It can be noted that no bandwidth reservation is made for the AC_BE class.

The WiMAX BS reserves bandwidth B_{ertPS} , B_{rtPS} and B_{nrtPS} for the three classes ertPS, rtPS and nrtPS respectively along with some spare bandwidth S_{bs} , using the similar method as described for WiFi APs.

5.1.2 Admission Control

Admission control is required to limit the traffic load at AP and BS within available bandwidth, and also to keep provision for the different traffic classes. Admission control is necessary when a MN first joins an WiFi AP or WiMAX BS. Admission control is also needed when a MN performs handover between APs and BS.

Admission Control at WiFi AP

A MN joining the WiFi AP for the first time with a traffic class is admitted only if the bandwidth occupied by the class does not exceed the lower limit of reserved bandwidth. Let a new flow of class C having a bandwidth requirement of BW_c needs to be admitted. It is admitted only if it satisfies Equation (5.5).

$$BW_c + O_c \leq L_c \quad (5.5)$$

Where L_c is the lower limit of the reserved bandwidth for class C and O_c is the bandwidth already occupied by class C . A MN failing to join an AP is expected to try joining other AP, and in case no AP admits the MN, the MN associates with the WiMAX BS.

A MN performs a horizontal inter-BSS handover from one WiFi AP to another, when it finds that it can not continue to communicate through the associated AP due to poor signal strength, measured in terms of received signal strength identifier (RSSI). The MN performs interleaved scanning, as discussed in Chapter 3, to find accessible APs and performs a horizontal inter-BSS handover to the AP having the best RSSI. During horizontal handover, a MN shall be admitted even if Equation (5.5) is violated. This is possible because of the reservation from the spare bandwidth S_{ap} within the WiFi AP. Let O_{vo} , O_{vi} and O_{bk} are the bandwidth occupied by flows of class AC_VO, AC_VI and AC_BK respectively. A MN, performing a horizontal inter-BSS handover, is admitted to

5.1 The Proposed Scheme: Load Balancing and Handover Policy

a WiFi AP if Equation (5.6) is satisfied.

$$BW - (O_{vo} + O_{vi} + O_{bk}) \geq S_{ap} \quad (5.6)$$

However, if after admission it is found that the lower limit of reservation for the traffic class of the incoming MN is exceeded, a load balancing procedure is initiated immediately, as discussed in the next subsection.

Admission Control at WiMAX BS

As discussed earlier, a MN with traffic class C initiates association with the WiMAX BS when the association with all the WiFi APs in its vicinity fails because of the unavailability of spare bandwidth for traffic class C at the WiFi hotspot. In the proposed scheme in this chapter, if C is a non BE class then the BS finds the total number of MNs with class C . Let this number be N_c . BS admits the MN with a probability \mathcal{P} where,

$$\mathcal{P} = 1 - \frac{N_c}{\frac{B_c}{B_{ac}}} \quad (5.7)$$

Where B_c and B_{ac} are the reserved bandwidth for class C and the average bandwidth demand of a MN with traffics of class C , respectively.

If C is a BE class, then the probability is calculated as;

$$\mathcal{P} = 1 - \frac{N_c}{\frac{(BW_{bs} - (O_{ertPS} + O_{rtPS} + O_{nrtPS}))}{B_{ac}}} \quad (5.8)$$

Where BW_{bs} , O_{ertPS} , O_{rtPS} and O_{nrtPS} are the total bandwidth of BS, bandwidth occupied by flow of class ertPS, bandwidth occupied by flow of class rtPS and bandwidth occupied by flow of class nrtPS, respectively.

5.1.3 Class Aware Load Balancing and Context Aware Handover

As discussed earlier, a WiFi AP admits a handover MN (either horizontal or vertical) of traffic class C , even when the lower limit of bandwidth reservation, L_c exceeds, however the total bandwidth occupancy for traffic class C is less than the upper limit of reservation, U_c . This indicates an impending load overflow for traffic class c , at that AP, say AP_a . Then AP_a initializes load balancing by instructing some of the MNs associated with it to perform a horizontal inter-BSS handover, if possible (spare bandwidth is available at one of the APs in its vicinity), to another AP in their vicinity. It can be noted that this may initiate a cascading procedure. To avoid handover flapping (continuous changes in

5.1 The Proposed Scheme: Load Balancing and Handover Policy

Algorithm 5.1 Load Balancing at WiFi APs

Require: A WiFi AP, AP_a , admits a MN with traffic class C . After admission, the total bandwidth occupancy for traffic class C at AP_a exceeds L_c , though less than U_c

Ensure: AP_a initiates the load balancing procedure

- 1: Let M_a^c denotes the set of MNs associated with AP_a , and have traffics of class C . AP_a selects a MN, $MN_s \in M_a^c$, that has the lowest RSSI. This can be achieved by broadcasting a probe request, and a probe reply from the MNs that carries the RSSI information.
- 2: The AP sends a *MoveRequest* message to MN_s .
- 3: MN_s scans different WiFi channels using the interleaved scanning method as discussed in Chapter 3, and builds a list of accessible APs into *APList*.
- 4: MN_s sends a *MoveResponse* message to AP_a along with *APList*.
- 5: AP_a sends a *LoadRequest(C)* message where C is the intended traffic class, using the back-haul or DS, to all the APs in the *APList*. Therefore this does not incur any signaling or control overhead at the wireless channel.
- 6: The APs receiving the *LoadRequest(C)* message, replies with a *LoadResponse(C)* message through the back-haul or DS, containing its current spare bandwidth for traffic class C .
- 7: Based on the replies it receives from the APs in *APList*, AP_a selects APs from *APList* to which the MN can handover as per Equation (5.9). AP_a keeps the list of such APs in *HCList*.

$$HCList = \{AP_i | AP_i \in APList \wedge O_c^i + M_c^s \leq L_c\} \quad (5.9)$$

Here O_c^i is the total bandwidth occupancy at AP_i for traffic class C , M_c^s is the bandwidth demand for MN_s with traffic class C , and L_c is lower limit for bandwidth reservation for traffic class c .

- 8: AP_a sends *HCList* in a message, *HTMmsg*, to the MN_s .
 - 9: MN_s selects the AP from *HCList*, from which it receives the strongest RSSI. Let this AP be called *APtarget*
 - 10: MN_s then sends a disassociation message to AP_a that contains the target AP address.
 - 11: AP_a sends context information of MN_s to AP_{target} and disassociates MN_s .
 - 12: MN_s changes channel to that of AP_{target} .
 - 13: MN_s sends re-association message to AP_{target} , and completes the horizontal inter-BSS handover to balance traffic load.
-

the AP association), a MN does not re-initiate an association with an AP from which it has received a disassociation request earlier. The load-balancing procedure at WiFi APs has been described in Procedure 5.1.

MN_s may not be able to handover to an WiFi AP as *HCList* may be empty. In that case MN_s performs handover to WiMAX BS. MN_a sends a disassociation message

5.1 The Proposed Scheme: Load Balancing and Handover Policy

to AP_a that contains the BS address. AP_a sends context information to BS through the back-haul or DS. On receiving context information for MN_s , BS prepares itself to allow MN_s to join. MN_s sends a association message to the BS, switches from WiFi to WiMAX interface, and performs a vertical handover.

To minimize cost-per-bit as well as average power consumption, MNs communicating through WiMAX prefer to switch back to WiFi whenever possible. MNs with high mobility may not want to switch to WiFi to avoid frequent handover and hence to reduce the possibility of incurring packet drop or connection drop during handover. The proposed procedure for handing over from BS to AP is similar to Procedure 5.1, except the BS performs the duties of AP_a and MN_s is the intended MN that checks for the possibility of switching back from WiMAX to WiFi. The MNs which are associated with the BS, checks for possible WiMAX to WiFi migration after a fixed timeout interval, called *WlanSwitchTimer*. In our implementation *WlanSwitchTimer* is set to 3 sec.

The load-balancing and handover (horizontal inter-BSS as well as vertical) policy designed in this chapter has following important properties.

Property 1. *As mentioned earlier, the proposed load balancing and handover policy can work on the top of any vertical and horizontal handover mechanism, that takes care of lower layer complexities, such as message decoding, differences in lower layer (MAC and physical) frame formats, signal decoding issues etc. In our implementation, the lower layer handover complexities are handled by the IEEE 802.21 MIH strategy [31], as it simultaneously takes care of both the horizontal and the vertical handover issues.*

Property 2. *A MN associates with the WiMAX BS only in two situations, either (i) there is no WiFi connectivity, or (ii) it does not get desired QoS/QoE because of traffic overload at all of the APs in its vicinity. Therefore, QoS/QoE degradation at a MN, when it is associated with WiMAX BS, indicates that the network is severely overloaded. As the APs are already overloaded, the BS does not have an opportunity to initiate a load balancing for instructing some of the MNs to migrate back to the WiFi.*

This chapter addresses the problem of designing a QoS/QoE based handover policy from the end users' perspective as well as from the network perspective. While the end users migrate from one network to another to balance between desired QoS/QoE profile and cost-per-bit/energy consumption, the network also initiates handover to instruct some of the users to migrate, on the vision of balancing the total traffic load throughout the network. This allows the users to maintain their choice, while the network allows

more number users to associate while maintaining their desired QoS/QoE profiles. The performance of the proposed scheme is analyzed in the next section with simulation results.

5.2 Simulation Results, Performance Evaluation and Comparison

The proposed scheme is implemented in Qualnet-5.0.1 network simulation framework with wireless and advanced wireless model libraries. The IEEE 802.21 MIH framework is implemented in Qualnet following the guidelines given in [85], and the proposed load balancing and handover policy is realized on the top of the MIH framework as a separate module that decides the handover points, outside the normal handover due to mobility, and sends a trigger to the MIH module to perform the lower layer tasks. Two other schemes are also implemented in Qualnet- the scheme proposed in [22] (denoted by “LIM” in this section) that designs a WiFi-WiMAX integration strategy on the top of MIH framework, and the scheme proposed by Ma *et al.* [44] (denoted by “MA” in this section) that designs a handover policy based on QoS/QoE profiles of the end users. Following parameters are used for QoS measurements: average end-users’ throughput with voice, video, background and best-effort traffics, average end-to-end delay experienced by the voice and video traffics and average jitter for the voice traffics. The QoE parameters measured are mean opinion score (MOS) for the voice traffic and structural similarity index (SSIM) for the video traffics. MOS for the voice traffic [86] gives the users’ perception of voice quality between numeric values 1 and 5, while 1 being the worst quality and 5 being the best quality. SSIM for video traffic [86] measures video quality through frame by frame comparison. The value for SSIM lies between 0 and 1, where 1 being the best quality. Additionally, we measure average per-user power consumption to show the effectiveness of the proposed scheme.

5.2.1 Network Scenario Setup

We consider a general network scenario as shown in Figure 5.1, where there is a single WiMAX BS, and under that there are 5 WiFi hotspots, denoted as HS-1 to HS-5. Every hotspot has different number of WiFi APs, however they are with sufficient overlapping, and forms a WiFi cluster. The communication range for WiMAX is considered to be 28 km, where each WiFi AP has a communication range of 200 meter. It can be noted that the large solid circle in the figure does not actually reflects the communication range of WiMAX, rather it is an indicative area only under which all MN lies. The radius of this

5.2 Simulation Results, Performance Evaluation and Comparison

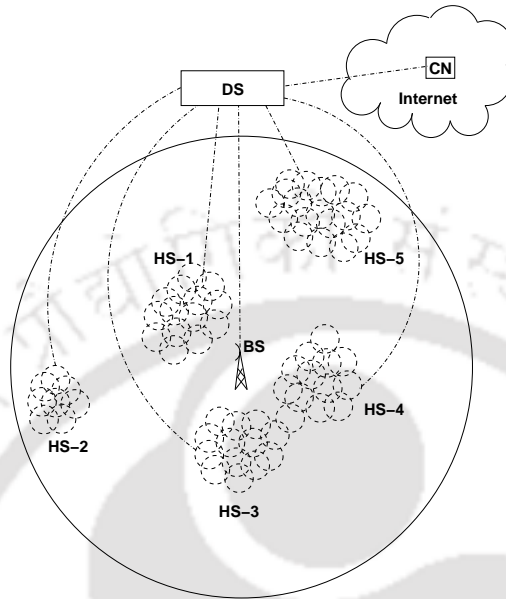


Figure 5.1: Simulation Scenario

area is approximately 2 km. The BS and the APs are connected with a DS through a 1 Gbps dedicated Ethernet link with 1 ms one way latency. The DS is further connected to the outside Internet through another 1 Gbps link with 2 ms one way latency. The corresponding node (CN) lies inside the Internet cloud, and all correspondences are to the CN. The communication parameters for the WiFi and WiMAX are set according to the Intel data sheet as given in [15].

We execute the experiments with varying number of MNs in the network. 40% of the MNs have zero mobility, whereas 60% of the MNs follow random direction mobility model with average speed 40 km/hour. It can be noted that in a commodity network environment, the service provider has prior knowledge of maximum number of users with different traffic classes. We use a network scenario where 40% users generate best effort traffics (like HTTP, FTP etc.), 15% users generate voice traffics, 20% users generate video traffics and rest other users generate background traffics (like telnet, SSH etc.). The maximum and minimum per user bandwidth requirement is set up based on the WiMAX standard, as given by WiMAX Forum, shown in Table 5.1.

In our setup, WiMAX supports 252 Mbps down-link data rate (4 sectors with each sector supports 63 Mbps down-link data rate), whereas each WiFi AP supports 54 Mbps physical data rate. The latency for WiMAX is 1 ms. The initial distribution of the MNs are Poisson over the target area, and every simulation setup is executed for 10 different

5.2 Simulation Results, Performance Evaluation and Comparison

Table 5.1: Per User Traffic Demand

| Traffic Class | Maximum (Kbps) | Minimum (Kbps) |
|--------------------|----------------|----------------|
| Voice (ertPS) | 75 | 25 |
| Video (rtPS) | 100 | 25 |
| Background (nrtPS) | 60 | 20 |

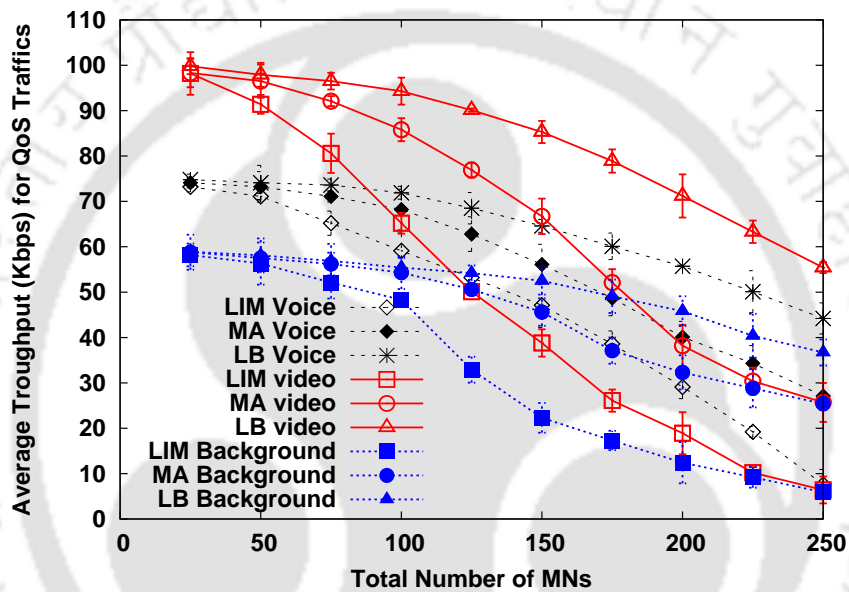


Figure 5.2: Average Throughput for QoS Traffics

times with different initial seed values, and last for 20 minutes. The average is taken to plot the graph. The confidence interval (difference between the maximum and minimum outcomes) is also shown in the graphs.

5.2.2 Analysis of QoS Parameters

Figure 5.2 shows the average per-user throughput for the QoS associated traffics, that is voice, video and background, with respect to the total number of MNs in the network. In the graphs, LB denotes the proposed load balanced handover policy. The simple handoff scheme proposed by Lim *et al.* fails to maintain minimum throughput guarantee for the QoS associated traffic classes, when number of MNs in the network is high. Though the scheme proposed by Ma *et al.* improves throughput and maintains the minimum bandwidth requirement, the throughput still drops with high traffic load. The proposed scheme significantly improves throughput for QoS associated traffic classes, by balancing

5.2 Simulation Results, Performance Evaluation and Comparison

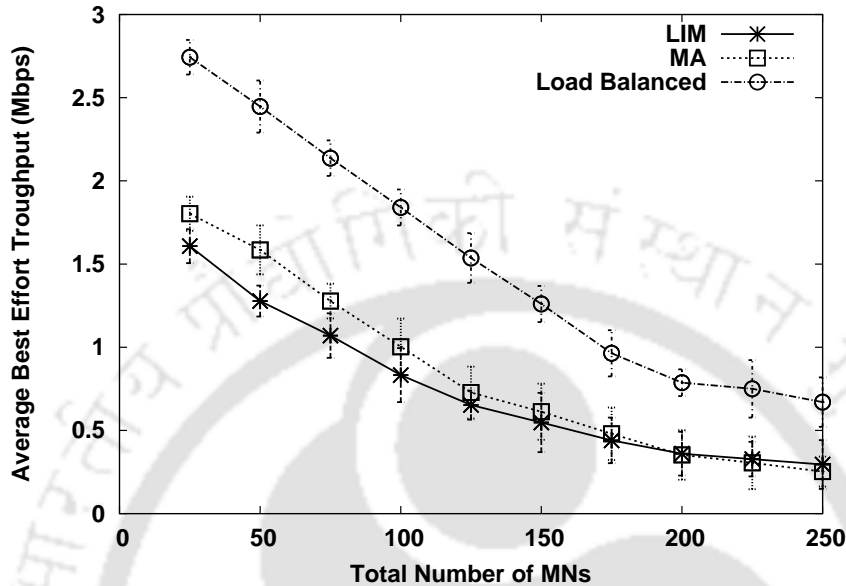


Figure 5.3: Average Throughput for Best Effort Traffics

the traffic load among the available WiFi hotspots and WiMAX network.

Figure 5.3 shows the average per-user throughput for the best effort traffic class. The proposed scheme significantly improves the throughput for best effort traffics by balancing the load throughout the network. It can be noted that the scheme proposed by Ma *et al.* fails to improve best effort throughput at high traffic load, compared to the scheme proposed by Lim *et al.* Though Ma *et al.* ensure QoS/QoE during vertical handover, their scheme fails to balance traffic load, and therefore the high priority traffic classes affect the performance of low priority best effort traffic class. This clearly shows the necessity of load balancing in a HetNet to maintain improved performance for all the traffic classes.

Figure 5.4 and Figure 5.5 show the one-way delay and jitter for the voice and video traffics. To ensure good voice and video quality the delay and jitter should be less than 150 ms and 10 ms respectively. The figures reveal that the schemes by Lim *et al.* and Ma *et al.* fail to provide this guarantee when traffic load is high. It can be noted that in the specific setup mentioned in this chapter, the network becomes almost saturated when number of MNs increases more than 250. Therefore the proposed scheme supports a large number of users while maintaining their QoS requirements.

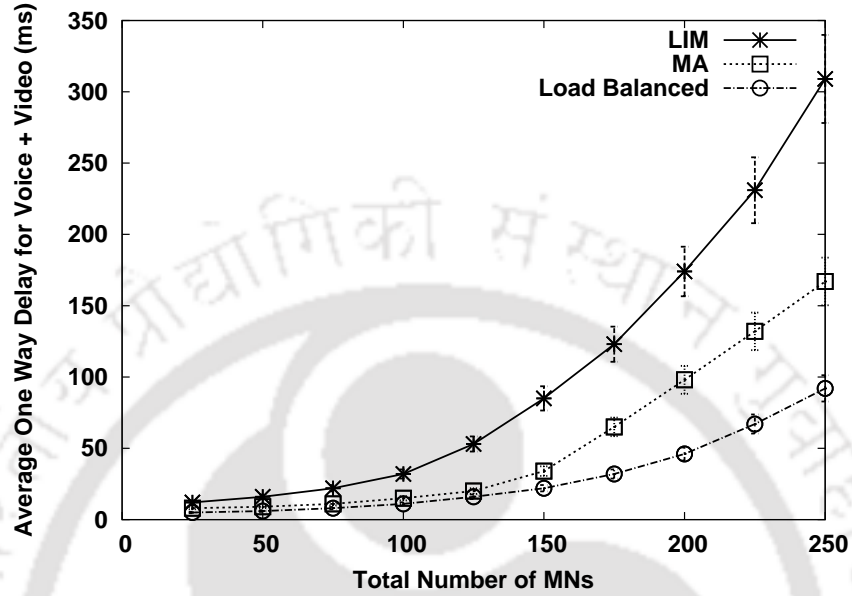


Figure 5.4: Average One-way Delay for Voice and Video Traffics

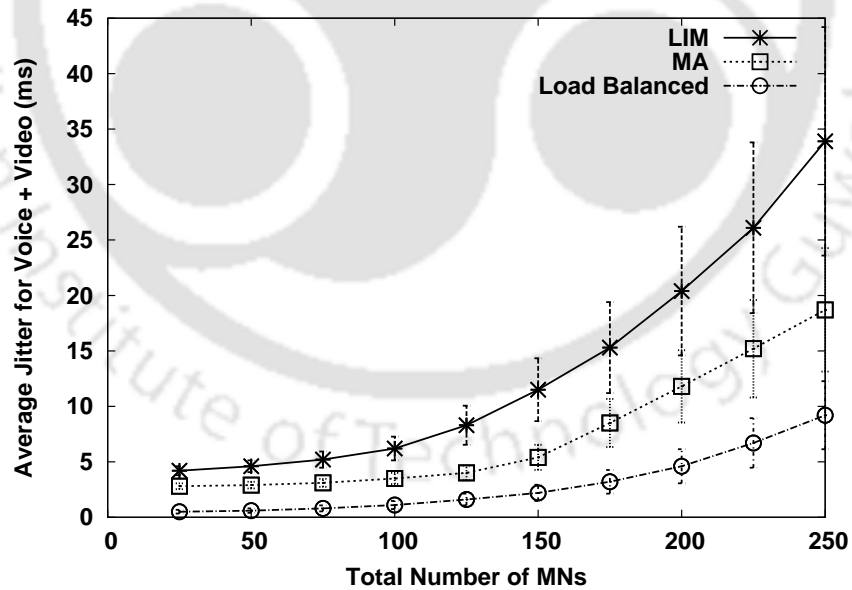


Figure 5.5: Average Jitter for Voice and Video Traffics

5.2.3 Analysis of QoE Parameters

Figure 5.6 compares the schemes for average MOS measured for the voice traffics, and Figure 5.7 shows the SSIM for video traffics. Both the graphs shows that the proposed

5.2 Simulation Results, Performance Evaluation and Comparison

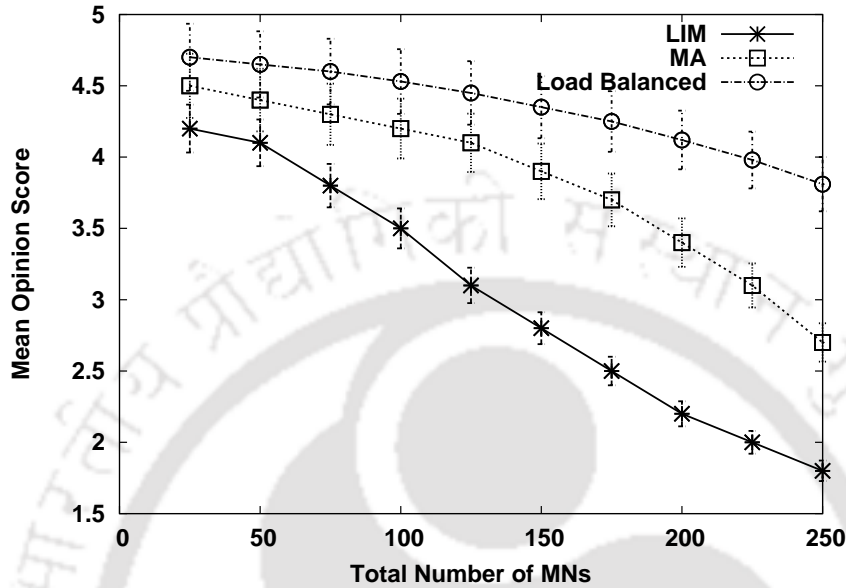


Figure 5.6: MOS for Voice Traffic

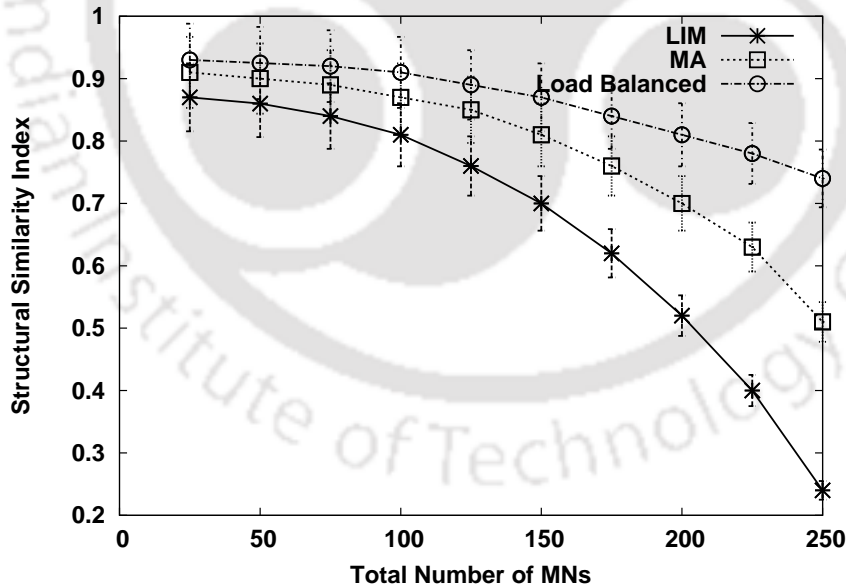


Figure 5.7: SSIM for Video Traffic

scheme significantly improves the users' perception for QoS associated traffic. As discussed earlier, voice traffic requires specific delay and jitter guarantee, while video traffic demands bandwidth and delay guarantee. The bandwidth reservation scheme proposed in this chapter ensures bandwidth guarantee for the voice and video traffic, while

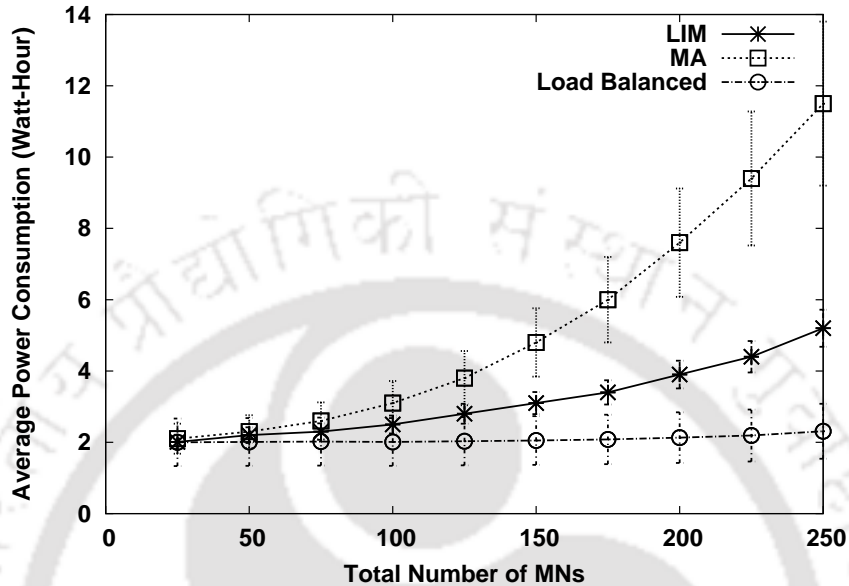


Figure 5.8: Average Per-User Energy Consumption

the load balancing scheme helps to ensure delay and jitter requirements at high traffic load. As a consequence, the proposed scheme significantly improves MOS for voice traffics, and SSIM for video traffics, even at high traffic load.

5.2.4 Average Energy Consumption

As discussed earlier, the proposed scheme aims to reduce average power consumption and costs per bit data communication, by allowing more number of users to communicate via the WiFi interface, while maintaining the specific QoS/QoE profiles for the traffic classes. Figure 5.8 compares the three schemes with respect to the average power consumption. From the previous figures, it can be seen that the scheme proposed by Ma *et al.* significantly improves QoS/QoE compared to the basic handover scheme as proposed by Lim *et al.* However with respect to the average energy consumption, the scheme proposed by Ma *et al.* shows poor performance. With the increase in traffic load in the network, the average energy consumption for the MNs, that uses the scheme by Ma *et al.*, increases almost exponentially. It has been observed from the simulation traces, that most of the MNs migrate to the WiMAX interfaces when traffic load is high. As the scheme does not balance traffic load across the WiFi APs, the WiFi hotspots fails to accommodate MNs for its full capacity, and the MNs prefer to use WiMAX interface to maintain the desired QoS/QoE profile. The proposed load balancing mechanism in this

5.3 Summary

chapter, along with the bandwidth reservation and the admission control policy, keeps the average power consumption almost constant, even at the high traffic load.

5.3 Summary

Designing an effective handover mechanism in a WiFi-WiMAX integrated HetNet environment should consider maintaining a proper handover policy to allow the users to migrate from one network to another based on their QoS/QoE requirements. This enables the end users to effectively utilize the capacity and advantages of both the networks. While the WiMAX technology provides assured QoS/QoE to an end-user, the WiFi interface is cost-effective and consumes less energy for communications. However, the traffic imbalance among WiFi APs prevents the users to associate with WiFi network while maintaining desired QoS/QoE profiles. This chapter proposes a bandwidth reservation and admission control policy based on traffic requirements for different QoS associated services. A class aware load balancing and context aware handover policy is designed, that maintains the bandwidth reservation for different traffic classes, and instructs the users to perform a horizontal or vertical handover to distribute traffic load uniformly throughout the network. This load balancing mechanism allows more end users to associate with the WiFi network, while using the WiMAX for smooth handover to maintain QoS/QoE. Simulation results show that the proposed scheme significantly improves the QoS/QoE for the end users with less communication cost.

Chapter 6

Conclusion and Future Directions of Works

This thesis proposed a context-aware inter-BSS handover management scheme along with QoS support to balance user traffic load across the WLAN, and therefore improves network capacity and QoE for the end users. The proposed horizontal handover for WiFi network is also extended to a vertical handover scenario between WiFi and WiMAX networks, where users perform a vertical handover between the two technologies to balance traffic load while maintaining the required QoS agreement. The contributions of this thesis are summarized as follows.

The first contribution of the thesis proposed a context-aware inter BSS handover management scheme to balance traffic loads across the network. For this purpose, the MNs as well as the APs periodically monitors performance metrics and the impact of traffic load over the MN performance. Whenever a MN experiences performance drop, it consults with the AP (its present point of attachment) to find out an alternate AP in its vicinity. To populate the list of APs with feasible connection opportunities, every MN uses an interleaved scanning mechanism that reduces the handoff latency. Based on the cooperative decision of an impending handover, traffic resources are reserved at the alternate AP, and once the MN experience serious performance drop, it performs a horizontal handover to migrate from one AP to another. The performance of this inter-BSS context-aware handover scheme is analyzed using theoretical modeling as well as results from simulation experiments. The proposed scheme is also evaluated using a practical WiFi testbed, and the performance metrics are compared with other existing schemes in the literature.

6 Conclusion and Future Directions of Works

The inter-BSS handover considers a single class of traffics only, during the pre-handover resource reservation. The second contribution of the thesis extends this mechanism for multi-class traffics, based on the standard IEEE 802.11e service class differentiations. The pre-handover bandwidth reservation is first modeled theoretically using a Markov process, and the effect of per-class bandwidth reservation is analyzed. Based on this theoretical analysis, a class-aware bandwidth reservation scheme is designed that augments the context-aware handover to balance traffic loads based on their service class requirements. The performance of the class-aware load balancing and context aware handover is analyzed using testbed results, and compared with other similar approaches proposed in the literature.

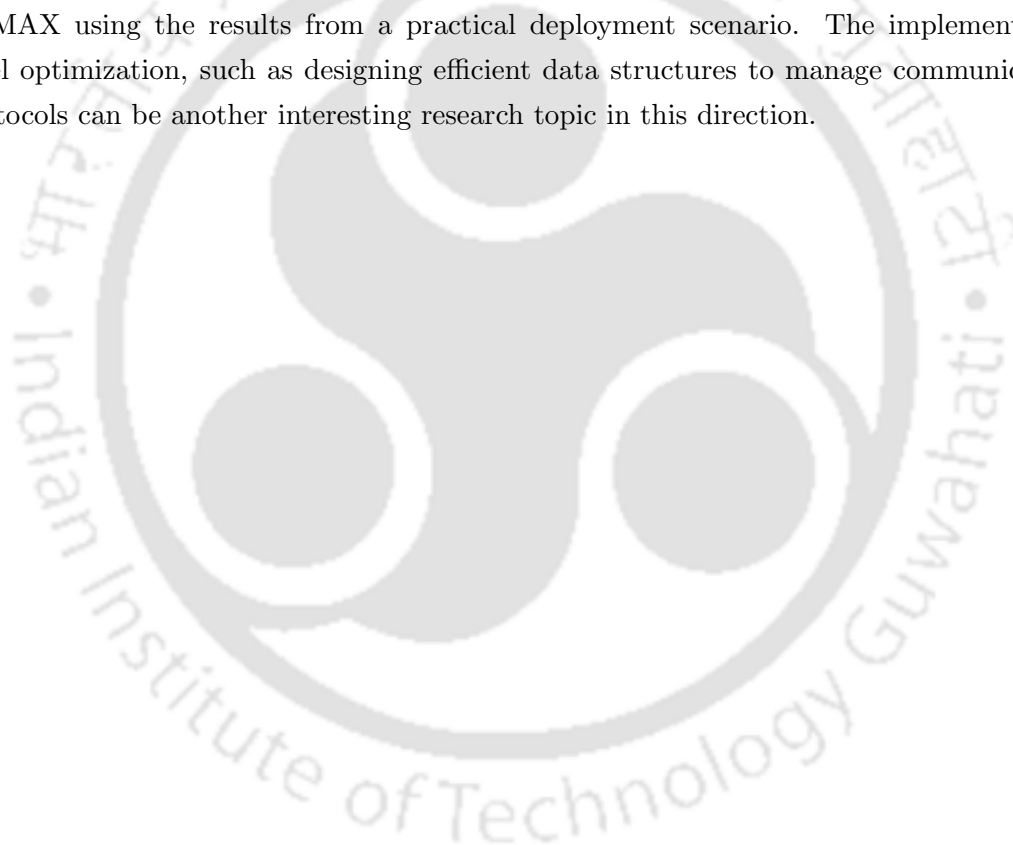
Towards this end, the third contribution of thesis extended the proposed context-aware horizontal handover mechanism for a vertical handover scenario between WiFi and WiMAX. In the vertical handover architecture, the WiMAX network supports a wide communication range like an umbrella, and the WiFi HotSpots create small connectivity areas under it. Because of the inherent differences between WiFi and WiMAX QoS provisioning mechanisms, a suitable mapping tabulation is designed to map between WiFi service classes to WiMAX service classes. According to this mapping, the class-aware load balancing and the context-aware handover mechanisms are augmented to support vertical handover between WiFi and WiMAX. The performance of this scheme is evaluated using simulation results.

The concepts and contributions of this thesis can be augmented and extended in several future directions of researches, as discussed next.

Cellular networks have gained widespread applications in today's data communication architecture with its migration to the fourth and fifth generation data cellular concepts. The interoperability between cellular networks and WiFi has also been well studied [87,88]. The existing literatures show that cellular network has also the inherent problem of load imbalance due to uneven channel assignment and connection patterns [89–93]. Therefore, the concept of context-aware vertical handover has immense opportunity to be designed, augmented and developed for vertical handover between cellular and WiFi networks. Further, cellular network shows heterogeneous connection patterns through micro, pico and femto cellular concepts, and a proper balancing of MNs is required to be designed to decide which MN should use what type of communication coverage, micro, pico or femto. Cellular networks also do not have proper QoS differentiation, and therefore the class aware load balancing for such type of networks can be an interesting future research topic.

All the concepts presented in this thesis are based on the assumption that a MN

generates a single class of traffics only. However, in reality an user may generate different classes of traffics simultaneously. For example, an user may browse Internet while a voice chatting is in progress. In this scenario, the user generates both voice traffic as well as best effort traffic. Therefore, the proposed schemes are required to be augmented in application point of view, where cross layer issues can play a major role in designing management mechanisms. Further, the proposed vertical handover scheme is evaluated through the simulation results only. It would be interesting to evaluate the performance of the context aware vertical handover and class aware load balancing between WiFi and WiMAX using the results from a practical deployment scenario. The implementation level optimization, such as designing efficient data structures to manage communication protocols can be another interesting research topic in this direction.





References

- [1] “IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications,” *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pp. 1–2793, March 2012.
- [2] A. Mishra, M. Shin, and W. A. Arbaugh, “Context caching using neighbor graphs for fast handoffs in a wireless network,” in *The IEEE International Conference on Computer Communications (IEEE INFOCOM)*, 2004.
- [3] R. J. Hulsebosch, E. H. Eertink, A. Prasad, H. Wang, P. Schoo, and M. S. Bargh, “Fast authentication methods for handoffs between IEEE 802.11 wireless LANs,” in *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, 2004.
- [4] D. P. Blinn, T. Henderson, and D. Kotz, “Analysis of a wi-fi hotspot network,” in *Proceedings of the 2005 workshop on Wireless traffic measurements and modeling*. USENIX Association, 2005, pp. 1–6.
- [5] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan, “Measurement-based characterization of 802.11 in a hotspot setting,” in *Proceedings of the 2005 ACM SIGCOMM workshop on Experimental approaches to wireless network design and analysis*. ACM, 2005, pp. 5–10.
- [6] B. Potter, “Wireless hotspots: petri dish of wireless security,” *Communications of the ACM*, vol. 49, no. 6, pp. 50–56, 2006.
- [7] M. H. Cheung, A.-H. Mohsenian-Rad, V. W. Wong, and R. Schober, “Random access for elastic and inelastic traffic in WLANs,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1861–1866, 2010.

REFERENCES

- [8] P. Hande, Z. Shengyu, and M. Chiang, “Distributed rate allocation for inelastic flows,” *IEEE/ACM Transactions on Networking*, vol. 15, no. 6, pp. 1240–1253, 2007.
- [9] “IEEE standard for information technology–local and metropolitan area networks–specific requirements–part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications - amendment 8: Medium access control (MAC) quality of service enhancements,” *IEEE Std 802.11e-2005 (Amendment to IEEE Std 802.11, 1999 Edition (Reaff 2003))*, pp. 1–212, 2005.
- [10] “IEEE standard for air interface for broadband wireless access systems,” *IEEE Std 802.16-2012 (Revision of IEEE Std 802.16-2009)*, pp. 1–2542, Aug 2012.
- [11] D. Niyato and E. Hossain, “Wireless broadband access: WiMAX and beyond - integration of WiMAX and WiFi: Optimal pricing for bandwidth sharing,” *IEEE Communications Magazine*, vol. 45, no. 5, pp. 140–146, May 2007.
- [12] H.-T. Lin, Y.-Y. Lin, W.-R. Chang, and R.-S. Cheng, “An integrated WiMAX/WiFi architecture with QoS consistency over broadband wireless networks,” in *proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference*, 2009, pp. 474–480.
- [13] W. Wang, X. Liu, J. Vicente, and P. Mohapatra, “Integration gain of heterogeneous WiFi/WiMAX networks,” *IEEE Transactions on Mobile Computing*, vol. 10, no. 8, pp. 1131–1143, Aug. 2011.
- [14] R. C. Meier, K. Dettloff, and J. G. Waclawsky, “System and method for integrated WiFi/WiMAX neighbor AP discovery and AP advertisement,” Mar. 2013, US Patent 20,130,070,641.
- [15] WiMAX, Intel, “Intel WiMAX/WiFi link 5350 and Intel WiMAX/WiFi link 5150,” 2014. [Online]. Available: http://www.intel.com/products/wireless/wimax/wifi/index.htm?iid=pr_hyrelease_link8
- [16] L. Taylor, R. Titmuss, and C. Lebre, “The challenges of seamless handover in future mobile multimedia networks,” *IEEE Personal Communications*, vol. 6, no. 2, pp. 32–37, 1999.
- [17] M. Kassar, B. Kervella, and G. Pujolle, “An overview of vertical handover decision strategies in heterogeneous wireless networks,” *Computer Communications*, vol. 31, no. 10, pp. 2607–2620, 2008.

REFERENCES

- [18] Y. Bejerano, S.-J. Han, and L. E. Li, "Fairness and load balancing in wireless LANs using association control," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, 2004, pp. 315–329.
- [19] Y. Bejerano and S. jae Han, "Cell breathing techniques for load balancing in wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, pp. 735–749, 2009.
- [20] R. Daher and D. Tavangarian, "Resource reservation and admission control in IEEE 802.11 WLANs," in *Proceedings of the 3rd international conference on Quality of service in heterogeneous wired/wireless networks*, 2006.
- [21] A. Balachandran, P. Bahl, and G. M. Voelker, "Hot-spot congestion relief in public-area wireless networks," in *proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*, 2002, pp. 70–80.
- [22] W.-S. Lim, D.-W. Kim, Y.-J. Suh, and J.-J. Won, "Implementation and performance study of IEEE 802.21 in integrated IEEE 802.11/802.16e networks," *Computer Communications*, vol. 32, no. 1, pp. 134–143, 2009.
- [23] G. Prasath, K. R. Raghu, and M. Ma, "Integration of WLAN and WiMAX with base station assisted QoS," in *proceedings of the 5th IFIP International Conference on Wireless and Optical Communications Networks*, 2008, pp. 1–5.
- [24] J. Jackson Juliet Roy, V. Vaidehi, and S. Srikanth, "Always best-connected QoS integration model for the WLAN, WiMAX heterogeneous network," in *proceedings of First International Conference on Industrial and Information Systems*, 2006, pp. 361–366.
- [25] Z. Dai, R. Fracchia, J. Gosteau, P. Pellati, and G. Vivier, "Vertical handover criteria and algorithm in IEEE 802.11 and 802.16 hybrid networks," in *proceedings of the IEEE International Conference on Communications*, 2008, pp. 2480–2484.
- [26] T. Ali-Yahiya, K. Sethom, and G. Pujolle, "Seamless continuity of service across WLAN and WMAN networks: Challenges and performance evaluation," in *proceedings of the 2nd IEEE/IFIP International Workshop on Broadband Convergence Networks*, 2007, pp. 1–12.
- [27] A. Balachandran, G. M. Voelker, and P. Bahl, "Wireless hotspots: current challenges and future directions," *Mobile Networks and Applications*, vol. 10, no. 3, pp. 265–274, Jun. 2005.

REFERENCES

- [28] L. Zheng and D. B. Hoang, "Overlapping impacts and resource coordination for high-density wireless communication," in *Proceedings of the International Conference on Computing and Communication Technologies*, 2009, pp. 1–7.
- [29] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the IEEE 802.11 MAC layer handoff process," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 2, pp. 93–102, Apr. 2003.
- [30] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless LAN," *SIGMETRICS Perform. Eval. Rev.*, vol. 30, no. 1, pp. 195–205, Jun. 2002.
- [31] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: media-independent handover services," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 96–103, 2008.
- [32] K. Gomez, D. Boru, R. Riggio, T. Rasheed, D. Miorandi, and F. Granelli, "Measurement-based modelling of power consumption at wireless access network gateways," *Comput. Netw.*, vol. 56, no. 10, pp. 2506–2521, Jul. 2012.
- [33] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," *Comput. Netw.*, vol. 52, no. 14, pp. 2690–2712, 2008.
- [34] H. Zhai, X. Chen, S. Member, Y. Fang, and S. Member, "How well can the IEEE 802.11 wireless LAN support quality of service?" *IEEE Transaction on Wireless Communications*, vol. 4, pp. 3084–3094, 2005.
- [35] G. Sawma, I. Aib, R. Ben-El-Kezadri, and G. Pujolle, "ALBA: An autonomic load balancing algorithm for IEEE 802.11 wireless networks," *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, pp. 891–894, 2008.
- [36] H. Gong and J. Kim, "Dynamic load balancing through association control of mobile users in WiFi networks," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 342–348, May 2008.
- [37] E. H. Ong and J. Khan, "An integrated load balancing scheme for future wireless networks," in *Proceedings of the IEEE GLOBECOM Workshops*, 2008.
- [38] S.-T. Sheu and C.-C. Wu, "Dynamic load balance algorithm (DLBA) for IEEE 802.11 wireless LAN," *Journal of Science and Engineering*, vol. 2, no. 1, pp. 45–52, 1999.

REFERENCES

- [39] F. Guo and T.-c. Chiueh, "Scalable and robust WLAN connectivity using access point array," in *Proceedings of the International Conference on Dependable Systems and Networks*, 2005, pp. 288–297.
- [40] Y. Lee, K. Kim, and Y. Choi, "Optimization of AP placement and channel assignment in wireless LANs," in *Proceedings of the 27th Annual IEEE Conference on Local Computer Networks*, 2002, pp. 831–836.
- [41] L. Du, M. R. Jeong, A. Yamada, Y. Bai, and L. Chen, "QoS aware access point selection for pre-load-balancing in multi-BSSs WLAN," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, 2008, pp. 1634–1638.
- [42] B. H. Park, Y. M. Kwon, S. J. Bae, and M. Y. Chung, "Implementation of connection management scheme of mobile stations in IEEE 802.11 WLAN APs considering transmission power control," in *Proceedings of the IEEE TENCON Spring Conference*, 2013, pp. 515–519.
- [43] S. Rashwand and J. Mišić, "Stable operation of IEEE 802.11e EDCA: Interaction between offered load and MAC parameters," *Ad Hoc Netw.*, vol. 10, no. 2, pp. 162–173, Mar. 2012.
- [44] D. Ma and M. Ma, "A QoS oriented vertical handoff scheme for WiMAX/WLAN overlay networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 598–606, 2012.
- [45] 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, 2005.
- [46] Y. Liu and M. Meng, "Survey of admission control algorithms in IEEE 802.11e wireless LANs," in *proceedings of the International Conference on Future Computer and Communication*, 2009, pp. 230–233.
- [47] Q.-l. Wu, Z.-j. Huang, and S.-y. Wang, "Heterogeneous voice flows-oriented call admission control in IEEE 802.11e WLANs," *International Journal of Electronics*, pp. 531–552, 2014.
- [48] C. M. Assi, A. Agarwal, and Y. Liu, "Enhanced per-flow admission control and QoS provisioning in IEEE 802.11e wireless LANs," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1077–1088, 2008.

REFERENCES

- [49] N. Chendeb Taher, Y. Ghamri Doudane, B. El Hassan, and N. Agoulmine, "Towards voice/video application support in 802.11e WLANs: A model-based admission control algorithm," *Computer Communications*, vol. 39, pp. 41–53, 2014.
- [50] F. U. Ahemd and S. K. Sarma, "QoS and admission controller in IEEE 802.11e WLAN," in *proceedings of 4th IEEE International Conference Intelligent Systems Modelling & Simulation*, 2013, pp. 468–471.
- [51] H. Velayos, V. Aleo, and G. Karlsson, "Load balancing in overlapping wireless LAN cells," in *proceedings of the IEEE International Conference on Communications*, vol. 7, 2004, pp. 3833–3836.
- [52] E. Garcia Villegas, R. Vidal Ferre, and J. Paradells Aspas, "Load balancing in WLANs through IEEE 802.11k mechanisms," in *proceedings of the 11th IEEE Symposium on Computers and Communications*, 2006, pp. 844–850.
- [53] G. Brandner, C. Bettstetter, and U. Schilcher, "Contention-based node selection with applications to relay communications and load balancing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–15, 2013.
- [54] H. Hwang and Y.-T. Kim, "QoS-aware fast BSS transitions for seamless mobile services and load balancing," in *proceedings of the IEEE International Conference on Consumer Electronics*, 2014, pp. 159–160.
- [55] S. Lee, K. Kim, K. Hong, D. Griffith, Y. H. Kim, and N. Golmie, "A probabilistic call admission control algorithm for WLAN in heterogeneous wireless environment," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1672–1676, April 2009.
- [56] D. K. Kim, D. Griffith, and N. Golmie, "A new call admission control scheme for heterogeneous wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3000–3005, October 2010.
- [57] G. Valenzuela, I. Ferreira, O. Nóbrega, and P. Cunha, "Vertical handover decision based on quality of experience in heterogeneous wireless networks," in *proceedings of the 6th Euro American Conference on Telematics and Information Systems*, 2012, pp. 67–72.

REFERENCES

- [58] B. S. Ghahfarokhi and N. Movahhedinia, “A personalized QoE-aware handover decision based on distributed reinforcement learning,” *Wireless Networks*, vol. 19, no. 8, pp. 1807–1828, Nov. 2013.
- [59] D. Rosário, E. Cerqueira, A. Neto, A. Riker, R. Immich, and M. Curado, “A QoE handover architecture for converged heterogeneous wireless networks,” *Wireless Networks*, vol. 19, no. 8, pp. 2005–2020, Nov. 2013.
- [60] A. Garg and K. C. Yow, “Determining the best network to handover among various IEEE 802.11 and IEEE 802.16 networks by a mobile device,” in *proceedings of the 2nd International Conference on Mobile Technology, Applications and Systems*, 2005, pp. 1–6.
- [61] M. Balazinska and P. Castro, “Characterizing mobility and network usage in a corporate wireless local-area network,” in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, 2003, pp. 303–316.
- [62] W. Wei, S. Jaiswal, J. Kurose, D. Towsley, K. Suh, and B. Wang, “Identifying 802.11 traffic from passive measurements using iterative bayesian inference,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 325–338, Apr. 2012.
- [63] J. Sommers and P. Barford, “Cell vs. WiFi: on the performance of metro area mobile connections,” in *Proceedings of the 2012 ACM conference on Internet measurement conference*, 2012, pp. 301–314.
- [64] B. Nunes and K. Obraczka, “Modeling spatial node density in waypoint mobility,” in *Proceedings of the IEEE 9th International Conference on Mobile Adhoc and Sensor Systems*, 2012, pp. 453–457.
- [65] R. Costa, P. Portugal, R. Moraes, and F. Vasques, “An admission control mechanism to handle real-time traffic in IEEE 802.11 networks in open communication environments,” in *Proceedings of the 9th IEEE International Workshop on Factory Communication Systems*, 2012, pp. 63–66.
- [66] *The Qualnet Network Simulator*. [Online]. Available: <http://www.scalable-networks.com>
- [67] *Engineering Statistics Handbook, Single Exponential Smoothing*. [Online]. Available: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm>

REFERENCES

- [68] R. Housley, B. Aboba *et al.*, “Guidance for authentication, authorization, and accounting (AAA) key management,” BCP 132, RFC 4962, July, Tech. Rep., 2007.
- [69] RaLink RT3352 series IEEE 802.11n routers-on-chip. [Online]. Available: http://www.mediatek.com/_en/01_products/04_pro.php?sn=1006
- [70] D. P. Blinn, T. Henderson, and D. Kotz, “Analysis of a Wi-Fi hotspot network,” in *proceedings of the 2005 Workshop on Wireless Traffic Measurements and Modeling*, 2005, pp. 1–6.
- [71] C. Na, J. Chen, and T. Rappaport, “Measured traffic statistics and throughput of IEEE 802.11b public WLAN hotspots with three different applications,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3296–3305, 2006.
- [72] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, “Usage patterns in an urban WiFi network,” *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1359–1372, Oct. 2010.
- [73] G. Divgi and E. Chlebus, “Characterization of user activity and traffic in a commercial nationwide Wi-Fi hotspot network: global and individual metrics,” *Wireless networks*, vol. 19, no. 7, pp. 1783–1805, 2013.
- [74] Y. Cui, T. Ma, J. Liu, and S. Das, “Load-balanced AP association in multi-hop wireless mesh networks,” *Journal of Supercomputing*, vol. 65, no. 1, pp. 383–409, Jul. 2013.
- [75] T. Oliveira, M. Silva, K. Cardoso, and J. Rezende, “Virtualization for load balancing on IEEE 802.11 networks,” in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2012, vol. 73, pp. 237–248.
- [76] Z. Sahinoglu and S. Tekinay, “On multimedia networks: self-similar traffic and network performance,” *IEEE Communications Magazine*, vol. 37, no. 1, pp. 48–52, 1999.
- [77] J. Abate, G. L. Choudhury, and W. Whitt, “Asymptotics for steady-state tail probabilities in structured markov queueing models,” *Commun. Statist.-Stoch. Mod*, vol. 10, pp. 99–143, 1994.

REFERENCES

- [78] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, “Stochastic properties of the random waypoint mobility model,” *Wireless Networks*, vol. 10, no. 5, pp. 555–567, 2004.
- [79] S. Garg and M. Kappes, “Admission control for VoIP traffic in IEEE 802.11 networks,” in *proceedings of the IEEE Global Telecommunications Conference*, vol. 6, 2003, pp. 3514–3518.
- [80] F. Hartanto and G. Carle, “Policy-based billing architecture for Internet differentiated services,” in *Broadband communications*, 2000, pp. 163–174.
- [81] Seagull open source multiprotocol traffic generator. [Online]. Available: <http://gull.sourceforge.net/>
- [82] D. Collins, *Carrier grade voice over IP*. McGraw-Hill New York, 2003, vol. 2.
- [83] Z. Wang, L. Lu, and A. C. Bovik, “Video quality assessment based on structural distortion measurement,” *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.
- [84] B. Bensaou, Z.-N. Kong, and D. H. Tsang, “A measurement-assisted, model-based admission control algorithm for IEEE 802.11e,” *Journal of Interconnection Networks*, vol. 10, no. 04, pp. 303–320, 2009.
- [85] A. Vulpe, S. Obreja, O.-E. Barbu, C.-D. Penciu, M. Buga, and B. Ciobotaru, “Qualnet implementation for mobility management in a MIH enabled system,” in *proceedings of the 8th International Conference on Communications*, June 2010, pp. 523–526.
- [86] K.-T. Chen, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, “A crowdsorceable QoE evaluation framework for multimedia content,” in *Proceedings of the 17th ACM International Conference on Multimedia*, 2009, pp. 491–500.
- [87] D. A. Joseph, B. S. Manoj, and C. S. R. Murthy, “Interoperability of Wi-Fi hotspots and cellular networks,” in *Proceedings of the 2Nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, 2004, pp. 127–136.
- [88] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, “Mobile data offloading: How much can WiFi deliver?” *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, Apr. 2013.
- [89] Y. Jin, N. Duffield, A. Gerber, P. Haffner, W.-L. Hsu, G. Jacobson, S. Sen, S. Venkataraman, and Z.-L. Zhang, “Characterizing data usage patterns in a large

REFERENCES

- cellular network,” in *Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, 2012, pp. 7–12.
- [90] S. Chamberland and S. Pierre, “On the design problem of cellular wireless networks,” *Wirel. Netw.*, vol. 11, no. 4, pp. 489–496, Jul. 2005.
- [91] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, “A first look at cellular network performance during crowded events,” *SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 1, pp. 17–28, Jun. 2013.
- [92] A. Gember, A. Akella, J. Pang, A. Varshavsky, and R. Caceres, “Obtaining in-context measurements of cellular network performance,” in *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, 2012, pp. 287–300.
- [93] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, “An untold story of middleboxes in cellular networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 374–385, Aug. 2011.

Publications Related to the Thesis

Journals

- **Abhijit Sarma**, Shantanu Joshi, Sukumar Nandi, “Context Aware Mobile Initiated Handoff For Performance Improvement In IEEE 802.11 Networks”, International Journal of Computer Networks and Communications, Vol. 3, No. 3, pp. 48-66, May 2011
- **Abhijit Sarma**, Sandip Chakraborty, Sukumar Nandi, Anchal Choubey, “Context Aware Inter-BSS Handoff in IEEE 802.11 Networks: Efficient Resource Utilization and Performance Improvement”, Wireless Personal Communications, Springer, August 2014, Volume 77, Issue 4, pp 2587-2614
- **Abhijit Sarma**; Chakraborty, S.; Nandi, S., ”Context Aware Handover Management: Sustaining QoS and QoE in a Public IEEE 802.11e Hotspot,” Network and Service Management, IEEE Transactions on , September, 2014, vol.PP, no.99, pp.1,1
- **Abhijit Sarma**, Sandip Chakraborty, Sukumar Nandi, “When do I Switch between WiFi and WiMAX? Deciding Handover Points in a Heterogeneous Network Maintaining QoS and QoE”, IEEE Transactions on Vehicular Technology (Under review after Second Revision)

Conference Proceedings

- **Abhijit Sarma**, Sukumar Nandi, “Zone Based Seamless Vertical Handoff Technique between WiFi and WiMax”, in the Proceedings of the Recent Trends in Wireless and Mobile Networks, CCIS, Vol. 162, Part 2, 305-314, 2011

Publications Related to the Thesis

- **Abhijit Sarma**, Sukumar Nandi, “Context Aware Handoff for Performance Improvement in a WiFi-WiMax Heterogeneous Networks”, in the Proceedings of the International Conference on Computer Science and Information Technology, February 2011
- **Abhijit Sarma**, Shantanu Joshi, Sukumar Nandi, “Context Aware Mobile Initiated Handoff for Performance Improvement in IEEE 802.11 Networks”, in the Proceedings of the 2nd International Conference on Networks and Communications, January 2011
- **Abhijit Sarma**, Rajat Gupta, and Sukumar Nandi, “A Zone Based Interleaved Scanning Technique for Fast Handoff in IEEE 802.11 Wireless Networks”, in the Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks, December 2009

Brief Biography of the Author

Abhijit Sarma was born at Mangaldai, Assam India on 1st April 1961. He completed his high school education from Mangaldai and completed pre university from Shillong, Meghalaya, India in 1978. Then he joined Jorhat Engineering College, Assam and completed his B.E. (Electrical) and passed in first class in the year 1984. In 1987 he joined Assam State Electricity Board as Assistant Engineer and subsequently in 1991 promoted to Assistant Executive Engineer. During this time he completed his MCA degree in first class, from Jorhat Engineering College in the year 1990. In 1993, he joined Department of Computer Science, Gauhati University, Assam as Lecturer and presently continuing as an Assistant Professor. His present research interests are in wireless networks and information security.



Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
Guwahati 781039, India