Aalborg Universitet

MANAGEMENT TECHNIQUES FOR CONVERGED WIRED AND WIRELESS NETWORK INFRASTRUCTURES

Ramkumar, Venkata

Publication date:
2013

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Ramkumar, V. (2013). *MANAGEMENT TECHNIQUES FOR CONVERGED WIRED AND WIRELESS NETWORK INFRASTRUCTURES*.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 26, 2020

MANAGEMENT TECHNIQUES FOR CONVERGED WIRED AND WIRELESS NETWORK INFRASTRUCTURES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING
OF
AALBORG UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

M. V. Ramkumar
May 31, 2012

Supervisor:
Associate Professor Neeli R. Prasad, Aalborg University, Denmark

The assesment committee:
Prof. Thomas Skjødeberg Toftegaard, Head of Department of Engineering, Århus University
Prof. Knut Øvsthus, Institutt for Elektrofag, Bergen University, Norway
Assoc. Prof. Tatiana K. Madsen, Aalborg University, Denmark (Chairman)

Moderator:
Assoc. Prof. Peter Fojan, Aalborg University, Denmark

Date of defence: May 31, 2012

Report: RXX-2012
ISSN: 0908-1224
ISBN: 978-87-92328-66-3

Copyright © 2014 by M. V. Ramkumar
All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author.

Dedicated to Lord Sri Rama

Abstract

The recent past has seen an exponential increase in the number of mobile terminals and applications used on these terminal. According to a recent survey from 2010 there are over 5 billion mobile phones, equivalent to around 70% of world's population using mobile phones. In three years more than 300,000 mobile apps have been developed and Apple App Store alone has distributed more than 25 billion apps in the relatively short lifetime of smartphones. This magnificent development translates into a huge demand on network operators to provide users with wireless services - anywhere and anytime. As a result, many new technologies and standards have been proposed to provide higher data rates and to use the scarce spectrum more efficiently. The increase in data rates is partly driven by further deployment of base stations, access points, routers and other data rate increasing network elements capable of improving the network coverage. All of these network elements much be managed in an efficient and autonomous manner in order to maximize the usage and this thesis takes up this major research challenge.

The work carried out in the scope of this thesis addresses many important areas of network management by identifying unsolved problems and proposing novel techniques to solve these problems. The goal is to propose methods for efficient network management in order to better utilize the network resources and network elements such that better services can be offered to the users and the operators will be able to reduce their capital expenditures (CAPEX) and operational expenditures (OPEX).

The first part of the thesis is concerned with radio over fiber (RoF) architectures as part of the convergence of wired and wireless infrastructures into next generation networks (NGNs). RoF architectures use fibre optic transmission as a backbone on which radio signals from various base stations are multiplexed. A novel RoF architecture based on OCDMA and a joint optical- and radio resource allocation framework are proposed. A RoF manager to manage a converged optical and cellular network is presented along with a proof-of-concept for authentication procedures using Triple-A.

In the second part of the thesis, novel approaches are presented for radio resource management (RRM) in homogeneous and heterogeneous radio access

technologies (RATs). A major challenge in RRM is admission control, which is triggered whenever a user initiates a call or when a user is handed off during a call. A fuzzy logic based technique for heterogeneous network scenarios, which includes cell and RAT selection procedures, is proposed along with a Markov-based cross layer technique for homogeneous RAT scenarios that takes into account the physical and medium access control (MAC) layers. A middleware platform, developed as part of the FUTON project, is described and more implementation details are given for a part this platform namely the common radio resource management (CRRM) module, which has been prototyped.

The third part of the thesis considers self-optimization in Long Term Evolution (LTE) networks mainly in the scope of interference control and capacity optimization. A major contribution of this part is the proposed Allocation, Assignment and Admission control (AAA) framework that addresses three important network management problems. The problems addressed under the AAA framework are resource allocation in a hybrid architecture, resource assignment based on application requirements and admission control based on a mean resource calculation. The results obtained from this framework show an increase in the throughput of the network and an increase in the total number of users that can be supported by the network while still satisfying the requirements for both new and existing users.

The last part of the thesis is focused on automated fault management through the proposal of a generic self-healing model, which includes modules for fault localization, key performance indicator (KPI) thresholding and monitoring, fault modeling and fault diagnosis. A survey of existing techniques from the literature is done along with a discussion of pros and cons in order to find suitable and applicable methods for implementing the individual modules. Simulation results are given for a technique for KPI thresholding and monitoring for LTE based networks.

In summary, this thesis addresses many important topics of network management including RRM and performance optimization in general as well aspects of security. Performance and fault management are addressed as modules of self-* management, which is a part of self-organizing networks. The frameworks, methods and techniques proposed in this thesis are, for the most part, applicable to any next generation wireless networks.

Dansk Resume

Den seneste tid har vist en eksponentiel stigning i antallet af mobile terminaler og applikationer, der anvendes på disse terminaler. If ølge en nylig undersøgelse fra 2010 er der over 5 milliarder mobiltelefoner, hvilket svarer til at omkring 70 % af verdens befolkning benytter mobiltelefoner. På tre år er der udviklet mere end 300.000 mobile apps og Apple App Store har alene distribueret mere end 25 milliarder apps i smartphones relativt korte levetid. Denne kæmpe udvikling giver sig udslag i et enormt pres på netoperatører for at forsyne brugerne med trådløse tjenester - hvor som helst og når som helst. Som et resultat er mange nye teknologier og standarder blevet foreslået for at opnå højere datahastigheder og for at anvende det sparsomme spektrum mere effektivt. Stigningen i datahastigheder er delvist drevet af stigende implementering af basestationer, accesspunkter, routere og andre netelementer, der kan øge datahastighederne og forbedre nettets dækning. Alle disse netelementer skal administreres på en effektiv og uafhængig måde for maksimere udbyttet og det er den forskningsmæssige udfordring som denne afhandling omhandler.

Forskningen, der ligger bag denne afhandling adresserer mange vigtige områder indenfor administration af network ved at identificere uløste problemer og foreslå nye teknikker til at løse disse problemer. Målet er at foreslå metoder til effektiv netværksadministration for bedre at udnytte netværksressourcer og netværkskomponenter, således at en bedre service kan tilbydes til brugerne og netværksoperatører vil være i stand til at reducere deres anlægs- og driftsudgifter.

Den første del af afhandlingen beskæftiger sig med radio over fiber (RoF) arkitekturer som en del af konvergensen mellem trådede og trådløse infrastrukturer imod næste generation af netværk (NGN). RoF arkitekturer udnytter fiberoptisk transmission som backbone, på hvilket radiosignaler fra forskellige basestationer multiplexes. En nye RoF arkitektur baseret på OCDMA og et fælles optisk- og radio ressourceallokering model foreslås. En RoF manager til at administrere et konvergeret optisk og mobil netværk præsenteres sammen med et koncept for godkendelsesprocedurer ved hjælp af Triple-A.

I den anden del af afhandlingen, er nye tilgange præsenteret for radio ressource management (RRM) i homogene og heterogene radio access-teknologier

(RATs). En stor udfordring i RRM er adgangskontrol, der udløses, når en bruger starter et opkald, eller når en bruger laver et handover under et opkald. En fuzzy logik-baseret teknik til heterogene netværk scenarier, som omfatter celle og RAT udvælgelsesprocedurer, foreslås sammen med en Markov-baseret krydslagsteknik for homogene RAT scenarier, der tager hensyn til det fysiske og mediumsadgangskontrol Control (MAC) lagene. En middleware-platform, udviklet som en del af FUTON projektet, er beskrevet og flere detaljer for implementering gives for en del af denne platform, nemlig det fælles radio ressource management (CRRM) modul.

Den tredje del af afhandlingen omhandler selv-optimering i Long Term Evolution (LTE) netværk primært indenfor interferenskontrol og kapacitetsoptimering. Et stort bidrag fra denne del, er den foreslåede Allocation, Assignment and Admission control (AAA) model, der adresserer tre vigtige netværksadministrationsproblemer. De problemer, der behandles under AAA modellen er ressourceallokering i en hybrid arkitektur, ressource tildeling baseret på anvendelseskrav og adgangskontrol baseret på en gennemsnitlig ressource beregning. Resultaterne for denne model viser en forøgelse as netværkstrafik og en stigning i det samlede antal brugere, som kan understøttes af netværket samtidig med at kravene fra både nye og eksisterende brugere opfyldes.

Den sidste del af afhandlingen er fokuseret på automatiseret fejl håndtering gennem en foreslået generisk selvhelende model, som omfatter moduler til fejlfinding og KPI tærskel og overvågning, fejlmodellering og fejldiagnose. En gennemgang af eksisterende teknikker fra litteraturen er foretaget sammen med en diskussion af fordele og ulemper med henblik på at finde egnede og anvendelige metoder til at implementere de enkelte moduler. Simuleringsresultater er givet for en teknik til KPI tærskelværdiansættelse og overvågning af LTE baserede netværk.

Sammenfattende omhandler denne afhandling mange vigtige emner indenfor netværksadministration, herunder RRM og performance optimering i almindelighed samt aspekter af sikkerheden. Ydelse og fejlstyring behandles som moduler i selv-* ledelse, hvilket er en del af selvorganiserende netværk. De modeller, metoder og teknikker, der foreslås i denne afhandling er, for det meste, gældende for alle næste generation af trådløse netværk.

Acknowledgements

I would like to thank Prof. Ramjee Prasad for giving me the opportunity to work at CTIF and pursue my PhD here. For this, I will be very grateful to him throughout my life.

I am thankful to my supervisor, Dr. Neeli R. Prasad, for guiding me through this work and keeping faith in me. Without her, I would never have been able to complete my PhD.

I would like to thank Rasmus Hjorth Nielsen for guiding and helping me in this work and for being a wonderful colleague. Also, it was a great pleasure working with Albena and Nicola in the initial part of this work. Further, I would like to thank all my colleagues in CTIF and NETSEC, especially Carlo, Andrei, Jimmy Nielsen, Nuno, Bayu and all my Indian friends in CTIF for making my stay at Aalborg University more joyful. Dorthe Sparre and the rest of the administrative staff did their best to help me out when needed.

Also a big thanks to all my colleagues from the FUTON project for having a great time in all the meetings, discussions, demos, etc. I would also like to thank the colleagues from the NICT project, especially Homare-san for his invaluable insight.

Lastly, I would like to thank my wife Chaithanya for being a great support for me and my son Hari for making me forget all the pressure with his innocent smile. Last but not least, I thank my parents and my wife's parents for being as a backbone to me.

Contents

List of Acronyms

3G 3rd Generation

3GPP 3rd Generation Partnership Project

AAA Allocation, Assignment and Admission control

AC Admission control

AMC Adaptive Modulation and Coding

AP Access Point

BER Bit Error Rate

BS Base Station

BW Bandwidth

CAPEX or capex Capital expenditures

CP Cyclic Prefix

CQI Channel Quality Indicator

CRRM Common Radio Resource Management

CU Central Unit

dB decibel

DHCP Dynamic Host Configuration Protocol

DL Downlink

FFR Fixed Frequency Reuse

FUTON Fibre-Optic Networks for Distributed Extendible Heterogeneous Radio Architectures and Service Provisioning

Hz Hertz

ICI Inter Carrier Interference

IP Internet Protocol

KBps Kilo byte per second

kbps kilo bits per second

KPI Key Performance Indicator

LRRM Local Radio Resource Management

LTE Long Term Evolution

MAC Medium Access Control

MADM Multiple Attribute Decision Making

mbps mega bit per second

MIH Media Independent Handoff

MIMO Multiple Input Multiple Output

MIP Mobile IP

MN Mobile Node

MS Mobile Station

ms millisecond

NE Network Element

NGN Next Generation Network

OFDMA Orthogonal Frequency-Division Multiple Access

OFDM Orthogonal frequency-division multiplexing

OPEX Operational Expenditures

OSI Open Systems Interconnection model

PER Packet Error Rate

PHY Physical Layer

PLR Packet Loss Rate

QAM Quadrature Amplitude Modulation

QoS Quality of Service

RAN Radio Access Network

RAT Radio Access Technology

RAU Remote Antenna Unit

RoF Radio over Fiber

RRM Radio Resource Management

SCM Service/Connection Manager

SFR Soft Frequency Reuse

SNR Signal-to-Noise Ratio

SINR Signal to Interference plus Noise Ratio

SON Self Organizing Network

SSID service set identifier

TCP Transmission Control Protocol

Triple-A Authentication, Authorization and Accounting.

UL Uplink

UMTS Universal Mobile Telecommunications System

VoIP Voice over IP

WCDMA Wideband Code Division Multiple Access

WiMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

XML Extensible Markup Language

1
Introduction

The goal of this chapter is to motivate network management and it is an important problem to address both from an academic and industrial point of view. The key issues, identified and addressed in this thesis, are explained in order to get an overview of the thesis. The contributions of this thesis are explained and the related publications provided. Finally, the outline of the thesis is provided giving an overview of the individual chapters.

1.1 Motivation

In a recent survey by CTIA [1], the numbers of wireless subscriber connections are 322.8 Million in 2011 whereas there are 28.1 Million connections in 1995. To provide the services for this huge increase of mobile users there is an increase in cell sites from 19844 to 256,920. Hence the wireless usage has increased from 37.8 Billion minutes to 2.2 Trillion minutes. The driver for this increase in the wireless usage is increase in the datarates, number of mobile applications and services shown in Figure 1.1.

Figure 1.1: Mobile applications [2]

To provide high data rate services to the users more efficiently, various standards and radio access technologies (RATs) have undergone rapid evolutions. By the evolution from 2G to 4G, from Global System for Mobile Communications (GSM) to Long Term Evolution (LTE), the data rates have increased from 9.4 kbps to 100 Mbps as shown in Figure 1.2. Each standard defines a specific RAT, multiple access techniques and a set of procedures and protocols for communication and supports certain physical parameters such as maximum cell radius and user velocity. As an example, WLAN has limited range and supports only limited mobility whereas Universal Mobile Telecommunications System (UMTS) has higher range and supports higher mobility and LTE provides even better coverage and higher data rates than UMTS and its other predecessors. Depending on the user requirements such as Quality of Service (QoS), user conditions and network conditions, different RATs are suitable for different types of deployment scenarios.

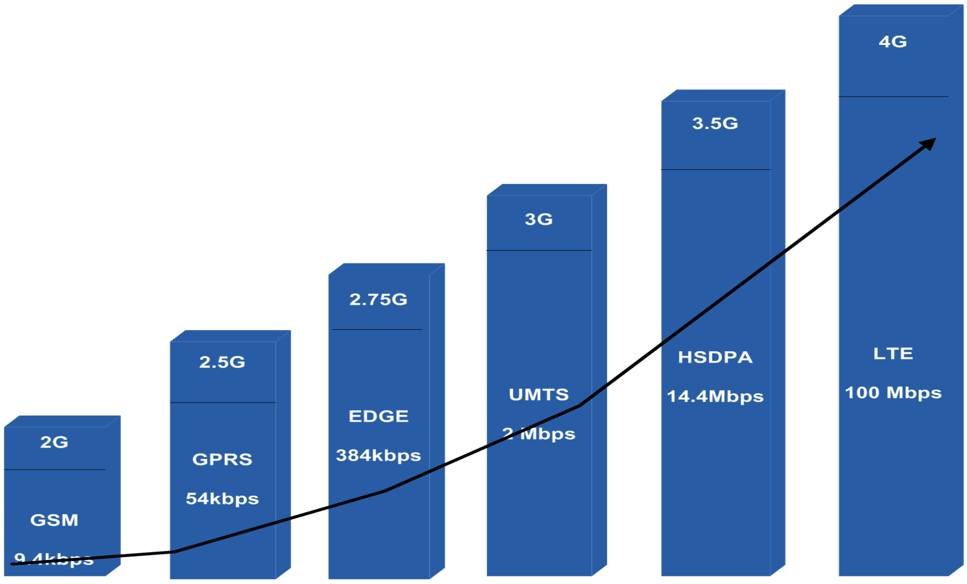


Figure 1.2: The wireless evolution from 2G to 4G [3]

Also, based on the environmental conditions, the user can be indoor or outdoor, in a rural or urban area and can move from one scenario to another. As a consequence, different types of cell deployments have been proposed including macro-, micro-, pico- and femto cells. Figure 1.3 shows the overlay of multi-tier cells sharing a common spectrum.

Hence, to provide services to users and to support applications with diversified requirements under diverse user and network conditions, next generation networks (NGNs) are heading towards integration of wired and wireless communication technologies, which is often referred to as a converged network infrastructure. Next generation converged network infrastructures are envisioned to support multi-radio technologies, multi-tier cells and distributed antenna systems and to support users with different user and network conditions. Figure 1.4 shows such an advanced network architecture.

Due to the complexity of NGNs, there are many challenges and problems that need to be addressed in order to use a common infrastructure in an efficient and seamless manner without significantly having to alter, change or update the deployment. One of the main challenges for an operator is to manage a network with a large degree of diversification and efficiently utilize the network resources, so that the overall spectral utilization is improved, which in turn can provide better services to the user.

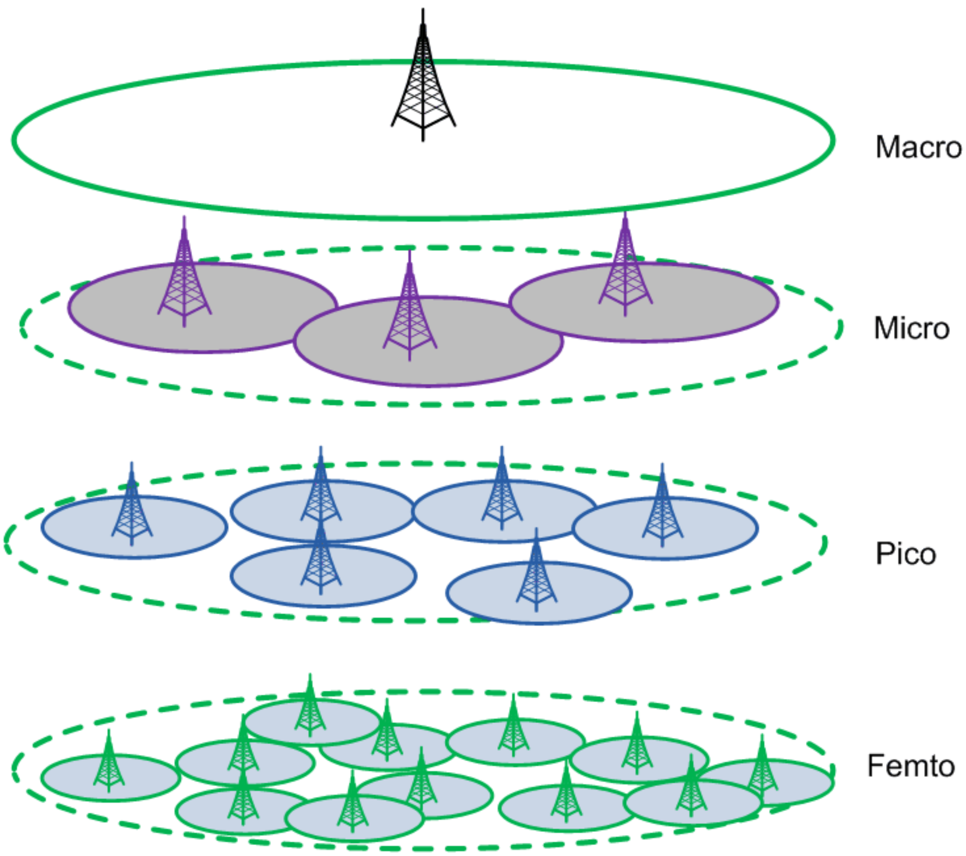


Figure 1.3: Multi-tier Network [4]

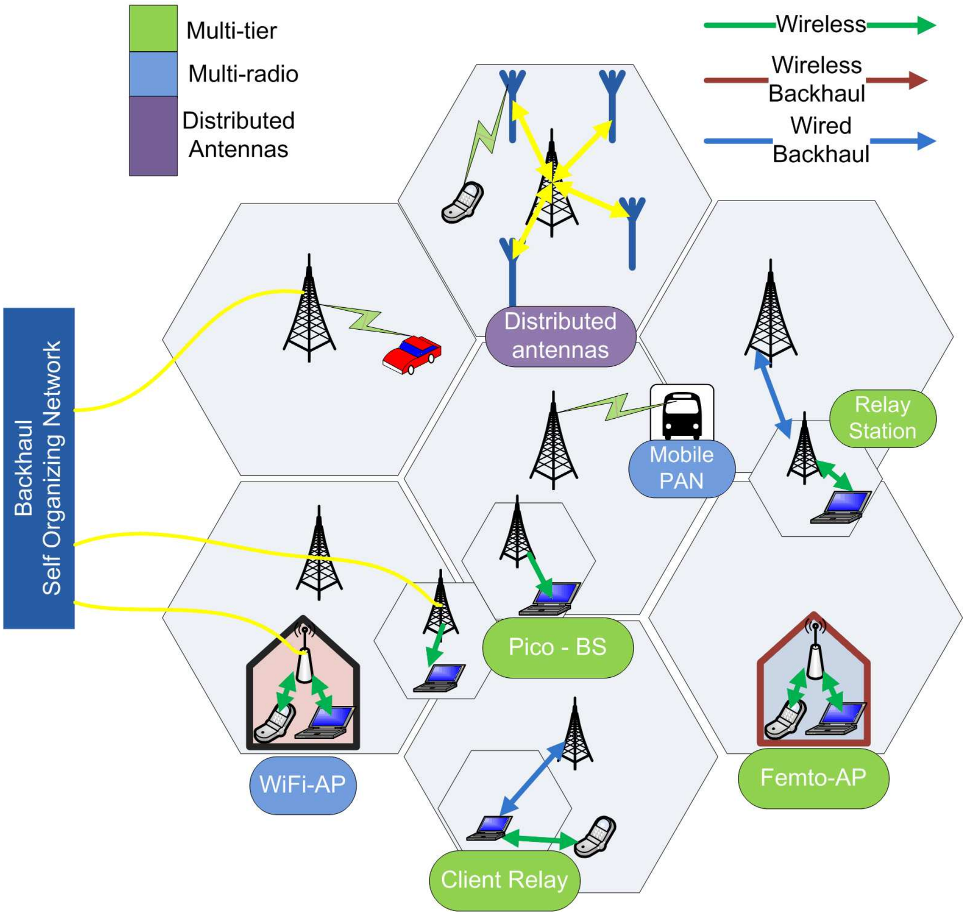


Figure 1.4: Envisioned advanced network architecture [4]

The problem considered in this thesis is, how to improve the overall performance of the network and QoS of the user with efficient network architectures and network management techniques and reduce the OPEX and CAPEX for an operator.

1.2 Challenges

The challenges relating to network management, investigated in the scope of this thesis, are listed below:

- Architectures for NGNs.
- Admission to new or handoff users for heterogeneous and homogeneous scenarios.
- Intercell interference minimization.
- Dynamic resource estimation and allocation.
- Automated fault management.
- Improvement of network throughput and users' QoS.

For the case of heterogeneous network scenarios, a network architecture is required for the integration of heterogeneous RATs. Also, as the cell sizes are decreasing, the number of base stations (BSs) that have to be deployed to cover a given geographical are is increasing, which increases the capital expenditure (CAPEX) for an operator. Due to heterogeneity of RATs, user applications and other network elements (NEs), network managers or management systems are only able make good decisions if information is available on a common platform. Hence, novel architectures are needed and must be able to integrate networks and to provide a platform for efficient network management while reducing operational expenditure (OPEX) and CAPEX for operators.

Another topic considered in this thesis, is admission control (AC), for heterogeneous and homogeneous network scenarios. A user need to be admitted into a cell whenever he or she starts a new call or continue the existing call in case of vertical or horizontal handoff. The user is either admitted or rejected by the system that makes its decision by considering the network conditions, the QoS of the user to be admitted and the QoS of the existing users. If a user is admitted and there is not sufficient resources available in the cell, it will cause network congestion and violation of the QoS for existing users. Similarly, a rejection of a user based on the wrong information causes wastage of bandwidth and resources and hence loss of revenue for the operator. In case of heterogeneous scenarios, each RAT provides different characteristics in terms of e.g. throughput, support for mobility and coverage. The AC should therefore consider the QoS requirements of the user and take a decision whether to admit or reject the user. After the decision, a suitable cell and RAT should be selected for the user.

Due to scarcity of spectrum and the increases in the number of users and applications, the available spectrum has to be reused in various cells. The reuse of the same frequencies, either in downlink or uplink, potentially, cause interference for the users. The simultaneous transmission or reception of two users on the same frequency causes interference to both the users and degrades the throughput. This problem is severe in case of cell edge users, as the received power of the interfering signal is high near the cell edge. Hence, interference minimization techniques, through proper network planning, is essential in order to improve the throughput of the network.

As part of the network planning, each cell in a network is allocated with resources and cell should adapt to changing network conditions. From the user point of view, the allocation should ensure that the users' QoS requirements are met and, from a network point of view, that the overall throughput of the network is maximized. Allocation of resources to each cell and to each user is therefore an important problem, as a wrong allocation increases the overall interference in the network. Another important problem in resource allocation is cell edge users and cell edge throughput as users near the cell edge are more prone to interference and also create more interference in the system compared to cell center users. The resource allocation should consider all these challenges when dynamically allocating the resources.

Fault management is another major problem in network management as a fault in the network increases the total downtime and manual diagnosis is usually time consuming and requires significant human effort, which as an issue both with respect to network availability and the OPEX experienced by the network operator. Self-diagnostic techniques are therefore essential as the size of the network grows involving more and more NEs and thus further complexity. In order to mitigate from the faults in the network, self-healing techniques become important as the next step and the challenges and problems encountered in relation to such are discussed in more detail in Section 8.2.

1.3 Novelty and Contributions

This PhD study contributes to solving the challenges listed in Section 1.2 by proposing novel methods applicable in the area of network management and Figure 1.5 provides an overview of the contributions presented in this thesis.

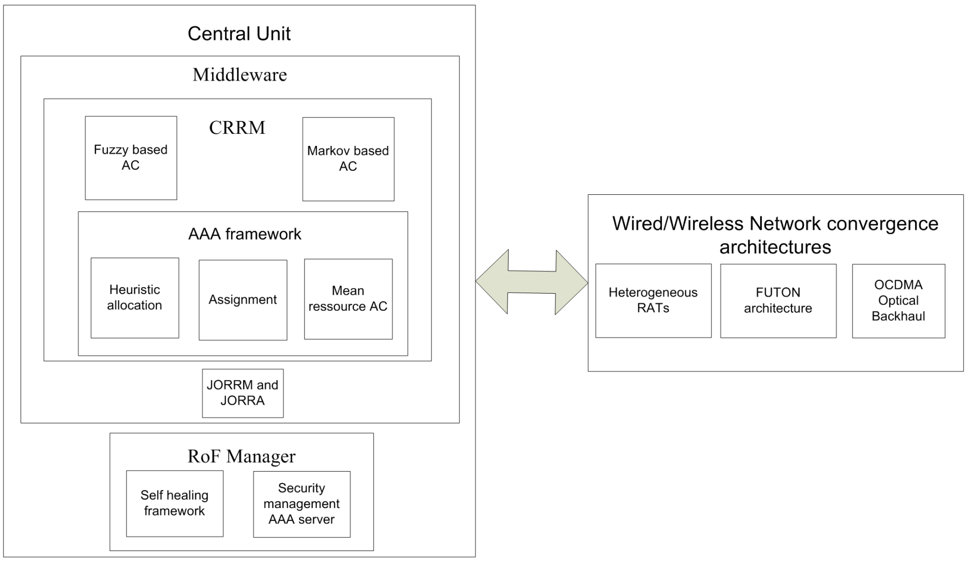


Figure 1.5: Overview of contributions

The first contribution of the thesis is network architectures for backhaul communication or distributed antenna systems using optical code division multiple access (OCDMA) multiplexing. The BSs are connected to a central unit (CU) using a passive optical network (PON) and OCDMA codes are allocated to each BS for transmission or reception. For a RoF network, the RoF manager is presented based the work carried out on FUTON.

The second contribution of the thesis is on AC for heterogeneous and homogeneous network scenarios. The first proposed method of AC is based on fuzzy logic for a heterogeneous network scenario. In this work, cell and RAT selection methods are proposed using fuzzy linguistic controller and fuzzy MADM methods respectively. The proposed method is validated using numerical analysis for the case of a wireless local area network (WLAN) and UMTS. The second proposed method is cross layer AC based on Markov modeling of the queue in the medium access control (MAC) layer and adaptive modulation and coding (AMC) in the physical layer. The proposed method is validated for an orthogonal frequency-division multiple access (OFDMA) based system with simulations in MATLAB and OMNET. The third proposed method of

AC is based on mean resource calculation, which is explained as a part of AAA framework.

The third contribution of the thesis is the proposal of a AAA framework, which considers three main aspects of resource management, namely Allocation, Assignment and Admission control (AAA). The interfaces between different modules of the framework are detailed and a semi-distributed method of allocation is proposed, which considers the geographical location of the users and the changing network conditions. A Markov-based method is applied for assignment, which calculates the number of resources required for a new user to fulfill the QoS requirements. The third aspect of AAA is AC, where a mean resource method is used, which increases the total number of users supported in the system using multi-user diversity based on buffer conditions. The AAA framework with the proposed methods is validated for LTE based networks and the results obtained are compared with state-of-art results.

Another contribution of this thesis is the implementation and design details of middleware and common radio resource management (CRRM) prototypes for FUTON scenario. The implementation is validated with vertical handoff between WLAN and UMTS.

The last contribution of the thesis is on self healing. In this work, the challenges in self healing are presented and a self healing model is proposed with four stages, namely fault localization, KPI thresholding and monitoring, fault modeling and fault diagnosis. For each stage, the various methods from the literature are presented and pros and cons for each method are explained.

The contributions have been, or are in the process of being, validated through peer-review and publication in journals and conference proceedings. The relevant publications are:

- Journal papers
  - M. V. Ramkumar, Neeli R. Prasad, Ramjee Prasad, " Middleware architecture for next generation heterogeneous networks," accepted for publication in Springer, Wireless Personal Communications
  - M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, "An Allocation, Assignment and Admission Control (AAA) Framework for Next Generation Networks," submitted for publication in Springer, Wireless Personal Communications
  - Santiago Carlos, Gangopadhyay, Bodhisattwa, Ramkumar, Venkata, Neeli R. Prasad, Artur Arsenio, "Network Management System

for (FUTON-like) Radio-over-Fiber Infrastructure,” International Journal of Computer and Communication Technology, vol. 1, no. 2,3,4

- Conference papers
  - M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, ”Efficient Admission Control for Next Generation Cellular Networks,” accepted for publication in ICC 2012
  - M. V. Ramkumar, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad, ”QoS-Guaranteed Admission Control for OFDMA-based Systems,” Globecom, pp. 606-610, Dec 2010.
  - M. V. Ramkumar, Albena Mihovska, Neeli R Prasad, Ramjee Prasad, ”Fuzzy-logic based Call Admission Control for a Heterogeneous Radio Environment,” Proceedings of Wireless Personal Multimedia Communications, WPMC, Sep 2009.
  - M. V. Ramkumar, Albena Mihovska, Neeli Prasad, ”Management of Distributed and Extendible Heterogeneous Radio Architectures,” NAEC 2009, Italy
  - Carlos Santiago, Bodhisattwa Gangopadhyay, Artur M. Arsenio, M.V.Ramkumar, Neeli R. Prasad, ”Next Generation Radio over Fiber Network Management for a Distributed Antenna System,” Wireless Vitae, pp. 182 - 186, May 2009.

In addition to the listed publications, contributions have been published as deliverables for the European project ”FUTON” and the joint NICT-CTIF project ”Research and Development on Converged Networks of Wireless and Wired Systems using Frequency-Sharing-Type Wireless Technologies”.

1.4 Thesis Outline

The following provides an outline of the thesis with a brief description of the individual chapters.

Chapter 2: Wired/Wireless Network Convergence Architectures

This chapter discusses the concept, framework and the features of FUTON architecture along with an explanation of the integrated optical radio architecture based on OCDMA for backhaul transmission. The joint optimization problem of optical and radio resource allocation for integrated optical-radio architecture is presented together with the RoF manager for RoF architectures and its main features.

Chapter 3: Radio Resource Management

This chapter introduces the concept of CRRM and its main features. The proposed AAA framework is then presented along with the individual modules of the framework, namely central and local allocation, the assignment and AC module and the priority based scheduler (PBS).

Chapter 4: Dynamic Resource Allocation

This chapter presents a novel heuristic hybrid allocation method for allocation of resources along with the results obtained from the simulation.

Chapter 5: Radio Resource Assignment

This chapter explains the assignment procedure, which is based on Markov modeling of the queue by considering the traffic and channel conditions of the user and estimates the resources required by a user to meet the QoS requirements. This method is validated on an LTE system and the results are presented.

Chapter 6: Call Admission Control

This chapter proposes three methods for AC: Fuzzy logic based, Markov based and mean resource based. Each of the three proposed methods is validated and the results obtained are discussed.

Chapter 7: Self-Healing

This chapter presents a self-healing model with four stages and each stage of the model is explained. For each stage, existing methods from the literature are discussed with pros and cons for applying it. The results obtained for a discretization method as part of KPI thresholding module is presented.

Chapter 8: Proof of Concept

This chapter explains the middleware architecture for FUTON architecture and the modules inside the architecture. The implementation details of the CRRM as a part of the middleware are discussed in detail and the results obtained from the CRRM implementation for the case of vertical handoff between WLAN and UMTS are shown. The implementation details of authentication as a part of the RoF manager are given and the results are explained.

Chapter 9: Conclusions and Future Work

This chapter provides the summary of the overall thesis and discusses future research work.

An overview of the thesis can also be seen from Figure 1.6, which shows the connection between individual chapters.

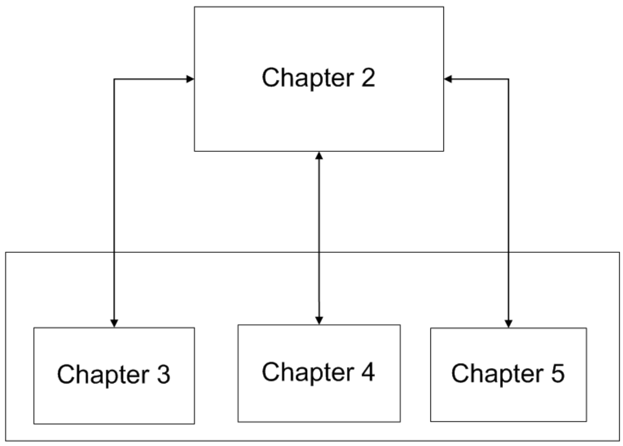


Figure 1.6: Organization of chapters

References

[1] "Background on CTIA Semi-Annual Wireless Industry Survey 2011," by CTIA, www.ctia.org

[2] "The Next Great Entrepreneurs-Mobile Applications," [teamaltman.com]

[3] LTE Evolution [telecominfo.wordpress.com].

[4] IEEE 802.16 Presentation Submission Template (Rev. 9), Document number: IEEE C802.16-10/0016r1, "Future 802.16 Networks: Challenges and Possibilities"

2

Converged Wired/Wireless Network Architectures

In this chapter the FUTON framework, which is a radio over fiber (RoF) architecture for the integration of heterogeneous wireless networks is explained and the features of FUTON are discussed. Then the integrated optical-radio network architecture with OCDMA multiplexing used for backhaul communication is proposed. The joint optimization problem of optical-radio resource allocation is stated.

2.1 Introduction

Network architecture is the design of a communications network, for the specification of NEs and their organization and configuration [1]. With a huge increase in the number of user terminals and the number of applications/services with different QoS requirements, next generation networks (NGNs) are moving towards heterogeneous RANs and RATs in the same geographical area. The user requirements depend on the type of application, for example video streaming, file download, VOIP call etc, and each application have different QoS requirements like datarate, delay, packet error rate (PER) etc. Each network may use a different RAT, for transmission and reception, which can cater a user with specific requirements. Hence a network architecture to integrate heterogeneous RATs in a single framework is required for efficient management of network, to achieve higher data rates, capacity etc. If the information related to each RAT, cell or BS is available at a central location, then it helps in taking global decisions more effectively.

Another challenge in next generation cellular networks is decrease in the cell size to achieve higher bit rates and higher capacity. The reduction of cell size increase the number of cell sites and hence increase the deployment costs for an operator. Also as cell size decreases the problem of inter-cell and intra-cell interference increases, which requires planning of radio resources such that the overall interference in the network is reduced. The parameters like signal to interference plus noise ratio (SINR), pathloss, and channel parameters, QoS requirements of users from various cells, networks and RATs help to make efficient network planning and CRRM. In this context, a network architecture for the management of heterogeneous networks is required to provide services like common AC, and vertical/horizontal handoff etc that can improve the overall performance of network while providing the QoS for existing and new users.

Also with multimode terminals supporting transmission and reception on different RATs, user has the choice to switch between RATs that gives the best service at the best price. From a network point of view, the capacity of the overall network should be maximized, increasing the revenue for an operator. Each RAT suits for some specific requirements, for example some RATs are suitable for low mobility, whereas some RATs are suitable for high mobility and some RATs may provide better datarate. Hence RAT selection is the key in heterogeneous network scenario, in case of a new user and also in case of a vertical handoff. Another important factor while doing handoff is delay, the effect of delay is related to the type of application. For example applications such as voice or gaming are delay sensitive, where as file download can tolerate more delay. In this context network architectures should provide flexibility to switch between RATs [1].

Due to heterogeneity of RATs and the infrastructure to support them, and increase in the number of user terminals and applications, there is an increase in the number of NEs, which provides need for security from bogus NEs. Hence to address above discussed challenges we present the Futon framework that provides a platform for integration of heterogeneous RANs. In Section 2.2, the Futon framework is explained and the main features of Futon are described briefly. In Section 2.3, an integrated optical wireless network architecture is proposed for backhaul communication with the help of OCDMA multiplexing.

2.2 Futon Framework [2]

The Futon framework provides a hybrid fibre-radio network that interconnects multiple RAUs, transparently carrying radio signals to/from a CU, where joint processing takes place. The generic architecture of Futon is intended to cover a geographical area that is divided into several serving areas, where the multifrequency RAUs are located. These RAUs are linked to a CU, using a transparent optical fibre system that can send/receive optical signals from different wireless systems; hence this can be used as a framework for the integration of heterogeneous wireless systems.

Figure 2.1 depicts the Futon scenario where several systems coexist in the same serving area, and shows the main modules of the CU to which all of systems are connected. From a practical point of view, Futon consists of a fixed infrastructure that provides flexibility to share its resources by a wide range of wireless systems and by fixed optical connections. The supported systems should include, among others, outdoor distributed broadband wireless systems (DBWS), where the infrastructure acts as a virtual multiple-input multiple-output (MIMO) enabler to achieve the targeted high bit rates, indoor distributed broadband systems, and the remote control of dedicated radio systems. DBWS is the technology envisioned within the scope of Futon to cope with the International Mobile Telecommunications-Advanced (IMT-A) requirements.

For distributed wireless systems, the availability of the radio signals at a common location allows the joint processing of the signals from different RAUs, enabling the development of virtual MIMO concepts for broadband wireless transmission and also inter-cell interference (ICI) cancelation. In addition, heterogeneous wireless systems can be co-located, which allows the use of a

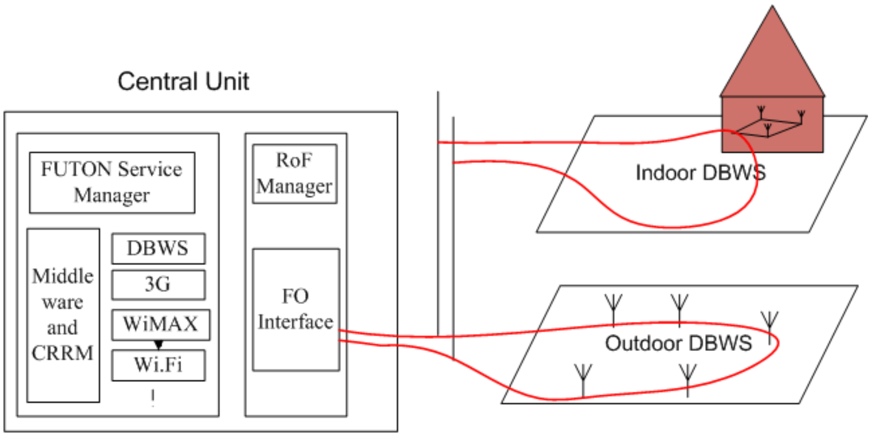


Figure 2.1: Generic Futon Architecture

common management platform. In this context, cross-layer algorithms may be developed to optimize the usage of the radio resources, and may be generalized to cross-system model, enabling the development of efficient CRRM algorithms.

Some of the features of the Futon framework are as below

1. High bit rates can be achieved by mobiles communicating simultaneously with several RAUs with perfect cooperation between them, which can be treated as virtual MIMO through distributed antennas.
2. As the cell size decreases, the problem of ICI increases, and to achieve high system capacity, interference cancelation is achieved by cooperation between RAUs.
3. With the reduction in the cell size to enhance the overall system capacity, the number of BSs is a major factor in the overall cost. In this context, usage of simple RAUs brings significant advantages both in terms of reduced installation and operation and maintenance costs.
4. Inter-system algorithms can be developed, due to processing of signals from heterogeneous systems at a single point.
5. Provides dynamic management of the resources at the CU, overcoming the shortcomings of an a priori static planning.
6. Provides a framework for efficient integration of wireless and wire line technologies.

2.3 Integrated Optical/Wireless Network based on OCDMA

2.3.1 Introduction

Due to the increase in the number of applications and number of mobile terminals, the need to have coverage anywhere also increases rapidly, which is leading towards reduction of cell size and increase in the number of BSs. These BSs are connected to the CU using fiber optic components e.g. as a passive optical network (PON) [3]. The optical signals from BS are multiplexed to transmit on a fiber optic link, which can be done e.g. using Wavelength Division Multiple Access, Time Division Multiple Access, Optical CDMA etc.

The advantages of OCDMA compared to other technologies are its asynchronous operation; hence the BSs can transmit/receive without the need for a global clock. As mentioned, due to smaller cells, there is a vast increase in the number of cells and network elements e.g. routers, switches, BSs etc. Hence, there is need for security, both for devices and data. OCDMA provides inherent security, which is very attractive in providing security for next generation wireless networks. Also, due to the soft blocking feature of OCDMA, the networks can be scalable and also survivable in case of network failures. Lastly, OCDMA allows transmission of traffic at variable bitrates and variable bit error rates (BER). Using OCDMA for backhaul transmission, by connecting BSs to CU is a next step towards increasing the performance of wireless networks, with cell size reduction and hence increases the overall capacity of the network by taking advantage of having central processing of data at a single point [3].

In the context of hybrid optical wireless network, the main challenges faced by a wireless network is ICI, which is caused due to overlap of resource usage by users in neighboring cells. With changing traffic conditions, and the dynamic conditions of the wireless users, the load of the network varies dynamically, which should be considered during resource allocation [4]. When BSs are connected to a CU, using optical link based on OCDMA, the resource allocation should consider the effects of network like load of the network, position of the BS, QoS of the users etc. Similarly the resource allocation in the wireless domain should consider the effects of the optical domain like OCDMA codes available, delay caused by optical link in providing QoS to the user etc. Hence a joint optical radio resource allocation is required, which does allocation of resources to optical and wireless domain simultaneously, by considering the effects of both.

JORRA deals with allocation of radio resources to the BSs in a network and in turn to the users and allocation of OCDMA codes to BSs, which involves finding the right radio resource to be used, amount of resources used by each BS and user, finding OCDMA code for each BS and length of OCDMA code to be used based on the load of the network, QoS of the users etc in every frame, such that the overall throughput of the network is improved and QoS is achieved.

In [5], the Orthogonal Frequency Division Multiple Access (OFDMA) resource allocation and wavelength assignment for RoF network based on Wavelength division multiplexing (WDM) is proposed, where picocells and nanocells are assumed, with each nanocell having 9 picocells. Intra-nano cell interference is eliminated by orthogonal allocation of OFDMA resource blocks (RBs), and inter-nano cell interference is minimized by resource allocation of OFDMA RBs. In [6], QoS aware scheduling algorithm was proposed for hybrid optical wireless network, to improve network throughput and transmission delay. In [7], a delay aware AC for integrated optical wireless network is proposed, where for every new connection request, the total delay in the optical and wireless network is estimated and based on this the new user is admitted, if the delay requirement of the user less than estimated delay. In [8], a dynamic BW allocation over Ethernet Passive Optical Network (EPON) is proposed; where each Optical Network Unit (ONU) gets a time slot for allocation and the excessive BW from lightly loaded ONUs is allocated to highly loaded ONUs. Also intra ONU scheduling is done, by considering services with different QoS. In [9], WDM based TDMA/TDD RoF network architecture with K transceivers (TRX) in Control station (CS) serving N BSs is considered, with each BS tuned to a single wavelength. Each TRX can only support a single BS at a time to avoid wavelength collision and stays tuned to certain wavelength for a time based on the allocated BW to a BS. The goal of the dynamic BW allocation in [9] is to find the amount of time each BS receives optical signal on a particular wavelength based on the load of BS. To the best of our knowledge, the joint resource allocation of optical wireless network based on OCDMA was not considered yet.

2.3.2 System model

In this section the OCDMA system model for back haul communication is explained. Figure 2.2 explains the scenario where fiber optic links using ODCMA multiplexing are used for backhaul transmission between BSs and CU. Here we assume the ONU is integrated with BS. In the downlink, traffic in form of packets, is sent from CU to each BS, which forwards to the corresponding subscriber station (SS). In the uplink traffic from various SS's are aggregated at BS and then forwarded to CU. Here we assume k BSs, which receive the traffic from CU and forward to n SS's. Optical splitter/combiner is used to multiplex and de-multiplex the optical signals in uplink and downlink respectively.

In the optical network each BS is assigned one or more OCDMA codes of different lengths. And each SS in radio domain is allocated with a radio resource.

For this network, a Joint optical- radio resource management (JORRM) manages the optical and radio network jointly, by addressing the challenges of both optical and radio network. The main goal of JORRM is to effectively manage the optical and radio resources and increase the overall throughput of the radio and optical network from network point of view, and satisfy QoS from the SS point of view. For example, from the optical network point of view by proper usage of OCDMA codes, MAI can be decreased and hence the number of BS's served by a CU can be increased. The main features of a JORRM module are Joint optical-radio resource allocation (JORRA), Joint

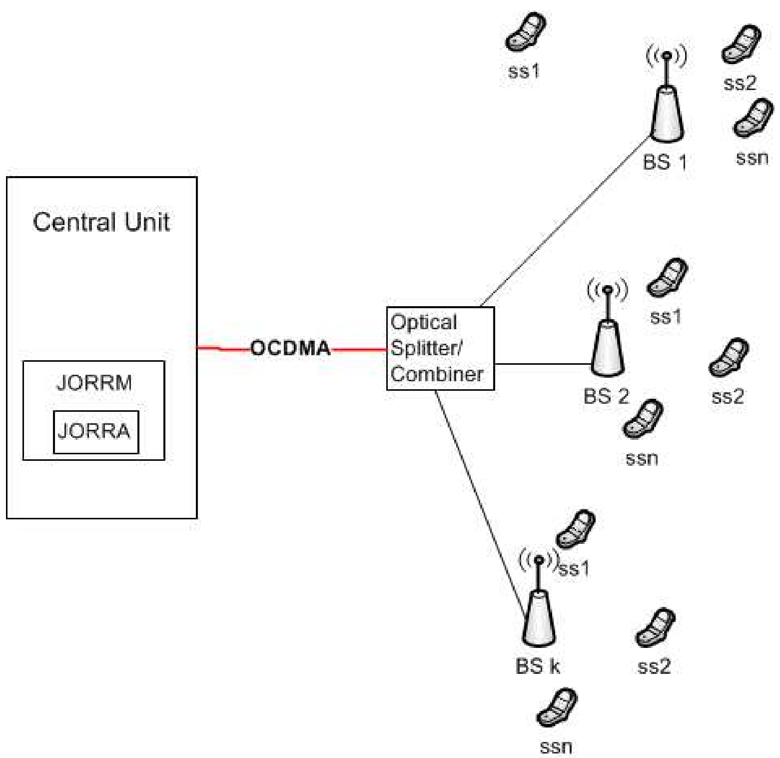


Figure 2.2: System Model

optical-radio AC, Joint optical-radio handoff and Joint optical-radio power control etc. In this section we consider JORRA, which deals with allocation of OCDMA codes to BS for transmission of optical signal from CU to BS and allocation of radio resources to each BS for transmitting radio signals to the SS in the downlink.

2.3.3 Problem definition

One of the main challenges of smaller cells is ICI. In the downlink, ICI is mainly from the neighbor BSs, where as in the uplink, ICI is from users within the cell or from neighbor cells. ICI is due to the overlap of resource usage by various cells in a network, which affects the QoS of the user and overall throughput of the network. If the resources are allocated orthogonally without any overlap between cells, ICI will be zero. As the percentage of overlap increases ICI will increase. The users near cell edge are sensitive to interference as they are close to the neighbor cell and also due to their weak signal strength due to path loss, where as the users near to the BS are not as sensitive to interference as cell edge users. Another challenge is the dynamic conditions of wireless users, such as fading, pathloss etc due to mobility. The dynamism from a network point of view is change in the load conditions, which should be considered by the resource allocation algorithms. Hence a dynamic resource allocation is preferred over a static allocation, which can deal with the changing conditions of the network. In this context, radio resource allocation, as shown in Figure 2.3, deals with finding the right resources for each cell in every frame, slot etc, such that ICI is minimized, overall throughput is maximized and QoS is achieved [4].

The Radio Frequency (RF) signals from various BSs are multiplexed on to optical fibre link to transmit the signals from each BS to a CU. Hence the allocation of OCDMA codes in every frame or super frame to each BS is a major challenge. The length of OCDMA code varies which decides the datarate over fibre link. OCDMA codes are allocated to each BS based on the load or throughput of the BS, position of BS etc.

One of the main concerns of OCDMA is MAI. As the number of simultaneous transmissions increase over a single fibre, each transmission interferes with other, which will reduce the performance of the transmission by increasing the BER. The Optical resource allocation finds the right OCDMA codes to be used by each BS considering the load of the BS, position of the BS and the total number of BS to be allocated. The main goal of optical resource allocation module is the selection of OCDMA codes to each BS, such that overall MAI is reduced, which will increase the capacity of the optical link and hence reduce throughput degradation of BS and QoS of the SS is achieved. Also OCDMA codes are allocated such that the total number of BS's in the network increase, which will increase the revenue for an operator point of view.

Radio resource allocation selects the resources to be used by each BS and by each user, by considering the performance degradation due to optical link while performing the allocation. The selection of resources takes ICI in to consideration and the goal of resource allocation module is to optimize the overall throughput of the network, and provide QoS to the users. This module also finds the number of resources required by each user, to achieve the QoS, by overcoming the optical as well as radio interference. The location of user is also taken in to consideration, as the cell edge users are mostly affected due to ICI compared to interior users. Radio resource allocation also considers orthogonal allocation of resources over various cells, so that two or more cells having resources non overlapped, can be combined to one OCDMA code, hence the number of cells that are served can be increased.

2.4 Conclusions

In this chapter, we discuss fiber optic based network architectures, which could cater some of the challenges like reduced cell sizes, CAPEX, security faced by the next generation cellular networks. The centralized architecture facilitates for efficient network management with the availability of data from all parts of the network. Lastly, a joint resource allocation problem in an integrated wireless-optical network is discussed.

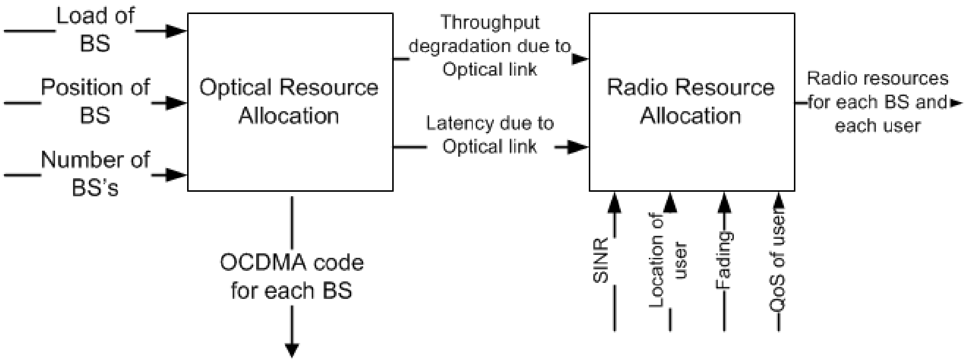


Figure 2.3: Joint optical-radio resource allocation

References

[1] "Heterogeneous network architectures", PhD thesis by Henrik Christiansen

[2] FP7 ICT Project FUTON, at http: www.ict-futon.eu.

[3] Yuanqiu Luo, Ting Wang, S. Weinstein, M. Cvijetic, " Integrating optical and wireless services in the access network," Optical Fiber Communication Conference OFC 2006, June 2006.

[4] M. V. Ramkumar, B. Anggorojati, A.L. Stefan, Neeli R. Prasad, R. Prasad, " QoS-guaranteed admission control for OFDMA-based systems," IEEE Globecom, pp. 606-610 , Dec 2010.

[5] Jingjing Zhang, Nirwan Ansari, " On OFDMA Resource Allocation and Wavelength Assignment in OFDMA-Based WDM Radio-Over-Fiber Picocellular Networks," IEEE Journal on Selected areas in Communications, vol. 29, no. 6, pp. 1273-1283, jun 2011.

[6] Yuanqiu Luo, Si Yin, Ting Wang, Yoshihiko Suemura, Shinya Nakamura, Nirwan Ansari, Milorad Cvijetic, " QoS-Aware Scheduling over Hybrid Optical Wireless Networks," Optical Fiber Communication and the National Fiber Optic Engineers Conference OFC/NFOEC, pp. 1-7, 2007.

[7] Y. Yan, H. Yu, H. Wessing, L. Dittmann, " Integrated resource management framework in hybrid optical wireless networks," Optoelectronics, Institution of Engineering and Technology, vol. 4, no. 6, pp. 267-279, dec 2010.

[8] Chadi M. Assi, Yinghua Ye, Sudhir Dixit, Mohamed A. Ali, "Dynamic Bandwidth Allocation for Quality-of-Service Over Ethernet PONs," IEEE Journal on Selected areas in Communications, vol. 21, no. 9, pp. 1467-1477, nov 2003.

[9] Hong Bong Kim, Adam Wolisz, " A Radio over Fiber based Wireless Access Network Architecture for Rural Areas," Proceedings of 14th IST Mobile and Wireless Communications Summit, jun 2005.

3

Radio Resource Management

This chapter introduced the concept of radio resource management and common radio resource management. The features and functionalities of RRM and CRRM are explained. Then a framework called AAA for join allocation, assignment and admission control is discussed. This framework performs three important features of RRM, which are related to each other. Hence considering them in a single framework is a next step towards effective RRM. The system model for the validation of AAA framework is presented and one of the important module of the system model, PBS, which provides fairness among users is presented.

3.1 Introduction

Radio resource management (RRM) provides a set of methods for the efficient usage of radio resources in the wireless and cellular networks [1]. Radio is defined as the wireless transmission of electromagnetic signals through space. Radio resources are defined as the resources used for transmission of these electromagnetic signals in space. The resources for transmission of electromagnetic signals in space are frequency, space, time and code. As the resources are scarce in nature, they are used by multiple users either simultaneously or at different times. Hence multiple-access is defined as method of sharing the resources by multiple simultaneous or non simultaneous users. The re-known schemes for multiple access are time division, frequency division, code division and orthogonal frequency division multiple access [2]. In each of the access schemes the frequency, time, code or space are shared by multiple users.

Radio resources are used by users for different kinds of services with each service having its own requirements. While transmitting the radio signals on radio resources, the signal power is varied depending on the type of modulation scheme used, position of the user from transmitting BS, receiver sensitivity, fading conditions and many other factors. Each user transmits or receives on these radio resources with different power levels and when two users transmits/receives in same frequency at the same time hence they interfere with each other [3].

Also due to the increase in the number of users and increase in the number of BSs the number of interferers in the uplink and downlink increase and hence the overall interference. The type of interference varies depending on the type of radio resource and the multiple access scheme. In CDMA multiple access scheme, the interference is mainly from the users in the own cell and this type of interference is called intra cell interference [3] . This type of interference is low in other access schemes. In case of FDMA or OFDMA the interference seen is mainly from other cells due to simultaneous usage of the same frequency, which is called inter cell interference (ICI). Also in FDMA schemes the interference seen as a leakage power from other frequency bands or sub channels is called adjacent channel interference (ACI). In case of OFDM system the interference seen in time domain due to jitter in the clock is called as inter symbol interference (ISI). These interferences can be minimized by proper utilization of resources, and by understanding the network conditions and user conditions. Hence the RRM deals with methods for proper utilization of resources in time, frequency, and space.

The proper utilization of radio resources can be done in many ways depending on the conditions prevailing and the goals to be achieved at the end. The goals to be achieved are normally defined by the operator using business policies. The conditions prevailing could be BW, carrier frequency, RAT, multiple access scheme, load of the cell, number of users in each cell, interference conditions from other cells and own cells, type of applications used by the users, fading conditions of the users, pathloss and shadowing of the users which normally depend on the geographic location, number of antennas supported at transmitter and receiver etc [4]. RRM techniques consider these prevailing conditions and plans for usage of resources such that capacity of the network is maximized and the users are satisfied by achieving their intended QoS.

Some of the functionalities of RRM are interference control, handoff, power control, AC, congestion control, cell selection based on the signal conditions and geographic location of the user etc [1]. The process of assigning a BS is called cell selection, which is done initially when the mobile is powered on. Once the user is assigned a BS, the user can be handed off to other BSs, which is taken care by RRM. Whenever a user is assigned to a BS, the user needs to be admitted, by checking the network conditions and user conditions. The process of admitting or rejecting a user is called AC.

Another important functionality of RRM is allocation of resource to each user and to each BS in the uplink as well as downlink [5]. The allocation of resources depends on the prevailing conditions, mentioned above. The process of allocating resource dynamically is called dynamic resource allocation, which can be changed on time basis or event basis. While allocating resources in downlink or uplink, the power control is also another feature that needs to be taken care by RRM. In fact the resource allocation and power allocation need to be addressed as a single problem as both are dependent on the signal or fading conditions and the location of the user from BS. For example the resources and power for users near cell edge are to be allocated with care as they are sensitive to interference. Apart from the above discussed features, other features of RRM are power control, link adaptation, congestion control etc.

This Chapter is organized as follows. In Section 3.2, the introduction to CRRM is given. In Section 3.3 a AAA framework for joint allocation, assignment and admission control is proposed. In Section 3.4, the priority based scheduler (PBS) for selecting users to be scheduled in the next frame is presented. The methods used for allocation, assignment and admission control for AAA framework along with PBS are validated in Chapter 4, Chapter 5 and Chapter 6 respectively.

3.2 Introduction to CRRM

Next generation wireless networks are envisioned to be heterogeneous, with multiple RATs existing in a single geographical area. Each RAT may support a specific standard like WLAN, Worldwide Interoperability for Microwave Access (WiMAX), LTE, UMTS etc [6]. Each RAT has some specific characteristics like BW, multiple access schemes, transmit/receive power etc. Hence the RRM becomes more challenging. All the features of RRM, like AC, congestion control, power control etc are also done by CRRM. Apart from these features, vertical handoff is an important feature of CRRM which is a mechanism to handoff a call from one RAT to another RAT. The advantage of addressing these issues in CRRM is the availability of information from overall network. Hence the decisions taken by CRRM are more effective from overall network point of view.

In 3rd Generation Partnership Project (3GPP), the CRRM mechanisms are proposed to be mainly used between UMTS and GSM. But these mechanisms can be extended to any heterogeneous wireless scenario. In 3GPP, two entities are considered for RRM, namely LRRM and CRRM [7]. LRRM entity resides in the RNC for UMTS and it is responsible for the management of resources of a specific cell/RAT/BS in a RAN. CRRM entity is the responsible for the coordination and management of overlapping/neighbour RATs controlled by different LRRM entities. Each CRRM entity controls a number of LRRM entities and may communicate with other CRRM entities, which is useful for collecting information about other LRRM entities that are not under its direct control.

In [7] three CRRM topologies are presented: CRRM integrated into every RNC/BSC, CRRM integrated only in some RNC/BSCs, and CRRM as a stand-alone server. The important factor to be considered in the topology of LRRM and CRRM is the overhead caused due to the exchange of information. Also the delay or latency cause due to the reporting of information by LRRM to CRRM and waiting for the decision from CRRM. This design criteria need to be kept in mind while developing the topology of LRRM and CRRM. If the delay is the main criteria, then distributed architectures are the better choice as the decisions are taken locally. But the amount of information available is less in centralized architecture and hence the decisions taken with minimum information may not be as efficient as in centralized management. Also hybrid architecture can be used with both distributed and centralized capabilities, which is a tradeoff between the both.

The interactions between LRRM and CRRM involve mainly two types of functions: Information report and RRM decision support functions. Also two interfaces are envisioned in this architecture, first one is between the CRRM and LRRMs that are under the control of CRRM and the other interface is between CRRMs. The interface between LRRM and CRRM support both the functions, where as the interface between CRRMs support only Information reporting functionality. The information report functionality is used when LRRM is requested by its responsible CRRM entity to report certain information with respect to its radio resources. The information reporting function can support immediate reply on request, or it can be trigger or event based or periodic reporting of information. When a CRRM receives information from its corresponding LRRMs, it takes decisions and RRM decision support function helps CRRM in influencing its decision.

The goals of RRM or CRRM can be varied and depend on the business policies. The most common goals are to improve the overall throughput of a cell in case of RRM or a network with heterogeneous RATS in case of CRRM. Another goal is to provide QoS requested by the user and maintain it throughout the call [8]. Apart from these, the other goals depend on the RRM mechanism

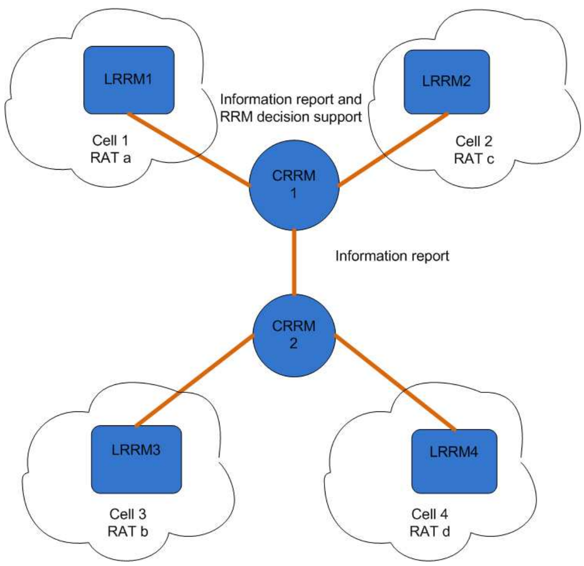


Figure 3.1: 3GPP CRRM functional model

applied. In case of resource allocation the goal can be to minimize the interference in the network. In case of AC the goal is to decrease the dropping or blocking probability of new users and doesn't violate the service level agreements of existing users. From an operator's point of view, the goal of CRRM could be to increase the number of users in the cell/network, so that the revenue can be increased. Some of the goals are dependent on each other, for example minimizing the interference in the network increase the throughput. Hence the design and choice of RRM mechanism should be done by considering the prevailing conditions and also by keeping in the mind the goal to be achieved by applying the RRM mechanism [9].

3.3 AAA framework

In this section, a new framework for allocation, assignment and admission control (AAA) [10] is proposed as shown in Figure 3.2. The main objectives of the proposed framework are:

- Dynamic and autonomous allocation of resources to each base station and to each user in the downlink by considering ICI or SINR, load of the network, location or downlink transmit power of the user etc. such that the overall network throughput is maximized.
- Assignment of resources to each user, which includes estimating the number of resources based on QoS requirements of the user, type of user, targetSINR, fading conditions, etc such that the QoS requirements of the user is met.
- Admission of a new or handoff user such that the network does not congest, QoS of existing users is not violated and QoS of the new user is achieved.

For dynamic resource allocation, the centralized architectures perform well at the expense of increased signalling overhead, where as the distributed approach has less signalling overhead at the cost of decreased performance. Hence a semi-distributed or hybrid approach is chosen, which is a tradeoff between signalling overhead and performance. The allocation of resources in the network is done at two time scales, super-frame and frame. The central allocation (CA) is done for every super-frame by predicting the SINR of the next super-frame from path loss and shadowing. Based on the predicted SINR, the CA allocates resources to each cell and to each user for the next super frame. The CA module receives inputs from the local allocation (LA) module, such as

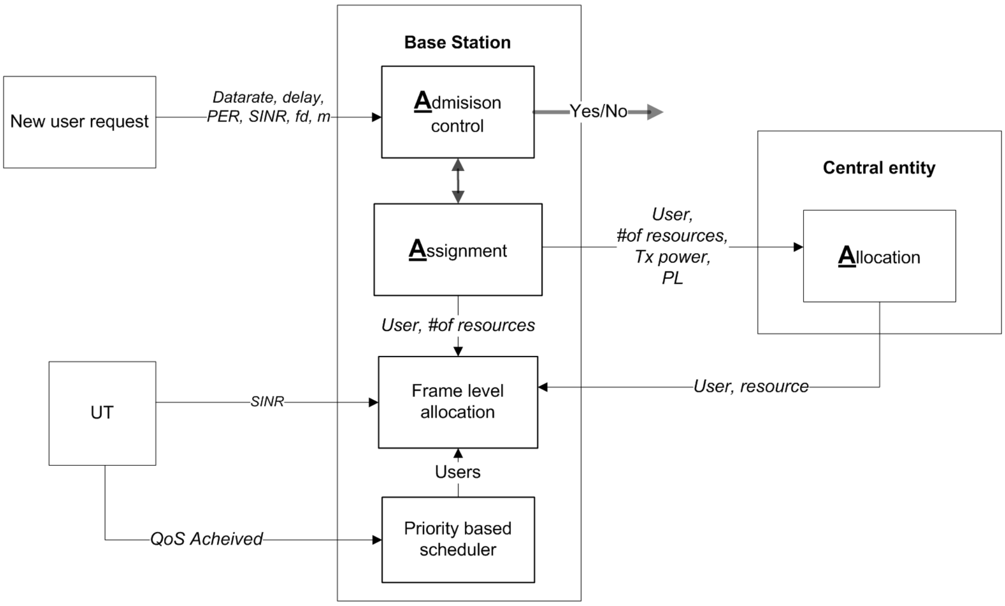


Figure 3.2: AAA Framework

transmit power and path loss from the assignment module, such as amount of resources to be allocated to each user. The SINR of the next super-frame is predicted from these inputs and, based on the predicted SINR the CA is performed.

To reduce the signalling overhead and to reduce the complexity, the CA is performed once in a super frame. The goal of CA is to improve the overall network throughput by reducing the interference and to improve the cell-edge throughput by providing lower reuse factors for cell-edge users compared to cell-center users. Due to the availability of data from all the cells in the network, the decisions taken by the CA are more effective from an overall network point of view.

The LA allocates the resources to each user in the particular cell, based on the channel and traffic conditions of the users, which is done in every frame. The CA recommends the resources for each user, but the LA may change this based on fading or buffer conditions of the user. The inputs received by the LA module from the other modules are

- The users to be scheduled in the next frame from the PBS
- The number of resources for every user from the assignment module
- Recommended resources to be used by the CA.

The LA allocates resources to the users based on the SINR reported by the user, such that the overall throughput of the cell is maximized. Unlike in the CA, the SINR values used in the LA are the measured SINR values by the user in the previous frame.

The next module of the AAA framework is the assignment module, which is triggered in case of a new user or a handoff user, where the admission control module sends the request from the user with the QoS requirements and channel conditions to the assignment module. The assignment module predicts the number of resources required by the user such that its QoS is met and it updates this information to the CA module once every super frame, and to the LA module in every frame.

The last module of AAA is admission control, which deals with admission/rejection of a new or handoff user. For each new request, the assignment module calculates the number of resources required by the user and forwards it to the admission control module. The admission control module estimates the mean number of resources used by existing users by taking advantage of multi user diversity based on buffer conditions. Based on the mean number of resources used by existing users and the number of resources required by a new user from the assignment module, admission to the user is performed. The mean resource method of admission control therefore increases the number of admitted users in the system, without violating QoS of existing users, and hence decreases the dropping probability.

Another important module of the system model in Figure 3.2 that interfaces with AAA framework is the PBS. The PBS selects the users to be scheduled in every frame based on the priority and sends this to the assignment module. The user terminal sends the QoS achieved to the PBS and, based on the level of user satisfaction, the PBS estimates the priority for each user. The users with highest priority is scheduled first in the next frame. By scheduling the users with least satisfaction some amount of fairness is obtained.

We consider three important aspects of resource management, namely allocation, assignment and admission control in a single framework. The interdependencies between the three main modules of AAA, which includes a semi-distributed CA and LA, assignment, admission control and PBS modules are explained. The proposed AAA is a generic framework which can be used for any radio access technology (RAT) or for heterogeneous RATs. The allocation, assignment and admission control modules and the methods applied in each module are explained more detail in Chapter 4, Chapter 5 and Chapter 6 respectively.

3.4 Priority Based Scheduler

The main function of the PBS [8] is to schedule the users based on the estimated priority, so that the user with highest priority is scheduled first. The amount of resources to be allocated to each user is estimated by the admission control algorithm based on the achieved QoS. For each user a satisfaction index (SI) is calculated, which gives the level of user satisfaction, and indicates how throughput and delay of a user is achieved w.r.t. the desired values. The desired values of QoS are assumed to be dependent on the type of service. From the SI values, the PBS calculates the priorities for each user, and sends them to the LA module. The SI is represented as a function of delay $\Gamma_u(t)$ or as a function of rate $\Psi_u(t)$ [11]. In either case, the lower the SI, the higher is the priority user will be assigned.

The delay component SI is expressed referring to the head of line (HOL) delay ω_u, which is the longest delay experienced by a packet at the HOL, and the maximum delay for service u, $T(u)$, as shown below

$$\Gamma_u(t) = \begin{cases} \frac{T(u)-\Delta T(u)}{\omega_u(t)}; & \text{if } \omega_u(t) < T(u) - \Delta T(u), \\ 1; & \text{otherwise.} \end{cases} \tag{3.1}$$

where $\Delta T(u)$ is a safety margin.

The rate component is expressed in terms of the average rate measured, η_u, and the desired data rate, $\hat{\eta_u}$:

$$\Psi_u(t) = \frac{\eta_u}{\hat{\eta_u} - \Delta\hat{\eta_u}} \tag{3.2}$$

where $\Delta\hat{\eta_u}$ is the margin coefficient.

The priority function has two components Φ_u^{rt} and Φ_u^{nrt}.

$$\Phi_u = \omega_u^{rt}\phi_u^{rt} + \omega_u^{nrt}\phi_u^{nrt} \tag{3.3}$$

where the weight coefficients ω_u^{rt} and ω_u^{nrt} are determined based on the service type.

The expressions for Φ_u^{rt} and Φ_u^{nrt} are

$$\Phi_u^{rt} = \begin{cases} R_u \frac{1}{\Gamma_u(t)}; & \text{if } \Gamma_u(t) \geq 1; R_u \neq 0, \\ 1; & \text{if } 0 < \Gamma_u(t) < 1; R_u \neq 0, \\ 0; & \text{if } R_u \neq 0 \end{cases} \tag{3.4}$$

$$\Phi_u^{nrt} = \begin{cases} R_u \frac{1}{\Psi_u(t)}; & \text{if } \Psi_u(t) \geq 1; R_u \neq 0, \\ 1; & \text{if } 0 < \Psi_u(t) < 1; R_u \neq 0, \\ 0; & \text{if } R_u \neq 0 \end{cases} \tag{3.5}$$

Using the above equations, the calculation of the priority function Φ_u is performed. From the above calculations, the users with least satisfaction is giver higher priority and scheduled first by which a fairness is achieved for each user.

3.5 Conclusions

In this Chapter the concept of RRM and CRRM are introduced. The problems addressed by RRM/CRRM and their functionalities, the architecture of CRRM in 3GPP is explained. Then a framework which address three important issues of radio resource management, dynamic resource allocation, estimation of resource to each user to satisfy its QoS requirements and admission control, is proposed. A semi-distributed or hybrid approach is chosen for dynamic resource allocation, to provide a tradeoff between overhead and the performance of allocation. The assignment module finds the minimum required number of resources for every user to satisfy its QoS. The admission control module admits or rejects a new user, with a goal to decrease the blocking or dropping probability of a new user or handdoff user. Finally the PBS improves the fairness of the users in the system. Hence this framework provides the overall improvement in the performance of the network, QoS and fairness of the users.

References

[1] http://en.wikipedia.org/wiki/Radio_resource_management

[2] Leijia Wu, Kumbesan Sandrasegaran, "A Survey on Common Radio Resource Management," 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, pp. 66-71, Aug. 2007.

[3] Andreas Mader, Dirk Staehle, "Interference estimation for the HSDPA service in heterogeneous UMTS networks," in EuroNGI Workshop on Wireless and Mobility, vol. 3883, pp. 145-157, 2006.

[4] Albena Mihovska, Tragos Elias, Mino Emilio, Luo Jijun, Mensing Christian, Fracchia Roberta, Horrich Sana, Hui Lin, Klockar Annika, Kyriazakos Sofoklis, " Requirements and Algorithms for Cooperation of Heterogeneous Radio Access Networks," Journal on Wireless Personal Communications, Springer, DOI: 10.1007/s11277-008-9586-y, Oct. 2008.

[5] I. Katzela, M. Naghshineh, "Channel Assignment schemes for cellular mobile Telecommunication systems: A comprehensive survey" IEEE Communications Surveys & Tutorials, vol.3, no.3, pp. 10-31, 2000

[6] FUTON Project Deliverable D4.1 "CRRM Functional Architecture and Interference Scenarios on FUTON Project" Dec. 2008, http: www.ict-futon.eu

[7] 3GPP. Improvement of Radio Resource Management (RRM) across RNS and RNS/BSS (Post Rel-5). TR 25.891, 3rd Generation Partnership Project, February 2003. Available from World Wide Web: http://www.3gpp.org/ftp/Specs/html-info/25891.htm [cited June 2008]. Release 6.

[8] M. V. Ramkumar, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad, "QoS-Guaranteed Admission Control for OFDMA-based Systems," Globecom, pp. 606-610, Dec 2010.

[9] M. V. Ramkumar, Albena Mihovska, Neeli R. Prasad, ”Management of Distributed and Extendible Heterogeneous Radio Architectures,” NAEC2009, Lake Garda, Italy, October 8-11, 2009

[10] M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, ”An Allocation, Assignment and Admission Control (AAA) Framework for Next Generation Networks,” submitted in International journal of Network management, John Wiley and Sons, Ltd.

[11] Qingwen Liu, Xin Wang, G. B. Giannakis, ” A cross-layer scheduling algorithm with QoS support in wireless networks,” IEEE Transactions on Vehicular Technology, vol. 55, no. 3, pp. 839-847, May. 2006.

4

Radio Resource Allocation

In this chapter the dynamic allocation of resources to basestation and in-turn to the users in the downlink is discussed. Here the allocation works in a semi-distributed fashion with central allocation (CA) and local allocation (LA) modules. A new heuristic method is proposed for CA to allocate resources to each cell in a network and in turn to users. The LA is based on the SINR reported by the user, which is done at the base station in every frame. The CA with LA is validated along with other modules of AAA framework and its observed that the over all network throughput and cell edge throughput improves compared to the state of art. The overhead analysis of the proposed method for different scenarios is also discussed in this chapter.

4.1 Introduction

Due to an exponential increase in the number of mobile terminals and variety of applications [1], next generation cellular networks are facing a challenge of satisfying QoS requirements of the users while improving the efficiency of the overall network. The main challenge is to efficiently provide a variety of services to the users in a wireless network, which is highly dynamic due to fading, user mobility, traffic conditions, network conditions etc. Also as bandwidth in wireless networks is a scarce resource, the available bandwidth has to be reused by various cells in a network. The overlap of resource usage by various cells in a network cause inter-cell interference (ICI), which decrease the Quality of service (QoS) and also cell throughput [2]. This problem is especially distinct for cell edge users, as the transmit power is high from the serving cell and interference is high from neighboring cells. Also ICI changes due to the mobility of users and changes in radio environment. Hence with all these challenges, there is a strong need for efficient management of resources in the next generation networks.

The management of resources includes, among all, allocation of resources to the cells in a network and in turn to the users, which deals with finding the right resources to be used by each cell and by each user [3] [4]. The allocation of resources should be done such that the overall throughput of the network is maximized and QoS of user is achieved. Prior to performing the allocation, the amount of resources to be allocated to each cell and to each user should be known, which depends on the load of the cell.

During allocation, if the resources are allocated orthogonally without any overlap between cells, ICI will be zero, but the overall throughput is not maximized. For 1:1 reuse the ICI is high which decreases the throughput, hence a tradeoff should be obtained on the percentage of overlap. Users near the cell edge are sensitive to interference as they are close to the neighbor cell and also due to their weak signal strength due to path loss, where as the users near to the BS are not as sensitive to interference as cell edge users. Hence the allocation of resources to each cell should take these effects into consideration.

In distributed allocation methods, each cell does not have much knowledge about neighbor cells hence the allocation performed independently by each base station may not be optimal [5] [6] [3]. The advantage of a centralized allocation is that it has knowledge of the overall network like load in each cell, user distribution in each cell etc [3]. However, the centralized allocation has disadvantages in terms of signaling overhead [5]. Therefore a hybrid approach could provide an optimized solution and tradeoff between the signaling overhead and optimal allocation [3].

The rest of this Chapter is organized as follows. Section 4.2 discusses the related work. Section 4.3 explains proposed heuristic method for allocation as a part of AAA framework proposed in Section 3.3. Section 4.4 presents the simulation setup and the results obtained in the validation of the proposed heuristic method as a part of AAA framework given in Section 3.3, along with assignment procedure given in Chapter 5, and mean resource based AC proposed in Chapter 6 respectively. Finally in the last section the chapter is concluded.

4.2 Related Work

The allocation of resources can be done in a static way using fixed reuse methods or in a dynamic way by considering the network condition. Dynamic channel allocation methods are less efficient than fixed allocation methods under high load conditions but provide more flexibility and traffic adaptability [5]. In fixed frequency reuse (FFR), PRBs are allocated without overlap. Hence, the ICI is less at the cost of reduced spectral efficiency. In [7] [8] [9], users in a cell are classified into different classes based on their geometrical position and different bandwidth allocation patterns are assigned for different user classes. The most promising approach divides the users into two groups: interior cell-center users and exterior cell-edge users. One third of the available bandwidth in each cell is fixed for cell edge users and the rest for the cell-center ones [7] [8], which is called soft frequency reuse (SFR). However, when the traffic load changes, it is desirable that the allocation of sub bands to cell-edge users does not have to be done statically, but in a dynamic way to take advantage of varying traffic load in the network, which is not addressed in [7] and [9].

In [8] an adaptive SFR scheme dynamically adapts to changing traffic load and user distributions among neighbor cells. In [10] [6], two methods for flexible spectrum usage (FSU) are proposed; spectrum load balancing (SLB) and resource chunk selection (RCS). In [11] partial frequency reuse (PFR) based on network load is proposed. In [3], as hybrid method for dynamic resource allocation is proposed. In all the above methods the allocation in the current frame is done based on the SINR measurements in the previous frame, which doesn't guarantee the optimal allocation in the current frame.

4.3 Heuristic Allocation Method

4.3.1 Central Allocation

In previous works [2] [4] [3], dynamic resource allocation is done based on the SINR measured by the user from the previous frame or slot. But with new allocations the maximum throughput is not guaranteed as the SINR with new allocation changes completely and does not depend on the previous SINR or previous allocations. Even if the allocation converges or stabilizes after a few frames, this cannot be guaranteed as optimal as path loss and fading change due to user mobility, and thus interference changes overtime.

Hence a new allocation scheme is proposed, in which interference is predicted by the central entity located inside the radio access network (RAN) [13], based on the current allocation of users. The allocation module receives input from the assignment module with a list of users and DL transmit power for each user, number of resources that need to be allocated to achieve the QoS and the path loss of the user. The output of the allocation module is allocation of resources for each base station and the recommended resources for each user. If the user has enough packets to send in the buffer or is not in a deep fade, then the base station may follow the recommendation of the CA module, otherwise the base station takes its own decision about the users to be scheduled in each frame and the resources to be allocated. This decision, taken locally at the base station, depend on traffic conditions of the user, the channel fading conditions and the user satisfaction levels.

The interference faced by a user in the next frame is estimated from the channel gain which is based on the path loss of the user from its own base station and neighboring base station. The allocation of resources starts with users having the highest DL transmit power in the system, which are most probably cell-edge users and creates most interference in the system. Once the user is allocated with resources the effect of its allocation on other users in the system is calculated on the fly, based on which the next user's allocation is done.

The reason for allocating the far off users first is that the interference created by far off users is higher compared to the nearby users. This way of allocating resources, based on geographic position, ensures a low reuse factor (e.g. 1/3) for cell-edge users and high reuse factor (e.g. 1) for cell-center users. The increase in reuse factor from cell-edge to cell-center increases, depending on the load of the network and distribution of the load in the network.

The proposed heuristic method of CA is explained for an orthogonal frequency division multiple access (OFDMA) based LTE system. Consider a network consisting of L base stations and a set of M users, with users served by base station l denoted as M_l, where $M = \bigcup_{l=1}^{L} M_l$. The base station $l \in L$ that serves user $m \in M$ is denoted as $l(m)$. Let $G_{m,j}$ denote long term channel gain, which includes path loss and shadowing, from base station $j \in L$ to user $m \in M$. The downlink transmit power of user m from $l(m)$ is denoted as P_m.

The modulation used is OFDMA with K PRBs in each frame, with user m allocated with a set of k_m PRBs, of length $|k_m|$ obtained from the assignment module where

$$\sum_{m \in M} |k_m| \leq K \tag{4.1}$$

The goal of the CA module is to find k_m for all $m \in M$.

The method followed for CA is as follows. All users in the network, M, are arranged in decreasing order of their downlink transmit power P_m and they are allocated in the same order. Let the user to be allocated from the list be i. The interference PI_m^i due to user i on user m in the system is predicted as

$$PI_m^i = P_i.G_{m,j} \tag{4.2}$$

where $m \in M, m \neq i$

Hence the total interference I_i seen by user i, from all users in Y that were already allocated is calculated as in eq. (4.3). Initially $I_i = 0$ on all the PRBs and when the users are allocated, I_i is updated by adding the interference from the allocated users on their corresponding PRBs. Till all the PRBs are allocated once, each user gets unused PRBs and interference is zero. Once all PRBs are allocated once, the PRBs are reused, in which case the next users in the list face interference from the users that were already allocated with the same PRBs and viceversa.

$$I_i = \sum_{y \in Y} PI_i^y(k_y) \tag{4.3}$$

Let $SINR_i^{CA}$ be the SINR of user i on all K PRBs from already allocated users Y and the received power of i from its own base station. The length of the $SINR_i^{CA}$ vector and the length of I_i are equal to K.

$$SINR_i^{CA} = \frac{P_i.G_{i,j}}{Noise + I_i} \tag{4.4}$$

where $i \neq Y, j = l(i)$

Once $SINR_i^{CA}$ and PI_m^i are calculated for user i, a ratio R_i is defined on each PRB which is the $SINR_i^{CA}$ of user i on each PRB divided by the total interference exerted by i on other users in the system on each PRB:

$$R_i = \frac{SINR_i^{CA}}{\sum_{m \in M, m \neq i} PI_m^i} \tag{4.5}$$

where $i \neq Y$

The PRBs to be allocated for user i, k_i, are selected, for which the ratio R_i is maximized, such that the user gets the best PRBs having good SINR conditions, and the total interference exerted by user i on the system is minimum. Once k_i is found, the interference due to user i on other users in M on k_i PRBs is updated, and Y is updated with i. For the next user in the list the same procedure is repeated, till all the users are allocated.

This method reuses the PRBs if the load in each cell increases, such that the interference from the users already allocated is minimum and the interference created by this user on other users in the system is also minimum. Hence this method guarantees, that when PRBs are reused, the users in the cell-center gets lower reuse factor and the users near the cell edge gets higher reuse factor. The proposed method can be applied to cells with uneven loads and uneven distribution of load in the cell. This heuristic approach improves the overall throughput of the network and guarantees cell-edge throughput, which is verified from simulation results shown in Section 4.4. This approach is suitable in a network with dynamic variation of traffic conditions and network load.

4.3.2 Local Allocation

Due to the signaling overhead, the CA module takes the path loss and shadowing into consideration but not the fading effects of users. Hence this module takes advantage of channel conditions, by allocating resources having high SINR to users. Even though the CA module in the RAN recommends the resources to users in every super frame, the LA may change the allocation by considering changes in signal conditions. This module takes the input from the PBS and from the assignment module and allocates resources to each user. The allocation of resources to the users is based on the $SINR_{i,k}^{LA}$ reported to the LA module of user i on PRB k, which takes channel gain due to fading into consideration. Hence the goal of this module is to maximize the overall $SINR_{i,k}^{LA}$ of the base station by allocating resources to the users having the best $SINR_{i,k}^{LA}$ conditions, which maximizes the overall throughput of the base station.

eq. (4.6) explains the SINR measured by the user where $h_{i,l(i)}^{k}$ is the channel gain due to fading between user i and its own base station $l(i)$ on PRB k. $|M_l|$ is the total number of users to be scheduled in base station l with user i allocated with $|k_i|$ PRBs and P_l^k is the downlink transmit power from base station l on PRB k.

$$SINR_{i,k}^{LA} = \frac{P_{l(i)}^{k}.G_{i,j}.h_{i,j}^{k}}{Noise + \sum_{l \in L, l \neq l(i)} P_l^k.G_{i,l}} \tag{4.6}$$

The goal of the LA is to find k_i PRBs for user i such that the overall $SINR_{i,k}^{LA}$ in each base station is maximized, as shown in eq. (4.7).

$$max\left(\sum_{i=1}^{|M_l|}\sum_{k=1}^{k_i} SINR_{i,k}^{LA}\right) \tag{4.7}$$

4.4 Simulation Results

4.4.1 Simulation Setup

The simulation setup used here is 4-Cell network, with users distributed randomly in each cell. LTE-TDD is used with 10MHz bandwidth at 3.5GHz center frequency. The number of PRBs used in each cell depends on the load of the cell and PRBs are distributed equally among users in the cell. The users are moving with 3 Kmph speed within the cell with 250m radius in urban scenario. The pathloss model used for urban scenario is C2 NLOS developed in Winner [14].

$$\begin{aligned} PL = &\big(44.9 - 6.55log_{10}(h_{BS})\big)log_{10}(d) \\ &+ 34.46 + 5.83log(h_{BS}) + 23(f_c/5) \end{aligned} \tag{4.8}$$

The transmit power of the base station is 43dBm, which is equally distributed on all PRBs. In each frame SINR is measured on each PRB by calculating the received signal from current base station and interfering signal from neighbor base stations on each PRB. Based on the measured SINR, LA assigns PRBs to each user, such that the overall SINR in each cell is maximized. Central allocation module allocates PRBs to each cell, based on the predicted SINR, for every super frame. Here we assume super frame consists of 20 frames. Proposed CA+LA is compared with existing method [3] in the literature which is also a hybrid approach, with LA alone and Random allocation. Basestation transmit to each user on the allocated PRBs, based on which SINR is measured in each frame and mapped to the Shannon throughput. The simulation parameters used in the simulation are given in the Table 4.1.

4.4.2 Cell throughput

Figure 4.1 shows the mean cell throughput and average cell edge throughput of the network with increasing load. The load is varied by varying the number of PRBs used in the cell. The existing hybrid method is compared with the proposed CA+LA method, and it can be seen that at 70% and 80% load, the

Table 4.1: SIMULATION PARAMETERS.

Parameter	**Value**
Carrier frequency fc	3.5Ghz
Deployment scenario	Urban macro
Intersite distance	500m
Pathloss model	C2 NLOS [14]
Mobility	3 Kmph
Bandwidth	10 Mhz
Antenna	Omni directional
DL Tx power	43 dBm
UL Tx power	24 dBm
targetSINR	15 dB
Noise figure	9dB
Buffer length K	[15, 30, 60] for Type 1,2,3
Packet length	84 bits
Packet arrival rate, λ	[3, 5, 12] for Type 1,2,3
Number of AMC modes	BPSK, 4, 8, 16, 32, 64 QAM
SINR thresholds [dB]	[7.2 10.1 12.5 16.1 18.7 22.2]

proposed approach performs around 2Mbps or 14% better than the existing method.

Also with CA+LA the performance of mean cell throughput is around 4Mbps better than LA alone at 60 and 70% loads. Central allocation recommends the PRBs to be used by each cell for every super frame, as load reaches 100% CA has to recommend all PRBs to each cell; hence the performance of CA+LA and LA alone merge at 100% load. At lower loads as the effect of interference is low, the difference in the performance between CA+LA and LA alone is less when compared to higher loads. Also the CA and LA are compared with random allocation. It can be seen LA alone and CA+LA performs around 5Mbps better than random allocation at 100% load. At 90% load CA+LA performs 6Mbps better than random allocation, where as LA alone performs 4 Mbps better than random allocation. The cell edge throughput is defined as outage throughput of users below 5%. The performance of cell edge user for the proposed approach is better by 125kbps or around 30% compared to the existing method [3] at 60% load. As explained above the difference in performance is better at medium loads compared to 100% load, where the performance of CA+LA and LA alone are same. At 90% load proposed CA+LA performs around 400 Kbps better than random allocation, where as LA alone performs 300 Kbps better than random allocation.

4.4.3 Overhead Analysis

CA module receives inputs from different basestations in the network. The amount of overhead information sent by each BS is estimated for four scenarios. In scenarios 1 and 2 less information, which includes userids, number of resources, DL transmit power, pathloss, is sent once for every frame and super-frame respectively. In scenarios 3 and 4 full information, which includes the channel gain information on all the PRBs apart from information in scenario 1 or 2, is sent once for every frame and super-frame respectively. eq. (4.9) and eq. (4.10) shows overhead with less information for scenarios 1 and 2, OV_{less},

Table 4.2: Overhead comparision

Scenario	Overhead in kbps
1 - less info every frame	920
2 - less info every super-frame	46
3 - full info every frame	6400
4 - full info every super-frame	320

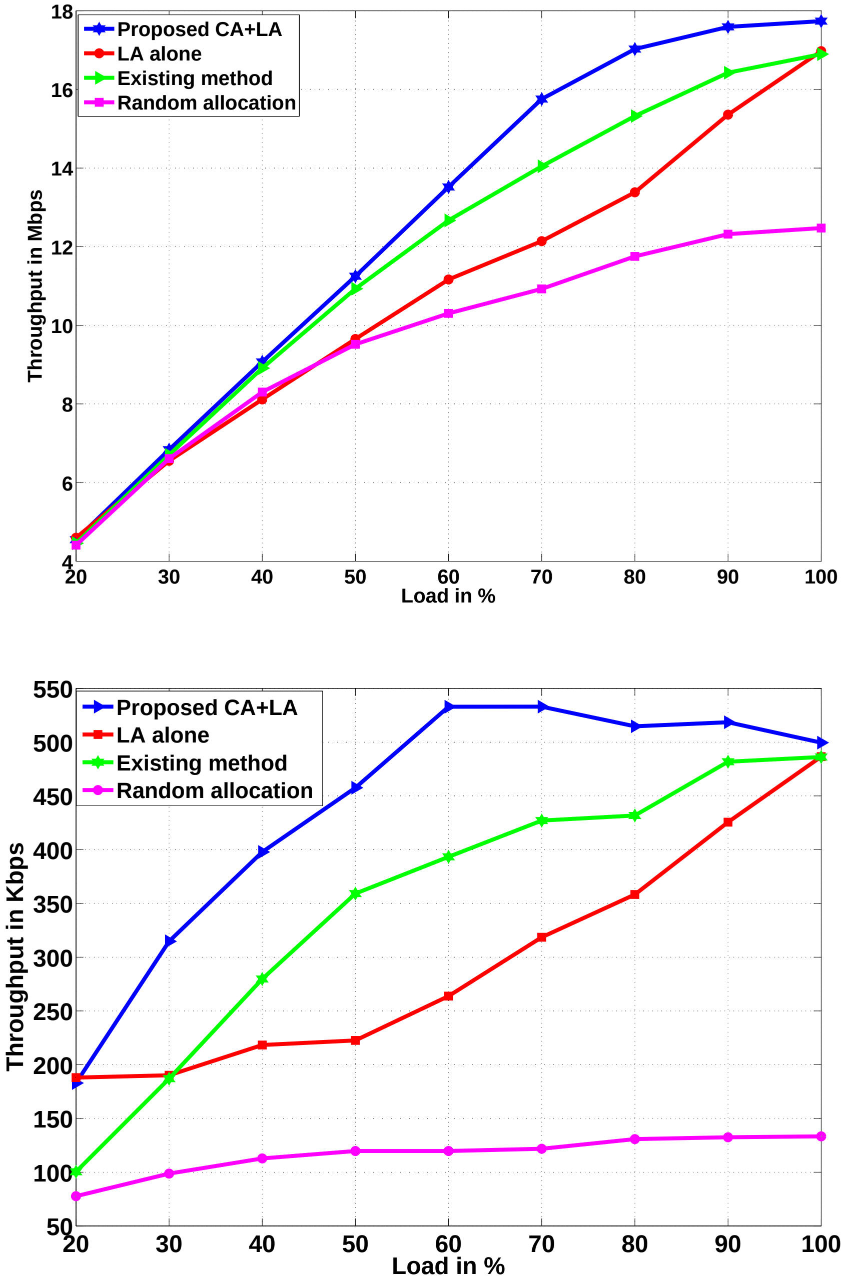


Figure 4.1: Comparison of Mean Cell throughput and Cell edge throughput

and overhead with full information for scenarios 3 and 4, OV_{full} respectively. Here it is assumed that 64 bit IMSI (International mobile subscriber identity) is used as user id and 32 bits are used for floating point representation of pathloss and DL transmit power.

$$OV_{less} = 64 * |M| + floor(log_2(K)) * |M| \\ + 32 * |M| + 32 * |M| * |L| \tag{4.9}$$

$$OV_{full} = 64 * |M| + floor(log_2(K)) * |M| \\ + 32 * |M| + 32 * |M| * |L| + 32 * |M| * K \tag{4.10}$$

The overhead values calculated with parameters used in the simulation for four scenarios are shown in Table 4.2. It can be seen that scenario 2 has least overhead of 46kbps, which is considered in this framework. The overhead of scenario 3 is higher compared to other three scenarios, which is impractical from implementation point of view. In scenario 2, though there is a overhead of 46 kbps due to CA, the proposed method gives 4Mbps improvement in overall cell throughput when compared to LA alone. Hence the overhead of 46kbps can be justified.

4.5 Conclusions

The proposed heuristic allocation method is applicable for any next generation cellular network. This approach is suitable for networks with dynamic network conditions because of mobility, handoffs etc. Also the proposed method is suitable for networks with various load distributions with in a cell and guarantees improved overall network throughput and cell edge throughput compared to the state of art methods. Because of its semi distributed approach this method is a tradeoff between the efficiency and overhead. For the case of overhead, four scenarios are investigated and scenario 2, where basestation sends less information every super frame has least overhead compared to other scenarios. The other methods can be chosen to improve efficiency at the cost of overhead and delay.

References

[1] M. V. Ramkumar, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad, "QoS-Guaranteed Admission Control for OFDMA-based Systems," Globecom, pp. 606-610, Dec 2010.

[2] S.E. Elayoubi, O. Ben Haddada, B. Fourestie, "Performance evaluation of frequency planning schemes in OFDMA-based networks," IEEE transactions on wireless communications, vol. 7, no. 5, pp. 1623-1633, May. 2008.

[3] Chrysostomos Koutsimanis, G. Fodor, "Dynamic Resource Allocation Scheme for Guaranteed Bit Rate Services in OFDMA Networks," International Conference of Communication Workshops ICC, pp. 2524-2530, May. 2008.

[4] Guoqing li, Hui Liu, "Downlink radio resource allocation for multi-cell OFDMA system," IEEE transactions on wireless communications, vol. 5, no. 12, pp. 3451-3459, Dec. 2006.

[5] I. Katzela, M. Naghshineh, "Channel Assignment schemes for cellular mobile Telecommunication systems: A comprehensive survey" IEEE Communications Surveys & Tutorials, vol.3, no.3, pp. 10-31, 2000

[6] S.Kumar, Yuanye Wang, N.Marchetti, I.Z.Kovacs, K.I.Pedersen, P.E.Mogensen, "Spectrum Load balancing for Flexible spectrum usage in Local area deployment scenario," 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN, pp. 1-5, 2008.

[7] E3 project deliverable

[8] Xuehong Mao, A. Maaref, Koon Hoo Teo, "Adaptive Soft Frequency Reuse for Inter-Cell Interference Coordination in SC-FDMA based 3GPP LTE Uplinks," IEEE Globecom, pp. 1-6, Nov-Dec. 2008.

[9] R1-050507, "Soft frequency Reuse scheme for UTRAN LTE"; Huawei, 3gpp TSG RAN WG1 Meeting 41, Athens , Greece, May. 2005.

[10] G.Monghal, S.Kumar, K.I.Pedersen, P.E.Mogensen, "Integrated Fractional Load and Packet Scheduling for OFDMA Systems," International Conference of Communication Workshops, ICC, pp. 1-5, 2009.

[11] B. Krasniqi, M. Wrulich, C.F. Mecklenbrauker, " Network-load dependent Partial frequency reuse for LTE," International Symposium on Commnucations and Information Technology. ISCIT, pp. 672-676, Sep. 2009.

[12] M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, "An Allocation, Assignment and Admission Control (AAA) Framework for Next Generation Networks," submitted in International journal of Network management, John Wiley and Sons, Ltd.

[13] IEEE 1900.4 "Architectural building blocks enabling network-device distributed decision making for optimized radio resource usage in heterogeneous wireless access networks".

[14] IST-4-027756 WINNER II, D1.1.2, WINNER II Channel Models part I, Sep. 2007.

5

Radio Resource Assignment

In this chapter the radio resource assignment, which is defined as estimating the amount of resources needed by a user to satisfy the Quality of service (QoS) requirements, is discussed. For this a Markov based approach is presented, which consider the effects of queue in MAC layer and AMC in physical layer and calculate the dropping probability of packets. From the dropping probability, the user throughput and delay are calculated and mapped to required throughput and delay to estimate the required number of resources. The assignment module along with AAA framework is validated on the LTE network and shown that QoS for new users and existing users is guaranteed.

5.1 Introduction

Due to an increase in data rates there is an increase in different types of applications, with each application having different QoS requirements. Also based on the geographic location of the user, the user can be in indoor or outdoor location in a rural, urban or suburban scenario. Depending on the type of scenario and location, the channel faced by the user is different.

Assignment module is a part of AAA framework [1] estimates the number of resources required by each user based on the traffic, channel and buffer conditions and QoS requirements. While performing the assignment, QoS requirements of a new user should be guaranteed without violating the QoS of existing users. The QoS achieved by a user depends on many factors, most importantly the channel conditions of the user which cause packet errors, buffer conditions such as buffer overflow, which cause packet drops or increase delay in packet delivery [2]. Both channel conditions and buffer conditions effect the throughput and delay of a user. Traditional queuing models did not consider the effects such as fading and pathloss and channel models didn't consider the effects of queue [3]. For example when the channel is in deep fade, AMC selects lower modulation order, which reduce the outflow of packets from the buffer and thus throughput of the user reduces. On the other hand, as the number of packets going out of the buffer reduces, the dropping rate of the packets increase, which increases the delay. Hence the QoS of the users is dependent on the channel and the queue characteristics.

In the AAA framework, the assignment module is triggered when a new user is admitted by the AC module. From the inputs received by AC, it models the queue as Markov process to find the dropping probability of packets from the queue. From the dropping probability P_d, it estimates the expected throughput and delay for different values of b, where b is the number of resources. The minimum value of b that matches with the requested QoS is selected. In the AAA framework, assignment module receives the input from the priority based scheduler, about the users to be scheduled in the next frame and it sends number of resources required by each user to the LA module. It also sends the user information, with the number of resources needed by each user, downlink tx power of the user and the pathloss of the user to the CA module every super frame as shown in Figure 3.2.

The rest of this Chapter is organized as follows. In Section 5.2 the method for assigning the resources to the users based on Markov modeling of the queue is explained. This method is validated as a part of the proposed AAA framework proposed in Section 3.3 along with the allocation and admission control methods proposed in Chapter 4 and Chapter 6 respectively.

5.2 Method used for assignment

In the assignment method, the estimation of the amount of resources takes in to account the AMC in physical layer and effects of the queue in MAC layer [2] [4] [1]. Each user is given b number of resources in each frame and the goal of the assignment module is to find suitable value of b that guarantees the QoS to the user. By using Markov analysis of the queue as shown in Figure 5.1, the assignment module estimates the probability of dropping a packet (P_d) and from this it estimates the throughput and delay for different values of b. The most suitable value of b that matches with the requested QoS is selected.

Each state of a Markov chain is defined as (U, C) where U is the number of packets waiting in the queue and C represents the number of packets transmitted in the next frame and U can range from 0 to K. The number of packets transmitted in a frame depends on the number of resources allotted to the user and the AMC mode of the user in that particular frame:

$$C_n = bR_n \qquad n = 1, 2...N \tag{5.1}$$

where b is the number of resources allotted to a user and R_n depends on the AMC mode. Hence C can take any value in $C_1, C_2, ...C_N$ and N is the number of AMC modes.

Each user is allotted a buffer and the size of the buffer K depends on the type or class of user. For high data rate applications like video conferencing, the buffer size K is large compared to low data rate applications like voice. The number of packets waiting in the queue depends on the arrival process A, the service process and the buffer length of the user.

The arrival process is modeled with a Poisson distribution:

$$P(a) = \frac{\lambda^a e^{-\lambda}}{a} \tag{5.2}$$

where $a \geq 0$ and $E(A) = \lambda$ is the packet arrival rate, defined as the average number of packets arriving during one frame duration depending on the traffic model.

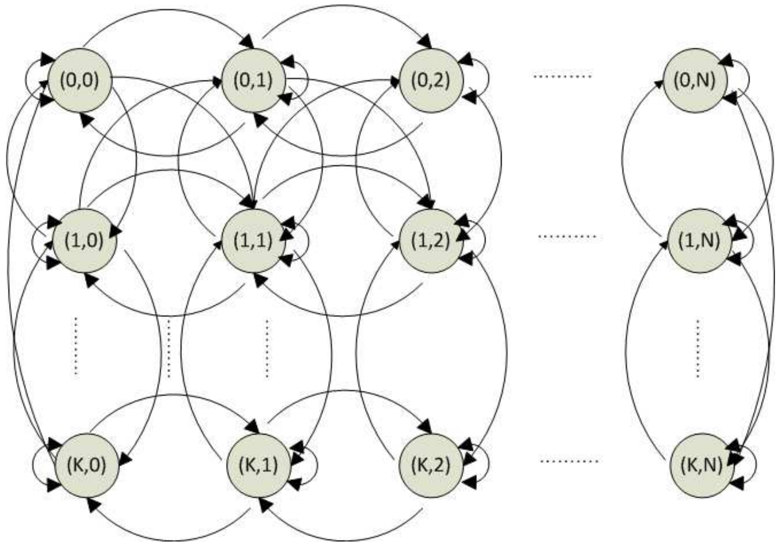


Figure 5.1: State transitions of two dimensional Markov chain of a queue

The service process depends on the AMC, and on the SINR of the user in the frame. The probability of the service process changing from one state to another depends on the transition probability of a user changing from one AMC mode to another, by assuming that the number of resources allotted to the user is fixed. In [5], SNR was divided into adjacent regions based on the desired BER. The transition probabilities between the various SNR regions were determined based on the level crossing rate (LCR) of the channel fading distribution.

Based on probability of packet arrival $P(a)$ from arrival process, and transition probability between AMC modes, the steady state distribution of 2-D markov chain $P(U = u, C = c)$ is calculated [6]. From $P(U = u, C = c)$, the expected number of packets dropped from queue $E(D)$ is calculated shown in eq. (5.3).

$$E(D) = \sum_{a\in A, u\in U, c\in C} max(0, a - K + max(0, u - c)) \\ P(A = a)P(U = u, C = c) \quad (5.3)$$

The dropping probability of a packet P_d from the queue is calculated from the expected number of packets dropped from the queue and expected arrival rate of packets in one frame duration, as shown in eq. (5.4), where T_f is the frame duration:

$$P_d = \frac{E(D)}{\lambda T_f} \tag{5.4}$$

From the above dropping probability P_d, the packet loss rate (PLR) which is defined as the probability that a packet is lost is calculated. The PLR is composed of two factors, PER and PDR as shown in eq. (5.5). The PER is the percentage of packets lost due to errors because of radio environment, noise, etc., and the packet dropping rate (PDR) is the percentage of packets dropped due to timeout in the queue or because of the finite buffer length K. The packet is assumed to be lost, when there are bit errors in the packet and/or when the waiting time of a packet in the buffer is more than a timeout and hence packet is dropped. PLR is calculated as shown below

$$PLR = 1 - (1 - P_d)(1 - P_0) \tag{5.5}$$

where P_0 is the PER due to channel fading and P_d is the dropping probability. From PLR, the prior throughput is estimated as:

$$\eta_{prior} = \lambda(1 - PLR) \tag{5.6}$$

From Little's Theorem [7], the expected prior delay is estimated as:

$$\tau_{prior} = \frac{N_w}{E(A)(1 - P_d)} \tag{5.7}$$

where N_w is the average number of packets waiting in the queue plus the average number of packets transmitted in one frame obtained from the steady state probability $P(U = u, C = c)$.

η_{prior} and τ_{prior} are calculated for different values of b. The minimum value of b is selected, for which the required throughput and delay of new user is greater than the estimated throughput and delay. Here it is assumed that delay in the transmission is only due to waiting time in the buffer, hence only P_d is considered for the delay calculations. The delay caused due to retransmissions because of CRC errors is not considered.

5.3 Simulation Results

The markov based method of assignment is validated using the system model for AAA framework explained in Section 3.3. For allocation and admission control, the methods proposed in Section 4.3 and Chapter 6.4 are used. The simulation setup used in the simulation is explained in Section 4.4.1.

Assignment module estimates the number of PRBs required by each type of user for a given targetSINR. We assume three types of users with data rates 400kbps, 800kbps and 2Mbps. For each type of user the number of PRBs required to obtain the required throughput is estimated by assignment module, using the parameters shown in Table 4.1. For AMC if the SINR is below 7.2dB, then there will be no transmission.

By using the above values the assignment module estimates the probability of dropping the packets for various values of b, where b is the number of PRBs allotted to each user. From the dropping probability the throughput is calculated. Figure 5.2 shows the average throughput of type-1, type-2 and type-3 users in the system with required throughput of 400Kbps, 800kbps and 2Mbps respectively. Red line shows the required throughput of the user, blue line shows the achieved throughput. It can be seen that on an average all users in each type obtain the required throughput. Hence the assignment module guarantees the QoS for the user. The users are admitted in to the system till the load reaches 100%. It can be seen that the average user throughput, of three types of users, is maintained above the required level till the load reaches maximum. Hence the assignment module guarantees the QoS of user in all load conditions.

5.4 Conclusions

In this chapter a markov based assignment method estimates the number of resources needed by a user to satisfy its QoS requirements. This method considers the traffic conditions and channel conditions of the user and estimates the dropping probability of packets from the queue. Here AMC is considered which ensures efficient bandwidth utilization. This method considers every QoS requirement of the user like datarate, delay, targetSNR, PER, hence it can guarantee QoS of the users, which is shown for a LTE network.

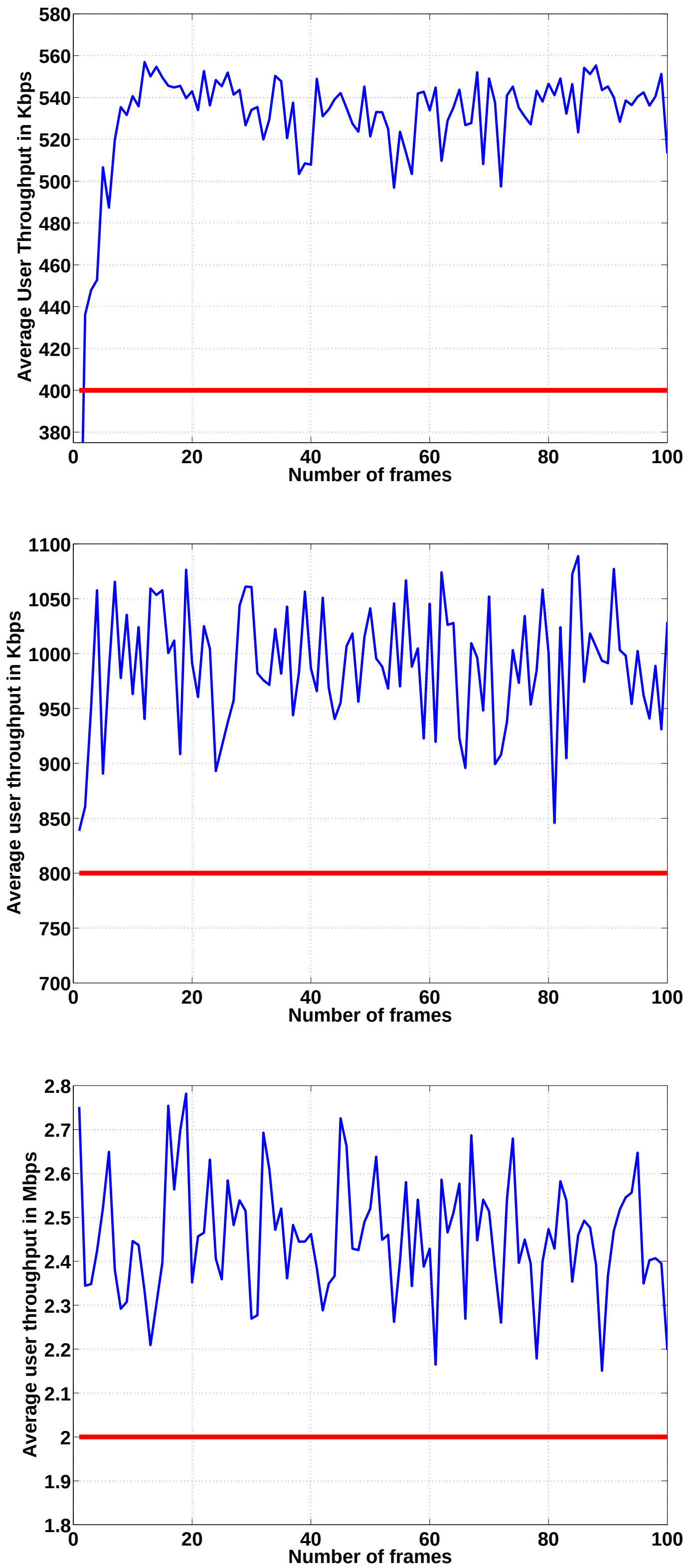


Figure 5.2: Average user throughput of Type-1, Type-2 and Type-3 users

References

[1] M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, "An Allocation, Assignment and Admission Control (AAA) Framework for Next Generation Networks," submitted in International journal of Network management, John Wiley and Sons, Ltd

[2] M. V. Ramkumar, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad, "QoS-Guaranteed Admission Control for OFDMA-based Systems," Globecom, pp. 606-610, Dec 2010.

[3] Qingwen Liu, Xin Wang, Giannakis GB. A Cross-Layer Scheduling algorithm with QoS Support in Wireless Networks. IEEE Transactions on Vehicular technology 2006, vol. 55, no. 3, pp. 839-847, DOI: 10.1109/TVT.2006.873832, 2006.

[4] M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, "Efficient Admission Control for Next Generation Cellular Networks," submitted in ICC 2012.

[5] Q. Zhang, S. Kassam, "Finite-state Markov model for Rayleigh fading channels," IEEE Transactions on Communications, vol. 47, no. 11, pp.1688-1692, Nov. 1999.

[6] Qingwen Liu, Shengli Zhou, Giannakis. G. B, "Cross-Layer Scheduling With Prescribed QoS Guarantees in Adaptive Wireless Networks," IEEE Journal on selected areas, vol. 23, no. 5, pp. 1056-1066, May. 2005.

[7] R. B. Cooper, Introduction to Queuing Theory, 2nd ed. Elsevier North-Holland, 1981.

6
Admission Control

This chapter discuss the management of radio resources for the case of heterogeneous and homogeneous radio environments. AC, which is one of the main features of RRM is considered and three approaches, fuzzy based AC, markov based AC and mean resource based AC, are presented for this problem. The results obtained for each approach are presented and discussed. Fuzzy based AC is applicable for heterogeneous network scenario and it also performs cell and RAT selection. In the second approach, a queue is modeled using a markov chain having different states depending on the traffic and channel conditions. Using this approach the QoS is guaranteed for different types of users. In the third approach, the mean resource based AC takes buffer conditions of the user and admits a new user by considering the mean resource usage of all the users. Hence using this approach the number of users admitted in to the system is increased and hence the blocking probability is decreased.

6.1 Introduction

Wireless systems brings dynamism in the propagation conditions and a variation of the network load due to fading and handoff of users from one cell/RAT to another cell/RAT. Also next generation wireless networks support various kinds of services with different QoS requirements. These services are provided through the wireless medium, which is dynamic due to fading and the mobility of the user. The main challenge for an operator is to efficiently provide a variety of services in dynamic propagation conditions.

For heterogeneous network scenario, the co-existence of multiple RATs to provide different applications/services to the users with different QoS requirements provides the need for efficient common radio resource management (CRRM) [1][2][3]. The main features of RRM or CRRM are handoff, AC etc. AC is the first thing that needs to be addressed in RRM, as even in case of handoff from one cell to another, the user has to be admitted into the cell before performing handoff. The user admitted in a cell can affect the system conditions and it also affects the QoS of the existing users. A wrong admission congests the network and also degrades the QoS of existing users.

The AC [4] is triggered when it receives an admission request from new user. The user sends a request along with different inputs, such as QoS requirements needed for its application like targetSINR, data rate, delay, PER and channel conditions like fading rate f_d and fading index m etc. Upon receiving these inputs the AC module checks whether the required resources are available in the network and hence takes the decision to admit or reject the user. The efficiency of AC can be defined in different ways by considering various parameters. Here, we consider efficient usage of system resources by the better utilization of the BW, QoS guarantee for existing and new users and fairness to all the existing users irrespective of the system conditions like load and channel conditions like fading etc.

In this Chapter various admission control methods are proposed for heterogeneous and homogeneous network scenarios, which provides better system capacity or guarantees QoS and fairness etc. In Section 6.2, the fuzzy logic based admission control is proposed for FUTON scenario, which can is applicable for any heterogeneous network. In Section 6.3 a Markov based admission control is proposed and its validated for OFDMA based system. In Section 6.4, a mean resource based AC which considers multi user diversity based on buffer conditions of the user to increase the capacity of the system is proposed and it is validated for LTE based system. In the last section the chapter is concluded.

6.2 Fuzzy logic based Admission control for Heterogeneous Networks

6.2.1 Introduction

This Section 6.2 explains a novel AC algorithm for futon architecture [4] [17] explained in Section 6.2.2, by considering various types of applications with different QoS requirements, and assigns the best cell and RAT to new users based on user conditions, user preferences, and network conditions that can provide the required QoS [16]. The proposed AC is evaluated for a heterogeneous radio access technologies (RATs) scenario. The QoS parameters vary depending on the type of application and the agreement between the service provider and the user.

The proposed AC is based on a fuzzy logic mechanism that comprises two stages; in the first stage the best cell in each RAT is selected by using a fuzzy linguistic controller, and in the second stage the best RAT based on the user preferences is selected using the fuzzy multiple attribute decision making (MADM) method. In the first stage the best cell in each RAT is selected using the cell selection algorithm and the cell admittance factor (CAF) values are obtained for each cell in a RAT. The best RAT is selected based on the weights assigned by the user to data rate, latency, BER, mobility, battery consumption and the service cost. The RAT that best fits the user needs is selected from the ranks assigned by the RAT selection algorithm to each RAT.

A number of AC mechanisms have been proposed, based on fuzzy logic, fuzzy MADM, measurement-based and parameter-based etc. The measurement-based AC takes the decision based on the current state of the network only; it has no prior knowledge about the traffic statistics. The parameter-based AC uses the parameters of the resource and service to decide whether the network can accommodate a new connection. The fuzzy logic based AC has been proposed in [5] when large imprecise data is involved. The Fuzzy MADM based AC [6] was used to combine different criteria e.g QoS, Price etc based on the weights for each criterion and to give ranking for all the alternatives. In [7], a two-tier AC based on the call level and packet level performance metrics for 4G networks, was proposed. This algorithm, however, did not consider the mobility, while allocating a RAT. In [8], a fuzzy-logic based AC with two stages was proposed, with the first stage selecting the best cell for each RAT, and the second stage using a fuzzy MADM to select the best RAT depending on the capabilities of the RAT and the user weights for each capability. However [8] did not consider the type of the call and the mobility of the user during the admission decision and RAT selection. In [5] a fuzzy-logic based AC is proposed, where for each cell the signal strength (SS), the resources available and the mobility of the user were taken into consideration for each RAT and the decision for each RAT was made based on these inputs. The main problem of this approach is the scalability. The number of combinations of input for 3 RATs in this approach was 432, and for 5 RATs this number would be 15,552 combinations.

In [6], fuzzy MADM approach is applied for network selection, based on price, BW, signal to noise ratio (SNR), sojourn time and seamlessness of the communication process. The fuzzy MADM method used in [6] operates in two steps. The first step is to convert the imprecise fuzzy variable into crisp values. The second step is to apply the classical MADM to combine the fuzzy variable and obtain ranking for each RAT. In [9] Chen used a fuzzy Q-learning based AC (FQAC) for Wideband code division multiple access (WCDMA)/WLAN heterogeneous network, where the AC was based on the number of users and the interference in WCDMA, and the network busy periods in WLAN. QoS based Q-learning provides the ability to adapt to the system dynamics and uses data rate, delay and BER.

The proposed method adopts the fuzzy logic approach and proposes a novel common AC architecture that is applicable to a heterogeneous RAT scenario. The algorithm selects the best RAT by taking into account the network status (e.g., load), the application requirements (e.g., data rate, latency, BER), the user parameters (e.g., velocity, SS).

The rest of the Section 6.2 is organized as follows. In Section 6.2.2, the system model for AC as part of FUTON architecture is described. Then the AC algorithm for the heterogeneous network scenario, which consists of cell selection and RAT selection algorithms, is proposed. In Section 6.2.3 the fuzzy-logic based Cell selection algorithm that resides in LRRM is explained. The RAT selection algorithm that resides in CRRM is explained in Section 6.2.4. Lastly the AC algorithm is evaluated in Section 6.2.5 and the proposed fuzzy based AC is concluded in Section 6.2.6.

6.2.2 System model

The scenario shown in Figure 2.1 is Futon architecture which is a hybrid fibre-radio network that interconnects multiple Remote Antenna Units (RAUs), transparently carrying radio signals to/from a CU, where joint processing takes place as part of a next generation communications scenario. The Futon architecture was proposed in [10] to serve a geographical area that is divided into several serving areas, where the multifrequency RAUs is located. These RAUs are linked to a CU, using a transparent optical fibre system, and can send/receive signals from different wireless systems.

LRRM is responsible for the local management specific to the RAT such as power control, scheduling etc. All LRRMs are connected to the CRRM. The CRRM is responsible for the global management of the resources through the LRRMs. The LRRM entity sends to the CRRM, a candidate cell list, together with measurements and the QoS requirements, and the CRRM entity selects the best target cell.

Figure 6.2 shows the stages of the proposed AC. The first stage of AC is cell selection, which is located inside the LRRM entity. LRRM can give an admission to the mobile user based on the output from the cell selection algorithm or it can forward the request with the CAF values to the CRRM, which takes a decision based on the outputs from all LRRMs and the other criteria such as the user preferences etc.

The selection of the best cell depends on the received SS reported by the mobile, and the resources available in that cell. The cell selection is done by using a fuzzy logic controller [11] with SS, resources available and type of call as inputs. The output of the fuzzy logic controller is the CAF. Two types of calls, handoff calls, either vertical handoff or a horizontal and new call, where handoff calls is given a higher priority than a new call.

The second stage of AC is RAT selection algorithm, which lies inside CRRM. Once the cell selection algorithm selects the best cell in each RAT, the most appropriate RAT, whose selected cell meets the cell selection criteria, is selected by the RAT selection algorithm. The RAT selection algorithm selects the best RAT depending on the application, QoS requirements, service cost of the RAT, battery consumption and mobility of the user. Fuzzy MADM method is used to combine the above inputs and take the decision of finding the best RAT.

6.2.3 Cell selection

The cell selection algorithm works inside the LRRM, to select the best cell in each RAT. Fuzzy logic [11][12][13] is used to select the best cell based on the SS and the resources available (RA) and the type of the call. Mamdani based fuzzy logic controller [11] consists of fuzzifier, inference engine and defuzzifier as shown in Figure 6.3.

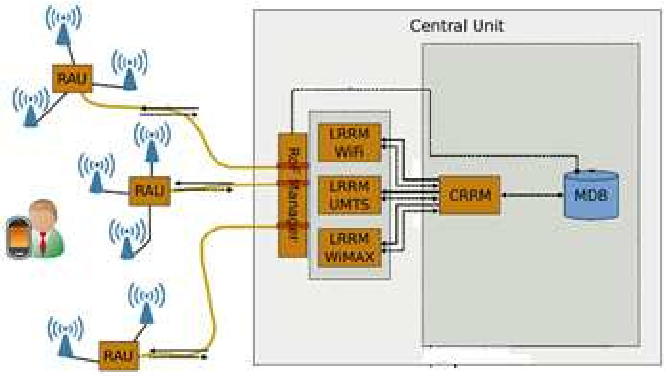


Figure 6.1: FUTON architecture with CRRM

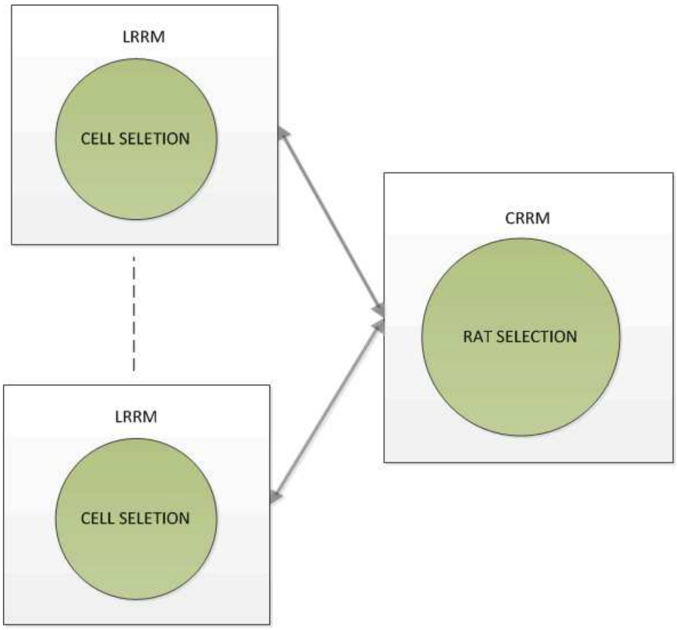


Figure 6.2: Proposed AC in FUTON

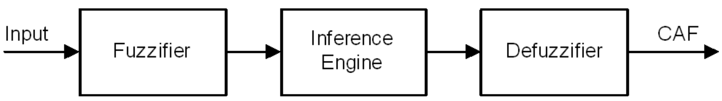


Figure 6.3: Fuzzy linguistic controller

The fuzzifier assigns a degree of membership for each input value, based on the fuzzy set it belongs to and membership function for the linguistic variable in that fuzzy set. Each of the input variables is assigned in to a fuzzy set; for example fuzzy set values of SS consist of linguistic terms: Strong, Medium, and Weak. The universe of discourse for the fuzzy variable SS is defined from -100dBm to -80dBm. The universe of discourse for resources available is from 0% to 100%, and the fuzzy set values for RA consist of the linguistic terms Very Low, Low, High, Very high. The membership functions used for the SS, Load and CAF variables are Triangular and Trapezoidal as shown in Figure 6.4- Figure 6.6. The output of the fuzzy logic controller is the CAF, which consists of the linguistic terms as No (N), Probably No (PN), Probably Yes (PY) and Yes (Y).

The inference engine executes some predefined if-then fuzzy rules, referred to as inference rules and determines the decision to admit the new or handoff call for each RAT. The predefined rules used in the cell selection algorithm are shown in Table 6.1. The decision varies depending on the type of call, where more priority is given to the vertical/horizontal handoff call than a new call.

The defuzzification involves the conversion of the fuzzy output in to the crisp output. There are many methods of defuzzifying the fuzzy output in to crisp value. Here we are using the centroid method [13]. The centroid defuzzification returns the center of the area under the curve. The output curves corresponding to each rule are added to get the final curve and then the centroid of the curve is found. Figure 6.7 shows the block diagram for cell selection algorithm.

For example, the specific fuzzy output that emerged from the composite fuzzy inference rule is N = 0, PN = 0; PY = 0.2 and Y = 0.8. Now the areas of PY and Y are combined and thus their contour becomes composite fuzzy

Table 6.1: Fuzzy inference rules for new/handoff call

Load	**Signal Strength**	**CAF for new call**	**CAF for handoff call**
VL	Weak	PN	PY
VL	Medium	Y	Y
VL	Strong	Y	Y
L	Weak	PN	PY
L	Medium	PY	Y
L	Strong	Y	Y
H	Weak	N	PN
H	Medium	PN	PY
H	Strong	PY	Y
VH	Weak	N	N
VH	Medium	N	PN
VH	Strong	N	PN

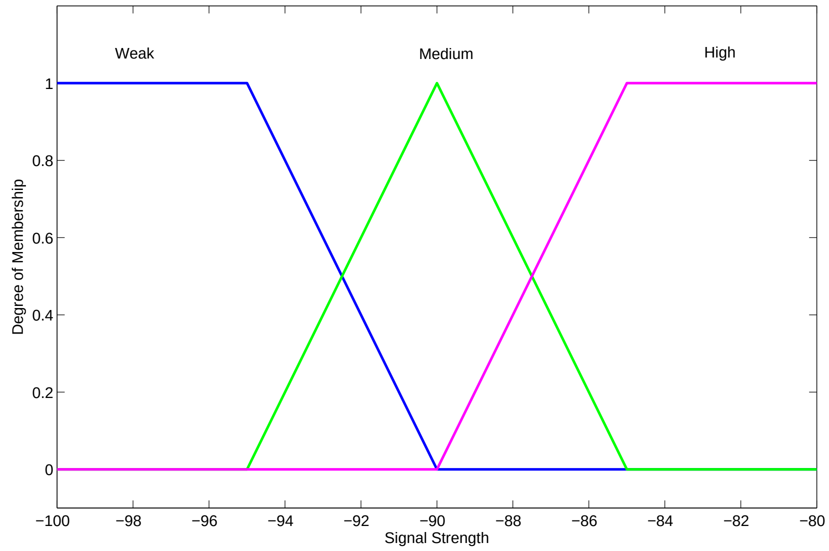


Figure 6.4: Fuzzy Membership functions

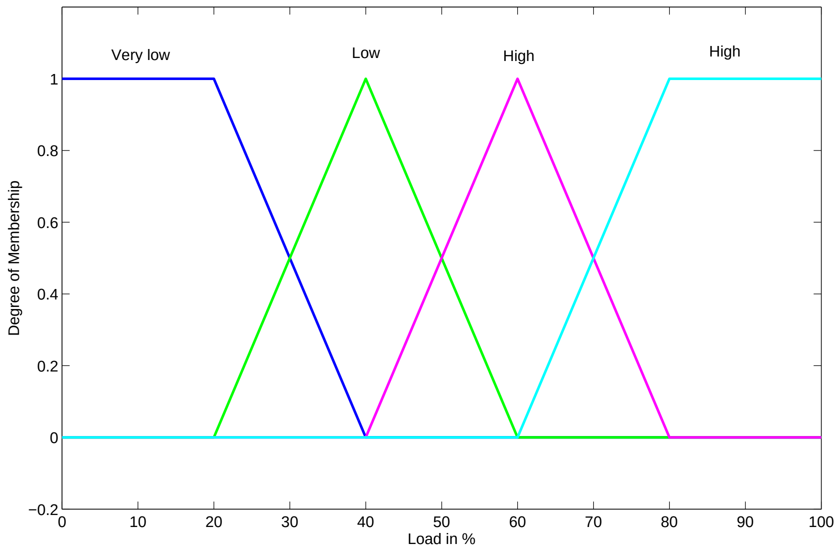


Figure 6.5: Fuzzy Membership functions

output. The method of defuzzification computes the centroid of this area using eq. (6.1).

$$\frac{\sum_{j=1}^{P} y_j \mu(y_j)}{\sum_{j=1}^{P} \mu(y_j)} \tag{6.1}$$

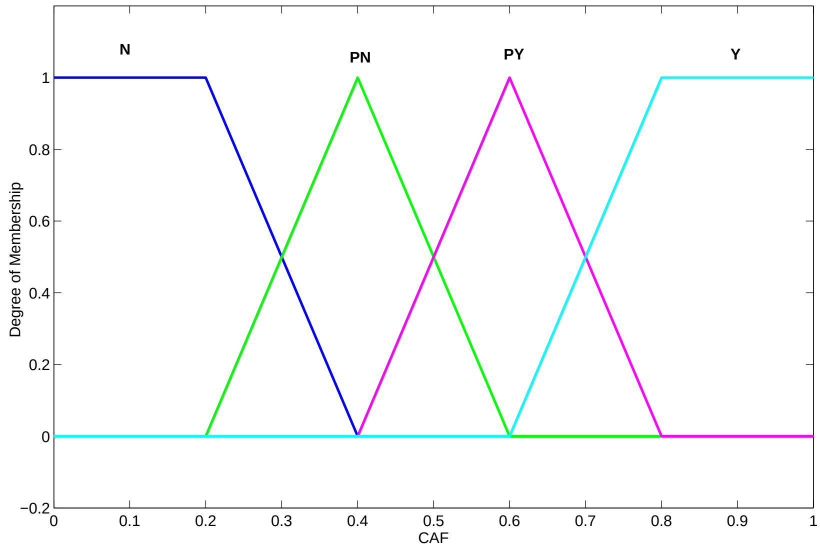


Figure 6.6: Fuzzy Membership functions

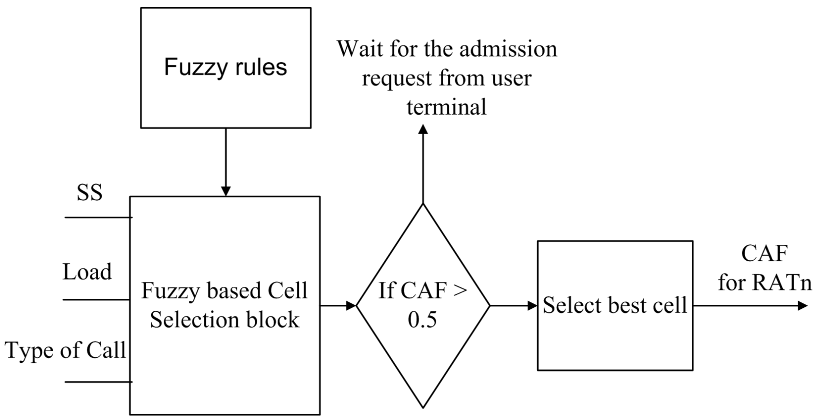


Figure 6.7: Cell Selection algorithm

where $\mu(y_j)$ is the area of membership function of PY and Y which are modified (i.e clipped) by the fuzzy inference result and y_j are the positions of the centroids of the individual membership functions PY or Y.

The cell selection algorithm inside the LRRM checks the SS received by the user from that particular cell of the RAT, and checks the available resources in the cell and type of call and calculated CAF values for each cell The cells with highest CAF values in each RAT are considered to be the best cells. The best cells with their CAF values are forwarded to RAT selection algorithm in CRRM.

6.2.4 RAT Selection

The RAT selection algorithm takes the input of best cells for each RAT from the cell selection algorithm and decides the best RAT to the user based on the inputs from the user and the suitability of each RAT to the user needs. The parameters taken into consideration for selecting the best RAT are the QoS parameters like datarate, latency and BER, mobility of the user, battery consumption and the service cost of each RAT. The reason for the selection of these parameters like QoS and mobility of the user to select RAT is because each RAT provides or supports different QoS and user mobility. For example WLAN provides higher BW compared to UMTS, where as UMTS supports higher mobility of user than WLAN. Also each RAT may have a different service cost in providing the service. Hence the goal of RAT selection is to provide the best QoS available to the user with lesser price possible. The preferences to these parameters can be selected by the user, for example some users may need best service irrespective of the service cost and some users may need best price irrespective of the QoS.

Hence RAT selection combines these parameters in a way to find the best RAT as per the user preferences. For this Fuzzy MADM (Multiple Attribute Decision Making) method is used to combine the parameters. In this approach there are two stages, the first phase is to convert the fuzzy data to crisp values and the second phase is to apply the classical MADM to find the ranking order for every alternative.

There are many classical MADM methods such as SAW (Simple Additive Weighting Method), AHP (Analytical Hierarchical Process), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), DEA (Data Envelopment Analysis) [6] etc. The most popular approach is the SAW method. In this section we illustrate RAT selection algorithm using SAW method. The SAW method consists of four parts, which are

- Alternatives
- Attributes
- Weights for each attribute
- Measures of performance of alternatives with respect to attribute

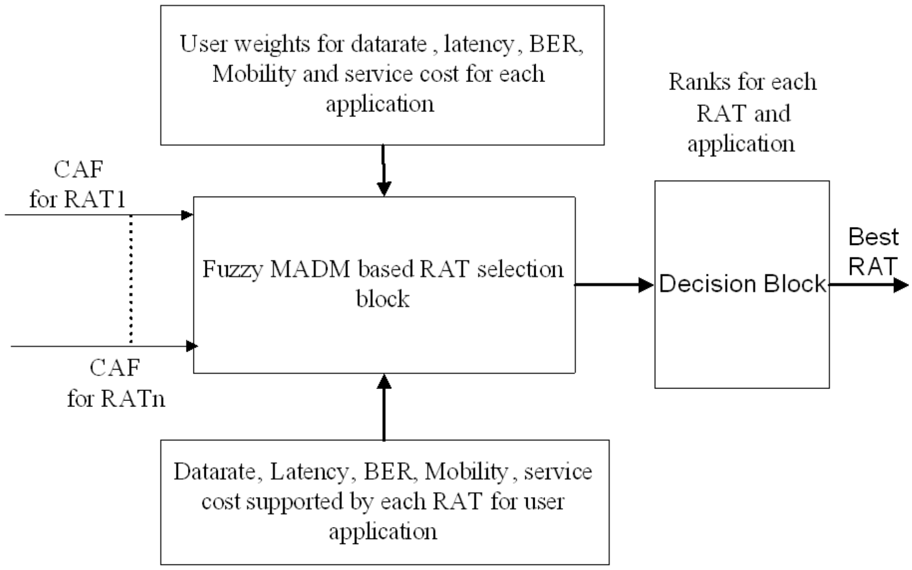


Figure 6.8: RAT Selection algorithm

If there are N alternatives each with M attributes and if w_j (for $j = 1, 2..M$) is the weight of the attribute and m_{ij} is the measure of performance ($i = 1, 2, ..., N$; $j = 1, 2..., M$) then the size of decision matrix is NXM. The SAW method finds the rank for each of the N alternatives using eq. (6.2).

$$P_i = \sum_{j=1}^{M} w_j (m_{ij})_{normal} \tag{6.2}$$

where $(m_{ij})_{normal}$ is the normalized value of m_{ij} and P_i is the composite score of alternative A_i. The alternative with the highest value of P_i is considered as the best alternative [6].

The attributes can be of two types, beneficial and non-beneficial. For a beneficial attribute (e.g. profit) its highest measures are required for given decision making problem. So the normalized value of m_{ij} is the ratio of each value with its highest attribute in N alternatives. For non beneficial attributes (e.g. cost) where lower measures are needed for decision making the normalized value is the ratio of smallest attribute with each value.

$$m_{ij} = \frac{m_{ij}}{m_{imax}} m_{ij} = \frac{m_{imin}}{m_{ij}} \tag{6.3}$$

In the RAT selection the number of attributes is five and the weights for each attribute are obtained from the users input depending on the application requirements, QoS requirements for the applications, cost requirements for the user etc. Using Fuzzy MADM the ranks for each alternative, here its RAT, is calculated based on the normalized attribute values and weights from the user preferences. The RAT with highest rank and the corresponding cell selected by the cell selection algorithm is assigned to the user. Figure 6.8 shows the block diagram, which explains the RAT selection algorithm.

6.2.5 Discussion

Consider a multimode terminal running two applications, voice and file download, needs the handoff due to mobility in moving from one cell to another. The user scans for all available cells and RATs and finds 3 UMTS cells (A_u, B_u, C_u) and 3 WLAN cells (A_w, B_w, C_w), and initiates handoff by sending a request with all the networks found and their signal strengths, and QoS requirements for each application being run on the terminal like datarate, latency, BER and the service cost it needs and also its velocity.

The QoS requirements for voice and file download are generally different. The voice application needs low data rate and low latency, while the file download needs high datarate and can tolerate more latency [14][15]. The algorithm assigns a weight for each user criterion according to the QoS needs of the application. The weights assigned by the user for voice and file download are W_v and W_d respectively. Based on the inputs from the user and the network conditions, the AC algorithm decides whether to admit or reject the user. If the user is admitted, then the cell and the RAT is allocated to the user as per the identified needs and network conditions.

In this example, the SS in dBm and load in % for UMTS in each cell A_u, B_u and C_u are $(-85, 25)$, (-80, 90), and (-95, 55). For WLAN, the SS and load in each cell A_w, B_w and C_w are (-85, 77.5), (-82, 20), (-90, 80), respectively. From cell selection algorithm, the CAF values for UMTS are estimated as 0.837, 0.4, and 0.458. For WLAN the CAF values are estimated as 0.847, 0.482 and 0.4.

Figure 6.9 is the plot showing the CAF curve based on SS and load for the new call and Figure 6.10 shows the plot showing the CAF curve for the handoff call. It can be seen that as CAF value decreases the load increases from 0 to 100%, which implies that the call admittance for a cell lower when the load is higher. Also as CAF decreases the SS reported by the user decreases, which means if a user reports a higher SS from a certain cell then the cell is given higher priority for admission. Also by comparing the CAF plots for new call and handoff call from Figure 6.9 and Figure 6.10, it can be seen that CAF values for a handoff call are higher compared to a new call, for same values of load and SS.

From CAF values for each cell, the best cells in each RAT, A_u and B_w are selected, as CAF values for A_u and B_w are higher compared to other cells. These CAF values are forwarded to RAT selection algorithm in CRRM.

The RAT selection algorithm selects the best RAT based on the datarate (m_{i1}), latency (m_{i2}), BER (m_{i3}), mobility (m_{i4}), price (m_{i5}) and battery power consumption (m_{i6}). The attributes of RAT1 and RAT2 corresponding to A_u and B_w are as given in decision matrix D.

$$D = \begin{bmatrix} m_{i1} & m_{i2} & m_{i3} & m_{i4} & m_{i5} & m_{i6} \\ low & low & medium & high & medium & 1 \\ veryhigh & low & low & low & low & 0.5 \end{bmatrix}$$

Since the user is running two applications voice and video, the preferences assigned by the user for each criteria are modeled as weights. The weight matrices for voice W_v and for file download W_d are shown below

$$W_v = \begin{bmatrix} low & low & medium & veryhigh & medium & low \end{bmatrix}$$

$$W_d = \begin{bmatrix} veryhigh & medium & low & veryhigh & low & low \end{bmatrix}$$

The linguistic terms in decision matrix D are converted to the crisp values using the conversion scale from Figure 6.11, and the decision matrix after conversion is

$$D = \begin{bmatrix} 0.283 & 0.283 & 0.5 & 0.717 & 0.5 & 1 \\ 0.909 & 0.283 & 0.283 & 0.283 & 0.283 & 0.5 \end{bmatrix}$$

The user preferences for voice and file download applications are also converted in to crisp values and normalized so that the sum is equal to 1. The normalized weights for voice W_v and file download W_d are shown in eq. (6.4) and eq. (6.5).

$$W_v = \begin{bmatrix} 0.1026 & 0.1026 & 0.1813 & 0.3296 & 0.1813 & 0.1026 \end{bmatrix} \tag{6.4}$$

$$W_d = \begin{bmatrix} 0.2870 & 0.1579 & 0.0894 & 0.287 & 0.0894 & 0.0894 \end{bmatrix} \tag{6.5}$$

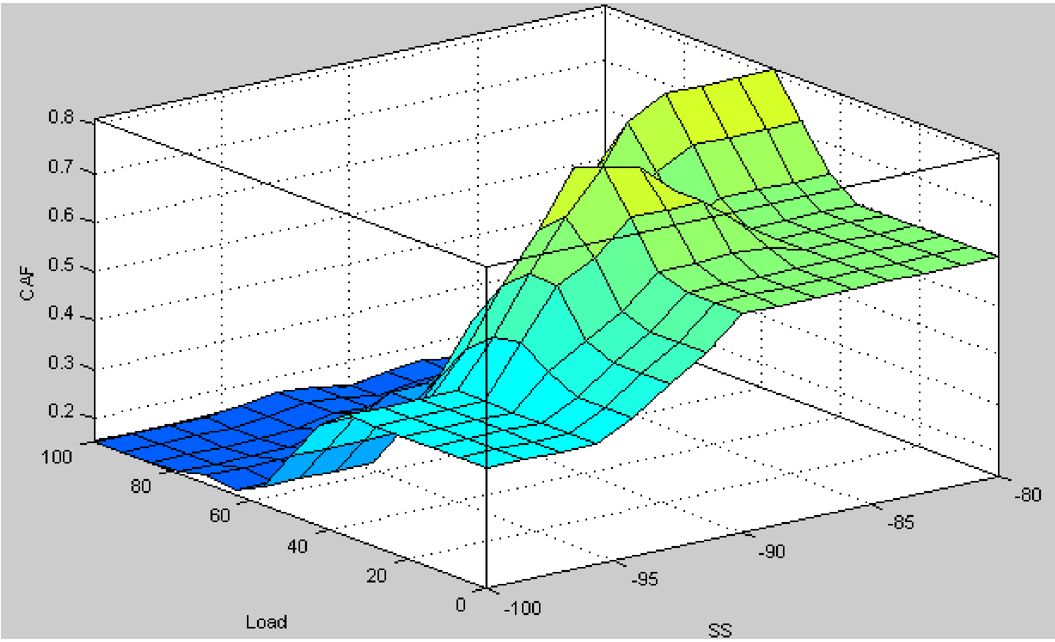


Figure 6.9: CAF plot for new call

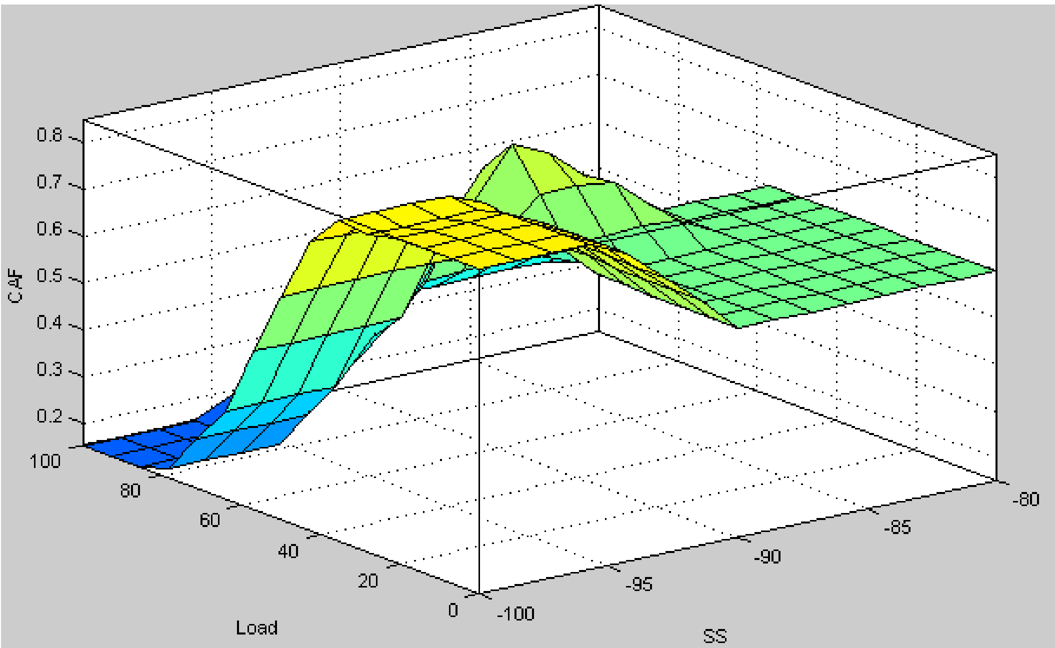


Figure 6.10: CAF plot for handoff call

The values of the decision matrix D are normalized according to eq. (6.3) based on the type of criteria, either benefit or non-benefit. In this case m_{i1},

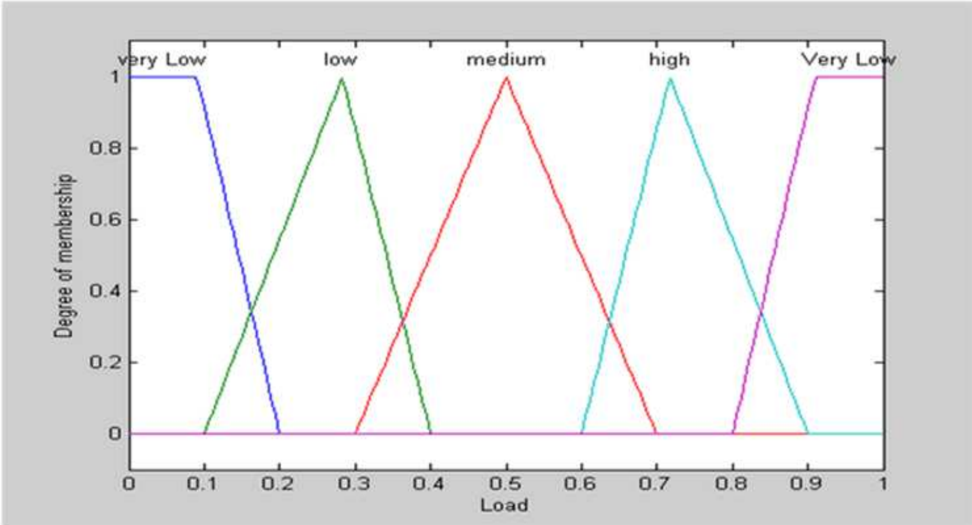


Figure 6.11: Linguistic term to fuzzy number conversion scale

m_{i4} are benefit criteria, m_{i2}, m_{i3}, m_{i5}, and m_{i6} are non-benefit criteria. The normalized decision matrix is

$$D = \begin{bmatrix} 0.3113 & 1 & 0.5666 & 1 & 0.5666 & 0.5 \\ 1 & 1 & 1 & 0.3947 & 1 & 1 \end{bmatrix}$$

Applying eq. (6.2) to decision matrix and weight matrices the final ranks of RAT1 and RAT2 for voice application are RAT1 = 0.7207 and RAT2 = 0.8005. For the file download application, the ranks for each RAT are as follows: RAT1 = 0.6801 and RAT2 = 0.8263. For both voice and file download applications, the rank obtained by RAT selection algorithm for RAT2 is greater than RAT1, therefore, RAT2 is selected. The Overall rank can be obtained by taking the average, assuming equal weights for either applications, or the user assigned priorities can be used for each application. The overall rank for RAT1 is 0.7 and for RAT2 is 0.81. Hence the user is admitted in to Cell B_w with WLAN as RAT.

6.2.6 Conclusion

In this section 6.2, AC algorithm based on fuzzy logic for FUTON architecture is proposed. This algorithm can be used for any next generation heterogeneous wireless network with heterogeneous RATs, applications, user requirements etc. The AC algorithm considers the application requirements, user conditions, network conditions user preferences and admits the user. While providing the decision to either admit or reject the user, the AC algorithm also selects the best cell and RAT. The AC algorithm is validated using numerical analysis for the case of WLAN and UMTS.

6.3 Markov based Admission control for OFDMA systems

6.3.1 Introduction

In this Section 6.3, the admission to a new user is done by considering these issues using a cross layer approach with adaptive modulation and coding (AMC) in the physical layer, scheduler in the medium access control (MAC) layer with resource allocation done in time and frequency domain. The AMC is used in the physical layer for the efficient BW utilization for a given error performance. The scheduler plays an important role in the provision of QoS. Here the design of the scheduler is based on the priority. Each user is given a priority, which is calculated based on the level of user satisfaction. In this work, the radio resource allocation is done on the downlink OFDMA, by which a better radio link capacity is achieved exploiting multiuser diversity based on channel conditions with a dynamic allocation in both time and frequency dimensions.

In [18] Jia Tang proposed a cross layer resource allocation for OFDMA systems, but did not consider scheduling, which is one of the important features for QoS. Also in [19], the author proposed utility based cross layer AC for OFDMA systems, but the resource allocation and scheduling were not considered. Also [18] and [19] are not for OFDMA based systems.

In [20], Qingwen proposed a cross layer scheduling algorithm with QoS guaranteed for the user. In this section we extend the work in [20] for an OFDMA based system with resource allocation done in time and frequency; hence the diversity is achieved in two domains. Also we use a multiple-input and multiple-output (MIMO) system, and hence the spatial diversity is also achieved. The diversities in temporal, spectral and spatial domains help to support diverse QoS. The proposed framework is not related to any specific standard and can be used in any next generation air interface technology.

This Section 6.3 is organized as follows. Initially the system model for AC is described in Section 6.3.2. Then the priority based scheduling based on user satisfaction and resource allocation for the OFDMA downlink are explained in Section 6.3.3 and Section 6.3.4 respectively. Further AC algorithm based on

Markov model is discussed in Section 6.3.5 and the results obtained based on the simulations in MATLAB are presented in Section 6.3.6. Finally the section is concluded with a follow up work proposal.

6.3.2 System Model

Figure 6.12 explains the cross layer system model used in this work. The objective of this model is to perform the AC algorithm for an OFDMA-based system, by taking into account the effects of the queue in the scheduler, AMC in choosing the right mode based on the received SNR and packet error rate (PER), resource allocation based on channel quality indicator (CQI) and priorities, and MIMO with 2x2 antennas.

The AC algorithm is triggered when a user sends a request for admission in to the cell, which happens when a new user initiates a call or when an existing user is handed off from one cell to an another cell. In both cases, the user sends a request with its service requirements and channel parameters such as doppler frequency and fading index. Based on the type of service, QoS and channel parameters the AC algorithm estimates the number of slots needed by the user to meet its requirements. If the estimated number of slots is available in the system then the user is admitted otherwise it's rejected.

The PBS receives the QoS achieved by the user, based on which it calculates the level of user satisfaction, called satisfaction index (SI). From the SI, the PBS selects the users to be scheduled in every frame, such that the user with the least SI gets the highest priority in the resource allocation. Hence the multiuser diversity based on the user satisfaction is achieved with the PBS. The user priorities are sent to the resource allocation module by the scheduler.

The resource allocation module finds the best slots for every user based on the priorities from the scheduler, the preferred slot input from the user, and the number of slots be to scheduled from AC module. The preferred slot is calculated by the user, from the CQI in each subcarrier, which is sent to the base station.

6.3.3 Priority based scheduler

A user may not achieve the requested QoS due to channel conditions like fading, traffic conditions or network conditions like load etc. So the users who are not receiving the requested QoS should be given higher priority for transmission

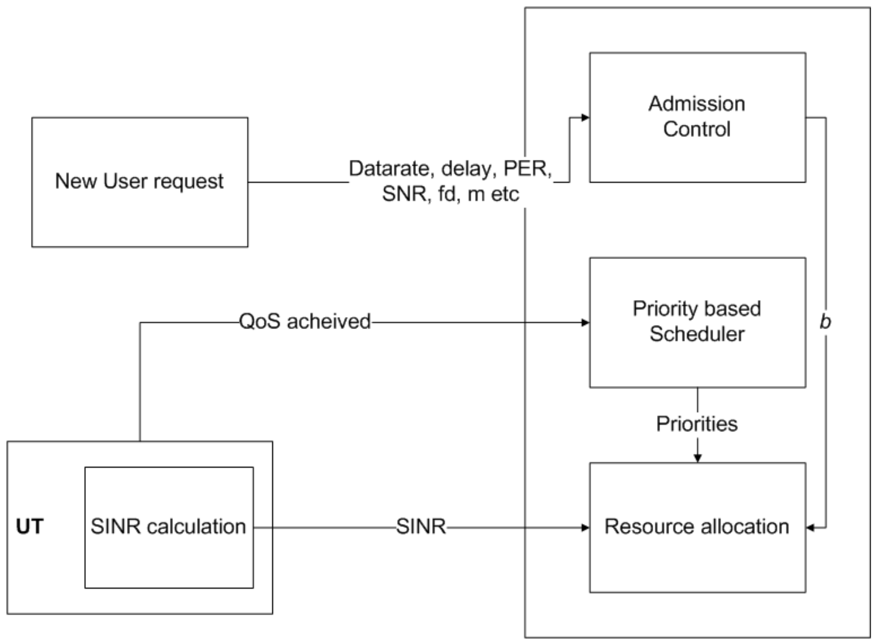


Figure 6.12: System model for AC

in the next frame. The main function of the PBS is to select the users to be scheduled in every frame based on the estimated priority, so that the user with highest priority is scheduled first, by which fairness will be achieved for each user. The amount of resources to be allocated to each user is estimated by the AC algorithm based on the QoS. For each user the SI is calculated, which gives the level of user satisfaction based on the throughput and delay achieved with respect to the desired values. The desired values of QoS are assumed to be dependent on the type of service. Based on the SI, the scheduler calculates the priorities for each user and send these to resource allocation module.

Here we consider four types of service classes, which are shown in Table 6.2. The class 1 users have equal weights for rate and delay like video applications, class 2 is for real time applications with low datarate like voice, class 3 users have more weight for the rate with more delay tolerance, like file download or browsing etc and class 4 users are the best effort users. The weights for each class are used in calculating the priority function.

The method for calculating the priorities based on the satisfaction index and weights is defined in Section 3.4. The priorities for each user are sent to the RA module.

Table 6.2: Service classes and SI Coefficients

Class	Rate	Delay	w_u^{rt}	w_u^{nrt}	notes
1	200 kbps	50ms	1	1	High rate and low delay
2	-	50ms	1	0	Low delay
3	200 kbps	-	0	1	High rate
4	-	-	0	0	Best effort

6.3.4 Resource allocation

The resource allocation module allocates the best slots in every frame to all the users in the system, such that the sum of the priorities of all the allocated slots is maximized. The resource allocation is done on the OFDMA grid of subcarriers spanning the time-frequency domain. The time-axis represents the OFDM symbols, where as the frequency axis represents the subcarriers [21], with the number of subcarriers depending on the length of fast fourier transform (FFT). For the two axes two partition factors - Λ_t and Λ_f are defined. In OFDMA, the minimum allocation unit is represented by a slot, where one OFDMA frame is composed of $\Lambda = \Lambda_t.\Lambda_f$ slots.

The allocation of slots to the users is based on the CQI calculated by the user. Each user calculates the CQI for each subcarrier and averages this information to find the CQI in each slot and sends this information to the base station. By receiving the CQI, the base station assigns the priority $\Phi_u(t, \Lambda_i, \Lambda_j)$ obtained from PBS to the preferred slot. This information is spread to the neighbor slots to calculate the priorities of the user in other slots, which is useful in case two or more users denote the same preferred slot. The spreading is based on the Euclidean distance (eq. (6.6)) between the preferred slot and all other slots.

$$l_{\hat{i}\hat{j},ij} = \sqrt{(\lambda_i - \hat{\lambda}_i)^2 + (\lambda_j - \hat{\lambda}_j)^2} \tag{6.6}$$

where $i \in [1, \Lambda_t]$ and $j \in [1, \Lambda_f]$

Taking in to consideration the spreading factor, the priorities for each slot are given by (eq. (6.7))

$$\Phi(t, \Lambda_i, \Lambda_j) = (1 - \alpha_\lambda . l_{\hat{i}\hat{j},ij}) \tag{6.7}$$

where α_λ is the coefficient denoting the loss by assigning the user to a slot other than the one considered the preferred one. If a user doesn't get the requested slot, then the user is given the best slot with priority value close to the priority calculated in the scheduler, with the help of spreading function. Hence the goal of the resource allocation is to maximize the sum of the priorities of all slots.

6.3.5 Admission Control

For each new request the AC calculates the number of slots required to meet the QoS requested by the user. For this we use Markov modeling of the queue which is explained in Section 5.2 to predict the number of slots b. Using the procedure explained in Section 5.2, η_{prior} and τ_{prior} are calculated for different values of b. The minimum value of b is selected, for which the required throughput and delay of new user is greater than the estimated throughput and delay.

Once suitable value of b is selected, if the number of slots needed by the user is available in the system, then the user is admitted otherwise the user is rejected, which is governed by eq. (6.8).

$$b + \sum_{j \in J} b_j \leq B \tag{6.8}$$

where B is the total number of slots in the system, J is the total number of users available in the system and b_j is the number of slots allocated to user j.

6.3.6 Results and Discussion

The simulation setup used to measure the performance of the AC is as follows. We consider four types of users with different QoS requirements is given in Table 6.2. For each service the target PER is taken as 0.05 and the target SNR is 27dB. Here three modes of transmission, which are 4-QAM, 16-QAM and 64-QAM are used. If the received SNR is below a certain threshold then mode 0 is used, where there is no transmission. The base station transmits to each user in one of the transmission modes, based on the received SNR and the target PER requirement of the service. The transmission mode can be different in different frames. A 2x2 MIMO system is used and between each antenna a Nakagami fading channel is simulated with $f_d = 10Hz$ and fading index $m = 1$.

The OFDMA PHY layer parameters are presented in Table 6.3. The PHY layer parameters are chosen according to the IEEE 802.16 standard and payload mapping is done according to adjacent subcarrier permutation, both defined in [26].

Table 6.3: PHY layer PARAMETERS.

Parameter	**Value**
FFT points	128
Pilot sub	12
Left guard sub	10
Right guard sub	19
Data sub	108
OFDM symbols	48
CP length	11.43 μs
OFDM symbol time	102.86 μs
Λ_t	12
Λ_f	16
Slot dimension	4*6

The BS transmits to each user, in their preferred slots based on CQI. The OFDMA modulation is used for the transmission, with a frame duration T of 5ms. Each downlink OFDMA frame has 48 symbols with each symbol having 128 subcarriers. Each slot is defined as 4 symbols of 8 subcarriers, which constitutes 32 subcarriers in each slot. The total number of slots that can be allocated to the users is 192.

The packet arrival rate and buffer length are the key parameters of the markov analysis to predict the number of slots needed by the new user. The values used are shown in the Table 6.4.

Figure 6.13 shows the average throughput of the users in class 1, 2 3, and 4 over time, without AC, where each user gets a fixed allocation of 5 slots in every frame. It can be seen clearly that for class 1 and 3 users the throughput is not satisfied, and it is around 120 kbps not fulfilling the required throughput of 200kbps. For class 2 and 4 users the average throughput is 80 kbps, which is above the required QoS. Hence the QoS is not guaranteed for class 1 and 3 users and for class 2 and 4 users the BW is not utilized efficiently.

Figure 6.14 shows the average throughput, with class 1 users alone in the system. The average throughput of class 1 users was plotted with an increase

Table 6.4: Packet arrival parameters

	Class 1 and 3	**Class 2 and 4**
Buffer Length (K)	40	10
λ	6 packets/frame	1 packets/frame

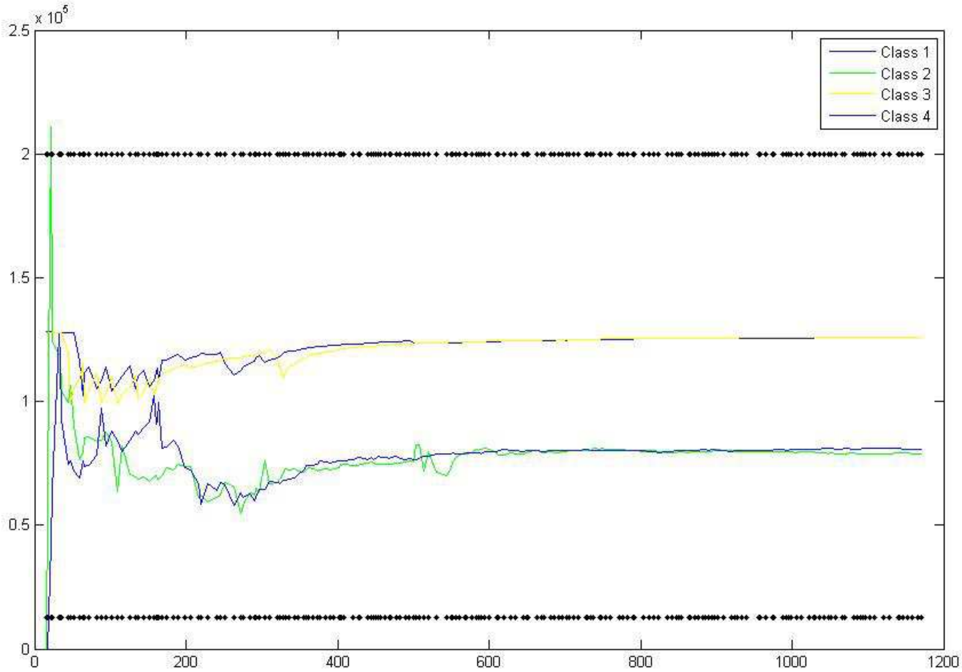


Figure 6.13: Throughput with fixed allocation

in the number of users in the system, as they are admitted by AC. It can be seen that as the load or number of users increase, the average throughput of the existing users is maintained consistently.

Figure 6.15 and Figure 6.16 shows the average throughput obtained by class 1,3 and class 2,4 users respectively. The system reaches its maximum load at the end of the simulation, where all slots are allocated to all users. When the system reaches maximum load, there are 12 users of class 1, 7 users of class 2, 9 users of class 3 and 7 users of class 4, in total 35 users in the system. Hence it can be observed that as the system reaches its maximum usage of slots the QoS for class 1 and 3 is maintained. For class 1 and 3 users the throughput is above the 200kbps consistently. The throughput achieved by class 2 and 4 users is shown in Figure 6.16 where it can be seen that the throughput achieved is consistently greater than 12.8 kbps. Hence, with the proposed AC, the throughput is guaranteed for each class of user.

In Figure 6.17 the delays achieved by class 1 and 2 users are plotted, as both are delay sensitive users. It can be observed that the delay starts with zero, in the start of simulation when there are less number of users and as the

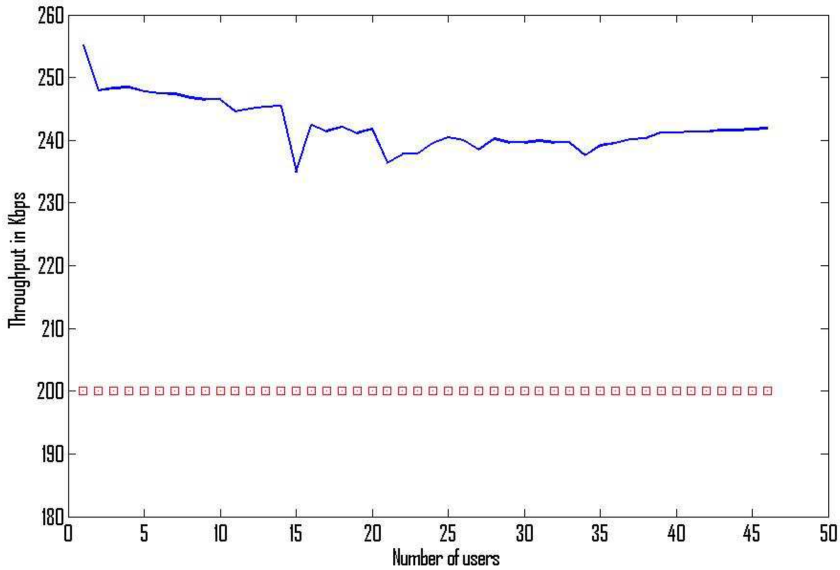


Figure 6.14: Throughput with AC for Class 1 users

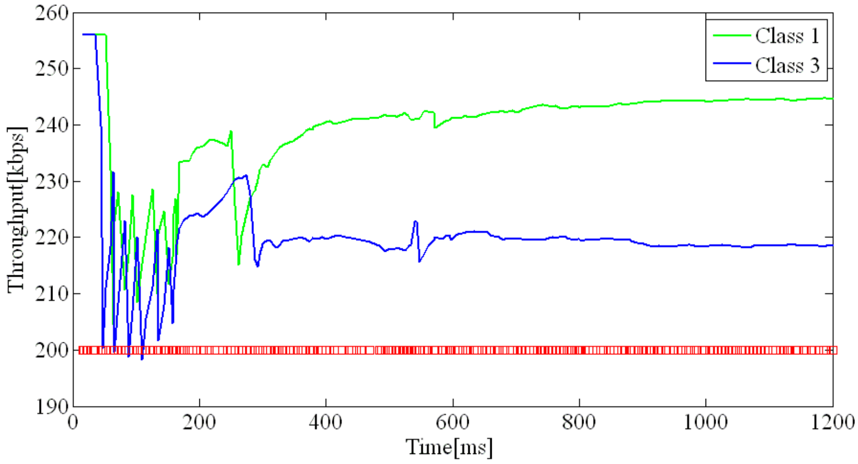


Figure 6.15: Throughput with AC for class 1 and 3 users

simulation progresses the load increases to maximum. The average delay of each class is within the required limit of 50 ms.

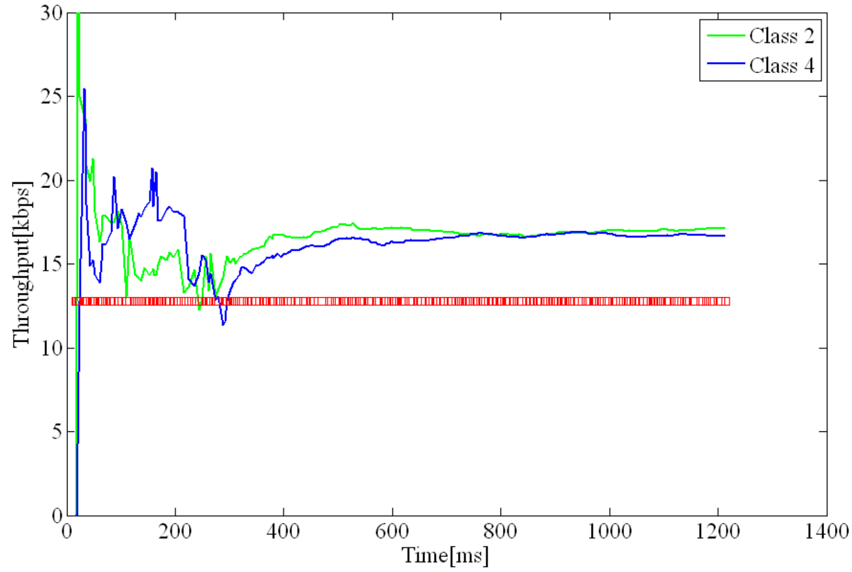


Figure 6.16: Throughput with AC for class 2 and 4 users

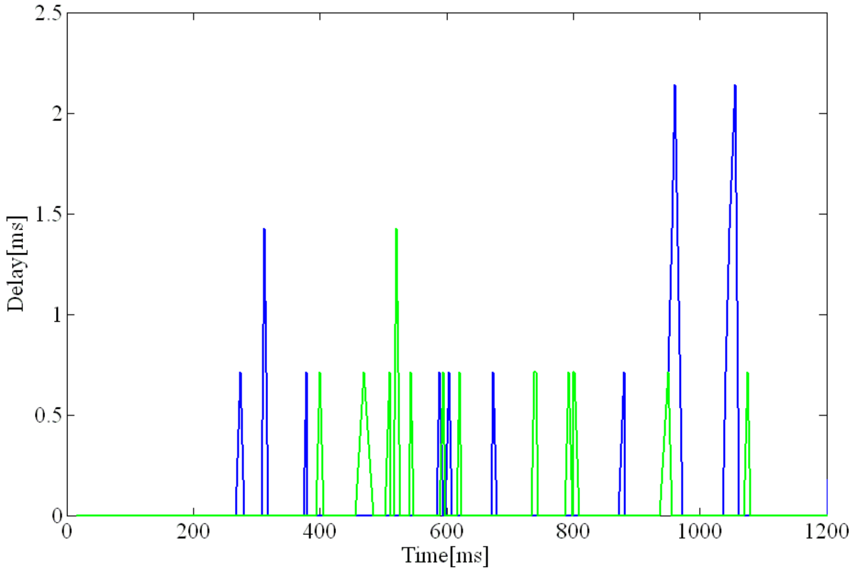


Figure 6.17: Delay profile for classes 1 and 2 with AC

6.3.7 Conclusion

In this Section 6.3, we proposed a cross layer AC algorithm by considering the queuing effects of scheduler in MAC layer, AMC and MIMO in physical layer and resource allocation based on priority and CQI. With AC algorithm, the number of slots required to meet the QoS requirement is estimated, hence the QoS is guaranteed to each user irrespective of the system and channel conditions. By using the AMC the BW is utilized efficiently in all the channel conditions. Diversity in time and frequency domain is achieved with resource allocation and spatial diversity is achieved with the MIMO. Hence the system can support users with varied QoS. The future work is to extend the current framework for heterogeneous wireless networks, where the AC should select the most suitable cell and RAT for the user that can cater the best QoS to the user while maximizing the overall capacity of the heterogeneous network. Also this work can be used in self optimization of the network, by taking into account the cross layer dependencies, like dynamic buffer allocation etc.

6.4 Mean resource based Admission Control

The Mean resource based AC algorithm increase the capacity of system and the number of connections that can be served by taking buffer conditions of each user in to consideration [29][20]. A user may not utilize all the resources allocated due to lack of packets in the buffer. Thus, in order to obtain an efficient utilization of the BW, a mean resource calculation is performed, which finds the average number of resources used by all the users in the system. The number of resources actually scheduled [28] can be expressed in the following way, depending on the current channel conditions and the previous buffer status:

$$k(U_{t-1}, C_t) = \begin{cases} 0; & \text{if } C_t = 0, \\ k_i; & \text{if } U_{t-1} \geq C_t, \\ floor(\frac{k_i * U_{t-1}}{C_t}); & \text{if } U_{t-1} < C_t. \end{cases}$$

where U_{t-1} is the number of packets that are in the queue for user i at the time moment $t-1$, k_i is the number of resources allocated for user i and C_t is the number of packets that can be accommodated in the next frame with the selected AMC mode. If k_i is the maximum number of resources that can be allocated to user i, then maximum number of resources that can be allocated to all the users in a system is $k_I = \sum_{i \in I} k_i$. The users are admitted till k_I reaches maximum number of resources in the system. The goal of mean resource allocation is to find the average value of k_I, such that more users can be accommodated in the system.

The mean number of resources used by all users, or average value of k_I, is estimated from the steady state distribution of k_I, which can be determined from the Z-transform $D_i(z)$ of k_i. The Z-transform of k_i can be expressed as:

$$D_i(z) = \sum_j P(k_i = j) z^{-j} \tag{6.9}$$

where $P(k_i = j)$ is the probability that the user i is allocated with j resources. The Z transform of k_I is expressed as $D_I(z) = \prod_{i \in I} D_i(z)$. By calculating the inverse Z-transform, the steady state distribution of k_I is obtained as:

$$P(k_I = j) = Z^{-1}\big\{D_I(z)\big\} \tag{6.10}$$

From the steady state distribution of k_I the mean number of resources k_I^{mean} used by all the users in the system is obtained from the kth moment of k_I. Based on the estimated value of k_I^{mean}, a new user i which requires k_i resources is admitted according to:

$$k_i + k_I^{mean} \leq N \tag{6.11}$$

where N is the total number of resources available in the system.

6.4.1 Mean Resource Evaluation

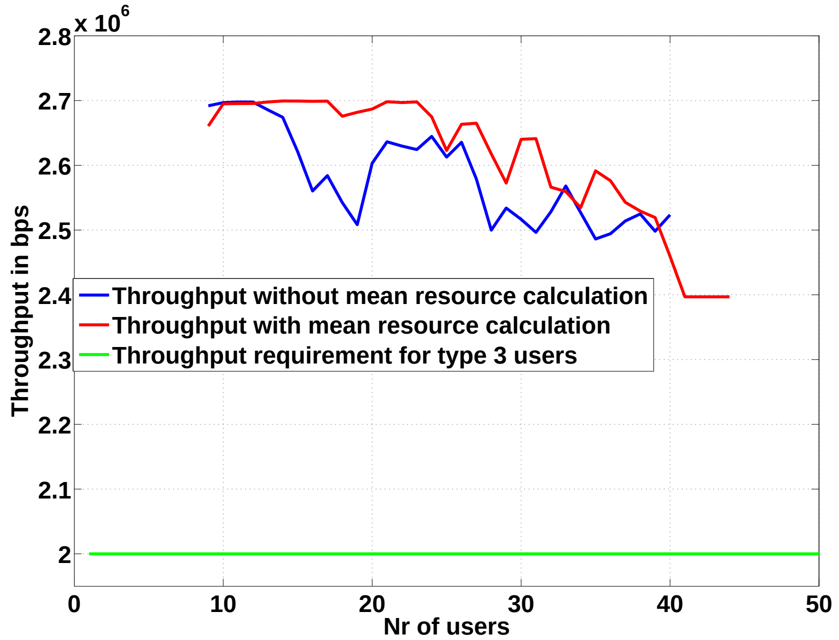
The proposed mean resource based AC is validated using the system model for AAA framework explained in Section 3.3. For allocation and assignment, the methods proposed in Chapter 4 and Chapter 5 are used. The simulation setup used in the simulation is explained in Section 4.4.1 and the parameters of the simulation are shown in Table 4.1.

In the simulation three types of users with 400kbps, 800 kbps and 2Mbps are assumed. For each type of user, the number of PRBs required to obtain the required throughput is estimated using the markov based approach explained in Section 6.3. In this simulation, the performance of mean resource calculation, is compared to a scenario that does not perform the mean resource calculation. In the scenario without mean resource calculation, each user is allocated with the resources estimated by using the markov based method throughout its call. The goal of mean resource calculation is not to improve the throughput of the user, but to improve the number of users admitted in to the system while maintaining the QoS of existing users in the system.

Figure 6.18 and Figure 6.19 shows the performance of admission control with mean resource algorithm. Figure 6.18 shows the average user throughput for type-2 and type-3 users with increasing number of users, with and without mean resource calculation. Without mean resource calculations, the maximum number of type-3 users that can be admitted in the network is 40 (10 users for each cell). Red curve illustrates the average user throughput with mean resource calculations and it can be seen that 44 users are admitted in the system, with 10% increase in number of users compared to without mean resource calculation. It can be seen that with mean resource calculation the number of users admitted in the system increase while maintaining the average user throughput. Hence, the AC algorithm guarantees QoS for the new user while maintaining QoS for existing users, which can be seen as the number of users increases till the load reaches 100%.

In case of type-2 users, it can be seen that with mean resource calculation the number of users admitted in the system is increased by 10% while the average user throughput is maintained and the minimum QoS requirement is met. In both cases, the average user throughput decreases marginally with an increase in the number of users. This is due to the fact that more users have to share the same resources. Nevertheless, the minimum required throughput is still achieved even in this case where more users are admitted to the system.

Figure 6.19 illustrates the dropping probability with the variation of the Erlang Load for type-3 and type-2 users. The Erlang load is defined as the ratio of the call arrival rates to the mean duration of call. It can be seen that by applying the mean resource calculation, the dropping probability is reduced by around 5.5% for type-3 user and 2.2% for type-2 user when compared to without mean resource calculation at 110% Erlang load. The increase in performance of mean resource calculation is highlighted better for type-3 users compared to type-2. This is due to the fact that type-3 users require more number of PRBs than type-2 users, which increase the probability of unused PRBs for type-3 users than type-2 users. Hence mean resource algorithm increase number of admitted users, which reduce the dropping probability.



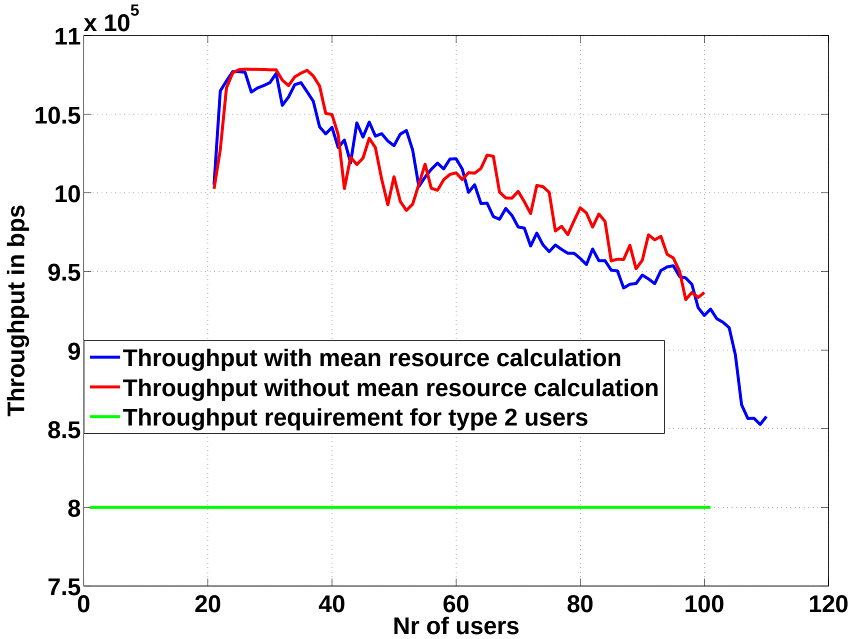


Figure 6.18: Mean resource performance for Type-2 and Type-3 users

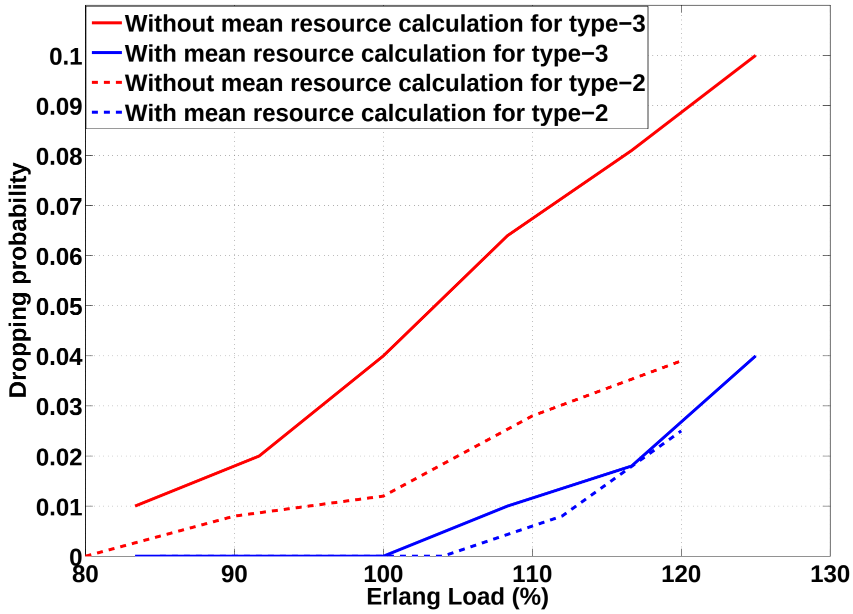


Figure 6.19: Dropping probability vs Erlang load for Type-3 users

6.4.2 Conclusion

In this section, we propose a mean resource based AC algorithm which guarantees the QoS for a new user and maintains QoS for existing users, while achieving an increase in the number of users admitted by 10%. The proposed algorithm was verified on an LTE platform, but this model can be extended to any next generation cellular network. Furthermore, the proposed algorithm increases the reliability of the system by reducing the dropping probability.

6.5 Conclusions

In this chapter, we propose three methods for admitting a new user based on fuzzy logic and markov process. THe fuzzy logic based AC is applicable for a heterogeneous network scenario and also performs cell and RAT selection. The markov based AC guarantees the QoS for a user irrespective of the load in the network. The last method increase the number of users in the network without violating the QoS of existing users in the network. The last method of AC is validated as a part of AAA framework.

References

[1] 3GPP. Improvement of Radio Resource Management (RRM) across RNS and RNS/BSS (Post Rel-5). TR 25.891, 3rd Generation Partnership Project, February 2003. Available from World Wide Web: http://www.3gpp.org/ftp/Specs/html-info/25891.htm [cited June 2008]. Release 6.

[2] Albena Mihovska, Tragos Elias, Mino Emilio, Luo Jijun, Mensing Christian, Fracchia Roberta, Horrich Sana, Hui Lin, Klockar Annika, Kyriazakos Sofoklis, " Requirements and Algorithms for Cooperation of Heterogeneous Radio Access Networks," Journal on Wireless Personal Communications, Springer, DOI: 10.1007/s11277-008-9586-y, Oct. 2008.

[3] FUTON Project Deliverable D4.1 "CRRM Functional Architecture and Interference Scenarios on FUTON Project" Dec. 2008, http: www.ict-futon.eu

[4] FP7 ICT Project FUTON, at http: www.ict-futon.eu.

[5] L. Giupponi, R. Agusti, J. Perez-Romero, O. Sallent , " Joint radio resource management algorithm for multi-RAT networks," IEEE Globecom 2005, pp. 3851-3855, Dec. 2005.

[6] Wenhui Zhang, " Handover decision using fuzzy MADM in heterogeneous networks," IEEE Wireless Communications and Networking Conference, WCNC, vol. 2, pp. 653-658, Mar. 2004.

[7] D. Niyato, E. Hossain, " Call Admission Control for QoS Provisioning in 4G Wireless Networks: Issues and Approaches," IEEE Network, vol. 19, no. 5, pp. 5-11, Sep-Oct. 2005.

[8] O. E. Falowo, H. A. Chan, " Fuzzy Logic Based Call Admission Control for Next Generation Wireless Networks," International Symposium on Wireless Communication Systems, ISWCS, pp. 574-578, Sep. 2006.

[9] Yung-Han Chen, Chung-Ju Chang, Ching Yao Huang, " Fuzzy Q-Learning Admission Control for WCDMA/WLAN Heterogeneous Networks with Multimedia Traffic," IEEE Transactions on Mobile Computing, vol. 8, no. 11, pp. 1469-1479, Nov. 2009.

[10] FUTON project Deliverable D1.1 "Project Presentation," Dec. 2007, http: www.ict-futon.eu.

[11] E. H. Mamdani, " Application of fuzzy logic to Approximate Reasoning Using Linguistic Synthesis," IEEE transactions on Computers, vol. C-26, no. 12, pp. 1182-1191, Dec. 1997.

[12] L. A. Zadeh, " Fuzzy sets," Information and Control, vol. 8, pp. 338-353, 1965.

[13] Amit Konar, Computational Intelligence, 1st ed. , Springer, May. 2005.

[14] T. Bejaoui, V. Veque, S. Tabbane, " Dynamic interference and timeout-based CAC scheme for multimedia cellular networks," IEEE 61st Vehicular Technology Conference, VTC , vol. 3, pp. 2086-2091, Jun. 2005.

[15] Universal Mobile Telecommunications (UMTS); "QoS concept and architecture," 3Gpp TS 23.107 version 5.8.0 Release 5, 2002.

[16] M.V. Ramkumar, Albena Mihovska, Neeli R. Prasad, Ramjee Prasad, "Fuzzy-logic based Call Admisison Control for a Heterogeneous Radio Environment," WPMC 2009, Japan, Sep. 7 -9

[17] M. V. Ramkumar, Albena Mihovska, Neeli R. Prasad, "Management of Distributed and Extendible Heterogeneous Radio Architectures," NAEC2009, Lake Garda, Italy, October 8-11, 2009

[18] Jia Tang, Xi Zhang, " Cross-layer design of dynamic resource allocation with diverse QoS guarantees for MIMO-OFDM wireless networks," Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks, WoWMoM, pp. 205-212, Jun. 2005.

[19] Hu Nan, Wand Xiaoxiang. Wu Weiling, " Integrated Cross-Layer Design of Utility-Based Connection Admission Control in Packet-switched OFDM Wireless Networks," International Workshop on Cross Layer Design, IW-CLD, pp. 152-156 , Sep. 2007.

[20] Qingwen Liu, Shengli Zhou, G. B. Giannakis, " Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks," IEEE Journal on Selected Areas in Communications, vol. 23, no. 5, pp. 1056-1066, May. 2005.

[21] Filippo Meucci, Albena D. Mihovska, Bayu Anggorojati, Neeli R. Prasad, " Joint Resource Allocation and Admission Control Mechanism for an OFDMA-Based System," 11th International Symposium on Wireless Personal Multimedia Communications, WPMC, Sep 2008.

[22] Qingwen Liu, Shengli Zhou, G. B. Giannakis, "Cross-layer modelling of adaptive wireless links for QoS support in heterogeneous wired-wireless networks," Wireless. Networking. vol. 12, no. 4, pp. 427-437, 2006.

[23] Q. Zhang, S. Kassam, "Finite-state Markov model for Rayleigh fading channels," IEEE Transactions on Communications, vol. 47, no. 11, pp. 1688-1692, Nov. 1999

[24] M. K. Simon, M.S. Alouini, Digital Communication over Fading Channels, 2nd ed. , Wiley-IEEE Press, Dec. 2004.

[25] R. B. Cooper, Introduction to Queuing Theory, 2nd ed. Elsevier North-Holland, 1981

[26] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001), pp. 1-857, 2004.

[27] Qingwen Liu, Xin Wang, Giannakis. G. B, "A Cross-Layer Scheduling algorithm with QoS Support in Wireless Networks," Transactions on Vehicular technology, vol. 55, no. 3, pp. 839-847, May. 2006.

[28] B. Krasniqi, M. Wrulich, C.F. Mecklenbrauker, " Network-load dependent Partial frequency reuse for LTE," International Symposium on Commnucations and Information Technology. ISCIT, pp. 672-676, Sep. 2009.

[29] M. V. Ramkumar, Andrei Lucian Stefan, Rasmus Hjorth Nielsen, Neeli R. Prasad, Ramjee Prasad, "Efficient Admission Control for Next Generation Cellular Networks," submitted in ICC 2012

7
CRRM Implementation

This chapter discusses the functionalities of FUTON Middleware and its internal modules. The implementation details of CRRM and the interfaces of CRRM with other modules of the FUTON middleware are explained. The CRRM framework is validated for the case of vertical handoff between UMTS and WLAN and the results are presented. Secondly the RoF manager, applicable for RoF network architectures is explained and the implementation of authentication, as a part of security management module is presented and the results are shown at the end.

7.1 Introduction

Futon framework provides a platform to integrate heterogeneous radio access networks. It is a hybrid fibre-radio network which replaces legacy BSs with simple remote antenna units (RAUs), which are connected to a central Unit (CU). As a part of the Futon framework, a generic middleware architecture is discussed in detail, which provides interoperability, cooperative management and service provisioning to both underlying radio access networks (RANs) and IP layer for a heterogeneous network scenario. One of the middleware's functionality is to provide vertical handover between heterogeneous IP-based radio access technologies (RATs) and to ensure seamless mobility and service continuity. The middleware is implemented on top of an IPv4/IPv6 Mobile IP (MIP) core to provide the interconnection of Futon components with the IP world and vertical handover between heterogeneous IP-based RATs and ensures seamless mobility and service continuity.

This Chapter 7 is organized as follows. A generic middleware architecture as a part of Futon framework for heterogeneous RANs is discussed in Section 7.1.1. The main modules of the middleware, namely CRRM, media independent handover (MIH), service/connection manager (SCM) and link selection (LS) are explained in Sections 7.1.2, 7.1.3, 7.1.5 and 7.1.4 respectively. In Section 7.2, the implementation design and details of CRRM and LRRM are explained in detail and results of vertical handoff from WLAN to UMTS and vice versa are presented. In Section 7.3 the RoF manager is explained and the implementation of authentication procedures inside the security management module of RoF manager is presented.

7.1.1 Middleware Architecture

As part of Futon architecture [1], an IP based middleware is developed for managing heterogeneous networks [2]. The basic operation of the middleware is to provide interconnection between lower layers of the OSI model with upper layers such as networking and transport layer. As NGNs are moving towards IP-based architectures, middleware functionalities are particularly dedicated towards interconnection of Futon components with the IP world. Apart from that, the middleware is intended to perform CRRM functionalities such as access control, network discovery, vertical/horizontal handover etc. Due to single point of operation and the advantage of having the data from different networks, the middleware will be in charge of vertical and media independent handover, cross-layer and cross-system optimization, CRRM, network discovery, authentication control, and seamless heterogeneous network provisioning.

The Futon middleware is envisaged to be responsible for interoperability, cooperative control and provisioning of services to both the underlying RANs and the IP layer. It includes the routing functionality that will implement a vertical handover between heterogeneous IP-based RATs (e.g. WLAN, UMTS, and WiMAX) and ensure seamless mobility and uninterruptable service provision (data, voice or media). This functionality is implemented on top of an IPv4/IPv6 MIP core.

As shown in Figure 7.1, the middleware entity consists of SCM, CRRM, MIH, LS and a middleware database (MDB) modules and it interfaces with MIP and Radio over fibre (RoF) manager [3]. The SCM controls all the services/functionalities provided by the middleware architecture and sends handover request to MIP core. The MIH module handles seamless mobility and session continuity at the middleware service platform. Another important module of middleware is the CRRM, which manages the network, with information available from LRRMs and takes global decisions to improve the overall performance of the network. The MIH module raises the trigger for handoff from user point of view if QoS of the user is not met, whereas the CRRM raises the trigger due to network reasons, and send to LS. The LS entity selects the best RAT, cell site, and link within RAT etc, for handoff by taking inputs from MDB, updated by both CRRM and MIH. All the four components of the middleware consistently update the MDB with important information, which can be used by other modules of the middleware.

MIP provides transparent internet mobility management services end ensures transfer of mobile terminals in a continuous and seamless way. The MIP core has the functionality of home agent and foreign agent and it resides in CU. The MIP client sits on the mobile terminal and implements mobile node (MN) functionality. There is also a direct interaction between the middleware entity and the RoF manager which monitors the fibre optical network by receiving triggers (or events) from software agents that are sitting on the network elements. The modules inside the middleware are explained in detail in this section.

7.1.2 CRRM

The CRRM platform is used in heterogeneous radio environments, which involves multiple RATs, for the management of radio resources with a common framework between the cells/RATs. The LRRM is used for the management

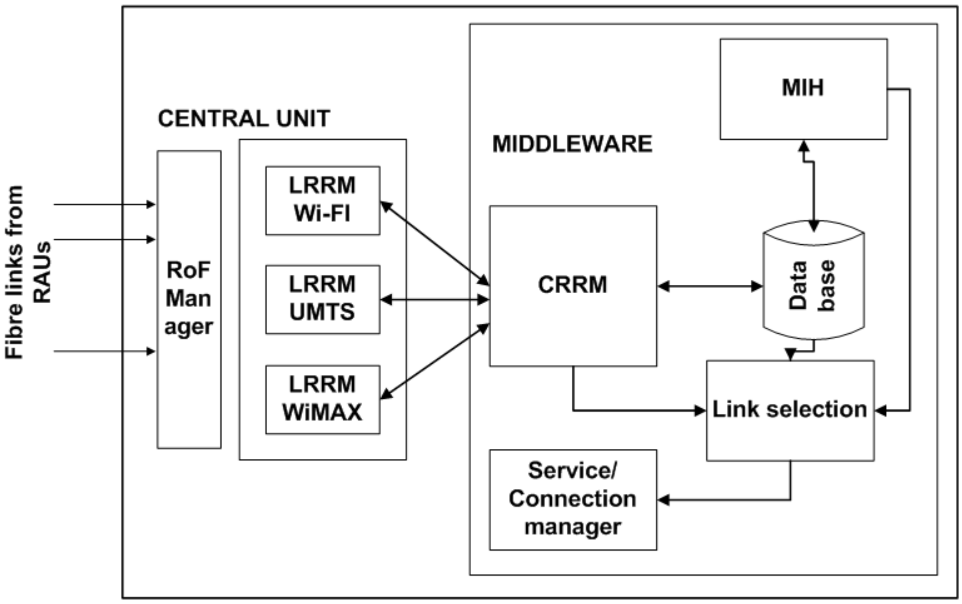


Figure 7.1: Generic Middleware Architecture

of a single RAT, whereas the CRRM is involved in the management of heterogeneous network with the help of LRRMs [4]. In Futon, both LRRM and CRRM are inside CU [5]. The management of the overall system resources is done in the CRRM, which is inside the middleware. Since LRRM and CRRM reside in the CU, the delay of control information between them is low; hence more functionality can reside in the CRRM and can take advantage of the availability of global information.

The LRRMs are responsible for the management of a specific RAT. It monitors the networks periodically and updates the CRRM, with the status of each network. The CRRM monitors the statistics of the overall network, which may consist of several RATs, and cells within the RAT. The CRRM manages the overall network, towards increasing overall capacity and coverage of the network, through vertical handoff, common admission control, common congestion control (CCC), load balancing, etc. Each LRRM manages the local Radio resource units (RRUs) based on the instructions from the CRRM and the local information of its cells. The CRRM concept in Futon is aligned with the 3GPP CRRM functional model. However, this particular implementation tries to take advantage of the centralized information processing from availability of data for various cells, networks and RATs, and the fact that LRRMs and the CRRM are placed inside the same CU.

The functionalities of the LRRMs are RAT-specific and independent from a central process, e.g. power control and the functionalities of the CRRM are designed as common, like common admission control, common congestion control, etc. Figure 7.2 shows the CRRM model in 3GPP. In a heterogeneous network scenario, the LRRM reports network measurements related to a RAT to its corresponding CRRM. Based on the measurements from LRRM, CRRM takes global decisions like accepting/rejecting new service flows, handover service flows, etc. Between CRRM entities, the information reports may be required to know the status of other LRRMs that are not directly connected to the CRRM. Thus all the radio network information is available at CRRM level and better decisions can be achieved.

In case of vertical/horizontal handoff, the CRRM triggers the handover due to network reasons and sends to LS module, the source RAT, and a candidate list of destination RATs in the order of their suitability. For example, handoff could be performed when the load in a network is above the congestion threshold. The execution of handoff will lead to load balancing between networks, which improves the overall capacity and coverage of the network and also mitigates interference. When the LS module receives a trigger, it selects the most suitable destination cell, RAT, link etc from the data available in the MDB,

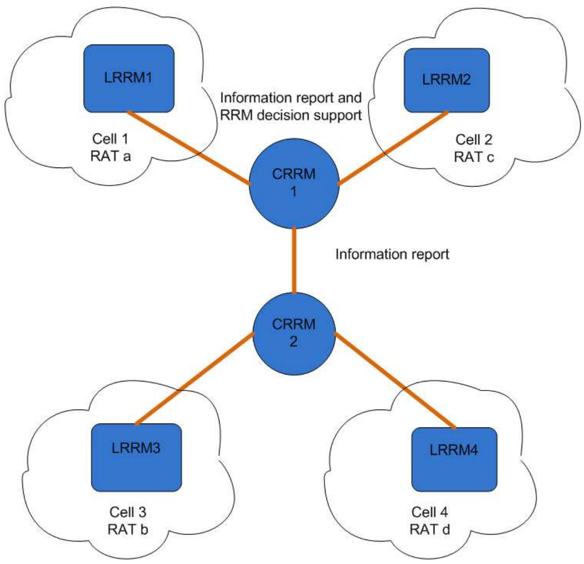


Figure 7.2: 3GPP CRRM functional model

updated by other modules of the middleware. The CRRM will take care of the VHO due to network reasons thus avoid congestion in any network.

Another feature of CRRM is AC [6], which admits or rejects a new user or a handover user. The main criteria for admission is, the service commitment made by the network to already admitted users is not violated, the network doesn't enter in to a congestion state and the QoS of the new user is met [7]. The decision made by AC should avoid bad admissions and bad rejections. In case of a bad admission, the user is admitted even though there are not available resources in the network, which degrades the QoS and user satisfaction. A bad rejection is rejecting a user even though there is a RAN that can meet the session requirements of the user, in which case the capacity is wasted and the operator's revenue is decreased.

Another feature of CRRM is CCC, which is triggered if AC fails or due to sudden increase in the network load exceeding a predefined threshold, which affects the existing services and the user satisfaction. The CCC works in two fashions, preventive and reactive. The preventive method ensures that the network does not get overloaded and remains stable, where as reactive method will react when congestion occurs and tries to bring back the network to stable conditions.

The design details of CRRM and LRRM implementation along with its interfaces to other modules of middleware, the tools and softwares used in the implementation, and the results obtained while performing vertical handoff between WLAN and UMTS are explained in detail in Section 7.2.6.

7.1.3 Media Independent handover (MIH)

The main objective of MIH [8] is to ensure multimedia service continuity across heterogeneous wireless access networks and guarantee QoS during handoff. The MIH mechanism handles the challenges of seamless mobility and session continuity at the middleware service platform. The MIH server is the key module of MIH functionality and it resides in the middleware. The MIH server hosts the handoff functionality (HoF) which triggers the handoff based on application and higher layer parameters. The MIH module interfaces with the LS module to trigger handoff from the user's point of view, where as CRRM raises a trigger for handoff due to network reasons.

IEEE Working Group has recently proposed the IEEE 802.21 standard to enable handover and interoperability between heterogeneous networks with context-awareness in mobile terminals. One of the main ideas behind IEEE 802.21 is to provide a common interface for managing events and control messages exchanged between network devices that support multiple interfaces, both wired and wireless. There is a synergy between MIH (IEEE 802.21) and MIP with fast handoff that will ensure QoS maintenance while network and physical conditions vary.

In IEEE 802.21 multiple services are deployed to optimize VHOs. In MIH, link layer intelligence and other network related information is provided to upper layers as shown in Figure 7.3. The services provided by the MIH include:

- Media Independent Event Service (MIES): It creates triggers when any change in physical, network or application parameters affect the ongoing session. MIES retrieves the necessary information from the MDB.
- Media Independent Command Service (MICS): It provides primitives to higher layers to control the lower layer functionality. MICS commands are used to obtain status of connected links and to execute higher layer mobility and connectivity decisions to the lower layers.
- Media Independent Information Service (MIIS): It provides a framework and corresponding mechanisms to obtain information about characteristics and services of existing networks, which optimize the handover decision.

The mobility management protocols combine dynamic information regarding link status and parameters, provided by the MICS, with static information regarding network status or other higher-layer service information provided by the MIIS, in order to help in the decision making.

7.1.4 Link Selection (LS)

The LS module selects the best RAT, cell site, link within a RAT, etc. for new user or handoff user based on the information available at the CU or reported by the mobiles, so that the QoS requirements of the new user and existing users are satisfied in the best possible way [9]. The LS interacts closely with the CRRM and MIH. The design and choice of the method for LS is a cross-layer and cross-system problem, and it is based on geographical position, signal strength, network load etc.

The LS finds the best RAT, cell or BS and if possible link for each user. The RAT selection [6] is done by considering QoS requirements, mobility, price factors etc. For example if a user has little mobility, but needs high datarate, WLAN is a good choice, and when a user is moving with high speed and needs low datarate, WCDMA might be the better option. While selecting the RAT, the cell selection should also be done, which is based on the signal strength received by the user from that particular cell and the network conditions of the cell etc. While selecting a RAT or cell the user preference and the cost factor could also be taken into consideration.

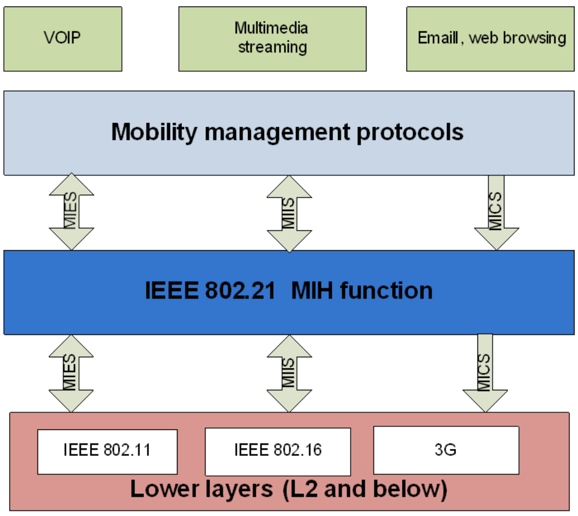


Figure 7.3: MIH framework

The CRRM initiates a trigger for a handoff due to network reasons e.g. traffic load, throughput etc, and send the trigger to the LS module with a list of available cells and RAT and rank them according to their suitability. The MIH module raises a trigger when a handoff is required from a terminal's point of view due to e.g. signal strength or QoS and sends it to the LS module. The LS module combines the information from the CRRM and MIH available in the MDB and selects the best link, cell site, RAT etc and sends this information to the SCM, which executes the handoff with the help of MIP.

In [6], we proposed a fuzzy logic based cell selection and fuzzy MADAM based RAT selection algorithm as shown in Figure 6.2 . The cell selection lies inside LRRM, which selects the best cell from the available cells based on signal strength, load of each cell and type of the call as shown in Figure 6.7. The list of best cells according to their rank or CAF (cell admittance factor), from cell selection module, is sent to RAT selection module in CRRM. The RAT selection selects the best RAT by taking the priorities from each user and the ranks from cell selection module as inputs.

7.1.5 Service/connection manager (SCM)

The SCM controls all the services provided by the middleware to provide seamless user connection between heterogeneous RANs and the IP world. Furthermore, it provides the connection management services including setting up of a session, continuity and termination of an application session and also responsible for managing the MDB

The SCM forwards the handover decision from the CU to the MIP core. It interfaces with the LS, which is the source of handover decision, the MIP core to send handover commands and the MDB to obtain information for converting the decision from the LS for a specific terminal to the specific network. The SCM also interacts with the RoF manager [3], which manages the network equipment on the optical front haul between the CU and all RAUs connected to the CU, as well as the communication links. It provides updated information to the SCM concerning the status of communication links and the occurrence of any fault, which is used by the SCM to initiate communication on new optical link.

7.2 Implementation of Futon CRRM

Figure 7.4 shows the system model used for the validation of CRRM and LRRM modules, which is explained in [2]. The system model includes LRRM, which manages WLAN and UMTS networks and updates CRRM periodically with network statistics, CRRM and other modules of middleware. A WLAN network is emulated with the help of an access point (AP), whereas UMTS network is simulated. Each module of the system model has an IP address to communicate with other modules using a TCP/IP socket connection. The LRRM and CRRM modules has local data bases to store local information, whereas the middleware contains a common database MDB to store the information from all modules of middleware.

In this implementation, both WLAN and UMTS networks are simulated/emulated, LRRM and CRRM modules are developed using C#, and the internal interfaces between WLAN-LRRM, LRRM-CRRM and external interfaces between CRRM-LS are developed using TCP/IP socket communication. The implementation of CRRM is validated by integrating with other modules of middleware developed by other partners of the project.

7.2.1 Tools and Software

The tools used in the implementation of CRRM include hardware tools like:

1. Access point (AP)
2. Router

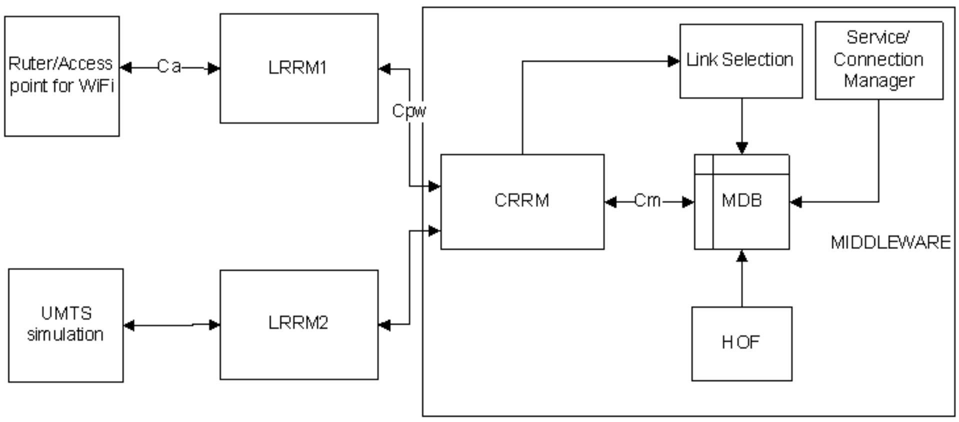


Figure 7.4: Implementation of CRRM

The software tools used in the implementation are as follows:

1. C# is used to develop the CRRM and LRRM
2. TELNET is used for communication with AP
3. Cisco IOs software is used to get network statistics from APs
4. MySQL database is used to store the read values from AP per cell
5. TCP/IP is used for communication between various modules
6. MATLAB for UMTS network simulation

7.2.2 WLAN Network emulation

A WLAN network is simulated using an AP that provides wireless connectivity using IEEE 802.11. WLAN is suitable for the case of limited mobility as the coverage range of WLAN access points is less compared to UMTS. The range of WLAN AP can increase upto 150 meters depending on the data rate and the protocol used. It can support the data rates from 1 Mbps until 54 Mbps, depending on the IEEE 802.11 protocol used. The basic settings done in the AP are IP address of the wireless interface and the service set Identifier (SSID). SSID is used by the mobile users and rest of the modules in the middleware to identify the network.

From a network point of view, the APs are connected to a switch, which provides an IP address to the interfaces connected to it using the DHCP server. The packets from the APs are routed to the next hop or the gateway using static routing. In this implementation the switch is the configured to have DHCP server, which assigns the IP address to the clients connected to it. The switch is also configured with Network address translation (NAT) to map the private IP addresses to public IP address.

In Figure 7.4, C_a is the interface between LRRM and AP used to obtain the statistics like load, throughput, and latency, total packets sent out and received, number of errors occurred etc., of a WLAN network. A TCP/IP socket connection was developed between AP and LRRM1 to receive the network statistics on periodic basis. LRRM obtain the network statistics from AP and updates the CRRM periodically.

7.2.3 UMTS Network simulation

The UMTS network is simulated to have realistic load and throughput of the network in every frame. Figure 7.5 shows the flowchart of the UMTS network simulator. A layout is generated, which involves generation of Node Bs with 1 km cell radius. In each cell the users are generated at random locations. Three types of services are considered as shown in Table 7.1, voice with 12.2kbps datarate and E_b/N_o requirement of 5dB [1], 144 kbps real-time data user with 1.5 dB of E_b/N_o requirement and non-real-time data user with 384 kbps with 1 dB requirement.

A service is assigned randomly to each user with equal probability. The users are moving with 30 km/h velocity in random directions. For each user the pathloss is calculated in every frame from current BS and from neighboring BSs using the C1 pathloss model [2] shown in eq. (7.1).

$$PL = \big(44.9 - 6.55log_{10}(h_{BS})\big)log_{10}(d) + 34.46 + 5.83log(h_{BS}) + 23(f_c/5) \quad (7.1)$$

Each user is allocated with a fixed number of codes depending on the data rate and length of the code or the spreading factor. The power transmitted by NodeB to each user is calculated and adjusted in every frame by $+/-1dB$ such that the received SINR at the terminal is equal to E_b/N_0.

In every frame, the downlink and uplink SINR are calculated for every user according to the equation shown below for user j

$$UL_{SINR} = \frac{(P_{ul}^j)}{\sum_{i=1,i\neq j}^{N} P_{ul}^i + Noise} \quad (7.2)$$

$$DL_{SINR}(at\ UE\ for\ user\ j) = \frac{(P_{dl}^j * PL^j)}{(I_{current}^j + I_{neighbor}^j) + Noise} \quad (7.3)$$

where P_{ul}^j is the power received at BS from user j, after pathloss and shadowing. P_{dl}^j is the downlink transmit power of user j. $I_{current}^j$ is the interference received

Table 7.1: Types of users

Service	**Datarate**	E_b/N_0
Voice	12.2 kbps	5dB
Real-time data user	144 kbps	1.5dB
Non real-time data user	384 kbps	1dB

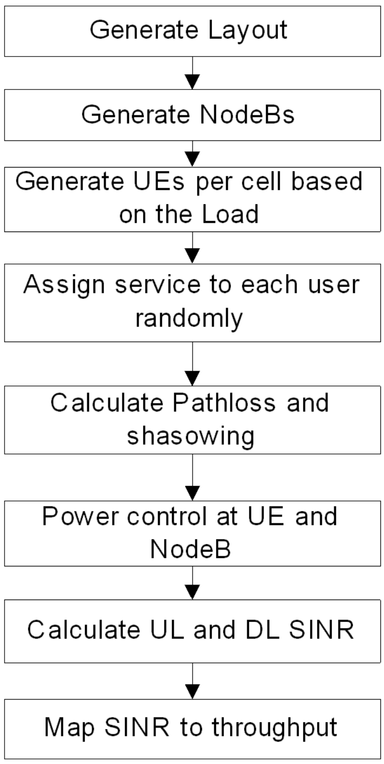


Figure 7.5: Flowchart for UMTS network simulator

by user j from current BS, whereas $I^{j}_{neighbor}$ is the interference received by user j from neighboring BSs. PL^{j} is pathloss of user j from current BS and N is the total number of users in the system.

From the SINR obtained for every user, the uplink and downlink throughput are calculated using Shannon mapping.

$$Throughput = BW.\log_2(1 + SINR) \tag{7.4}$$

The load of the UMTS network is calculated using the equation in [10] as shown below.

$$\eta_{DL} = \sum_{j=1}^{N} \frac{v_j.(\frac{E_b}{N_0})_j}{\frac{W}{R_j}}.[(1 - \alpha_j) + i_j] \tag{7.5}$$

where v_j is the activity factor, W is the chiprate, R_j is the bit rate of the user, i_j is the ratio of other cell to own cell interference and alpha is the average orthogonality factor and N is the number of users per cell.

7.2.4 CRRM

LRRM is a module used for the management of a specific RAT. The functionalities of LRRM are briefly classified as follows:

1. Monitoring each Cell in a RAT
2. Receive real time traffic measurements (RTTM's) for each Cell
3. Calculate KPIs
4. Store KPIs in local database
5. Forward alarms to CRRM
6. Enable local RRMs

In this implementation, LRRM receives RTTMs from each AP. LRRM opens TELNET connection with each AP and send commands (For example Cisco internetwork operating system (IOs) commands for Cisco APs) to receive the network parameters like load, throughput, etc. These parameters are parsed by LRRM and stored in the local database specific to LRRM.

CRRM is used for the global management of RATs. Each LRRM updates CRRM with network parameters in each cell with in the RAT in the form of extensible markup language (XML) messages. The functionalities of CRRM are:

1. Receives XML messages from each LRRM
2. Monitor network parameters
3. Triggers Vertical handoff and send to LS module
4. Updates middleware database periodically

The C_{pw} interface between CRRM and LRRM is a TCP/IP socket connection. CRRM receives XML messages on C_{pw} interface from LRRM periodically, and monitors the parameters extracted from XML messages on a periodic basis. If the load of any of the cell in a RAT exceeds the congestion threshold, then CRRM raises a trigger and sends to LS module. It also sends the candidate list of Cells in each RAT from network point of view. CRRM updates the middleware database with network parameters of each cell in a RAT.

7.2.5 Interfaces

The interface between CRRM-LS C_m is developed using TCP/IP socket connection. When there is a need for a VHO due to network reasons, or whenever the load, throughput, etc., exceeds the threshold in any cell inside a RAT, then CRRM raises a trigger. This trigger is sent to LS module, with source cell/RAT IP address and destination cell/RAT IP addresses in the order of the suitability. Each cell/RAT is given a rank according to the suitability based on load, throughput, etc. of the network. CRRM updates the $NetworkCell$ table in the middleware database with network parameters of WLAN and UMTS periodically. These parameters are used by other modules of the middleware to execute the VHO.

Figure 7.6 is the MySQL query used to store the parameters from CRRM in to middleware database.

```
INSERT INTO NetworkCell (CellID, CellType, CellLoad, CellLoadThreshold, AverageDataRate,
Capacity, MacAddress)
```

Figure 7.6: SQL query

7.2.6 Experimental evaluation of CRRM

The details of validation and demonstration of CRRM with other middleware components in FUTON environment is explained in [11]. The experimental setup for CRRM includes two WLAN access points. The users can transmit and receive 3G and WLAN. Due to unavailability of real 3G network data from operator, 3G network is simulated in MATLAB, which is a 4-cell layout with users randomly distributed and moved in random directions with 30 km/h speed. Table 7.2 shows the parameters used in the simulation of 3G network.

With this experimental setup two scenarios are verified to check the functionality of CRRM. In scenario 1, 3G network is congested, which is done by increasing the offered carried load in the simulation and the load of a WLAN network is reduced either by decreasing the number of users in the network or by decreasing the data rates of the existing users. In scenario 2, WLAN is highly loaded, which is done by increasing the number of users or by increasing the data rate of current users.

CRRM is executed with the experimental setup as shown in Figure 7.4 . Table 7.3 shows the snapshot of the network statistics obtained from wireless AP. Figure 7.7 shows the load profile of WLAN network, when the network is congested and existing user is handed off to the neighbor network.

Table 7.2: Simulation parameters

Parameter	Value
Center frequency (fc)	2 GHz
Frequency re-use factor	1
BW	5 MHz
Chiprate	3.84Mcps
Frame length	10 ms
Number of slots / frame	15
BS transmit power	43 dBm
Mobile station transmit power	24 dBm
Spreading factor	256
Mobile speed	30 kmph
Duplexing scheme	FDD
Number of NodeBs	4
Propagation Model	C1 Pathloss model [10]
Datarate for every type of user	[12.2 144 384] kbps
Eb/No for every type of user	[5 1.5 1] dB

Table 7.3: Network statistics from Wireless AP

Parameter	**Value**
Interface	Dot11Radio0
Hardware	802.11N 2.4Ghz
BW	1000 Kilobit
txload	61.5%
rxload	0.78 %
output rate bits/sec	616000 bits/sec, 48 packets/sec
Number of packets	472125
Output errors	241
Reliability	1

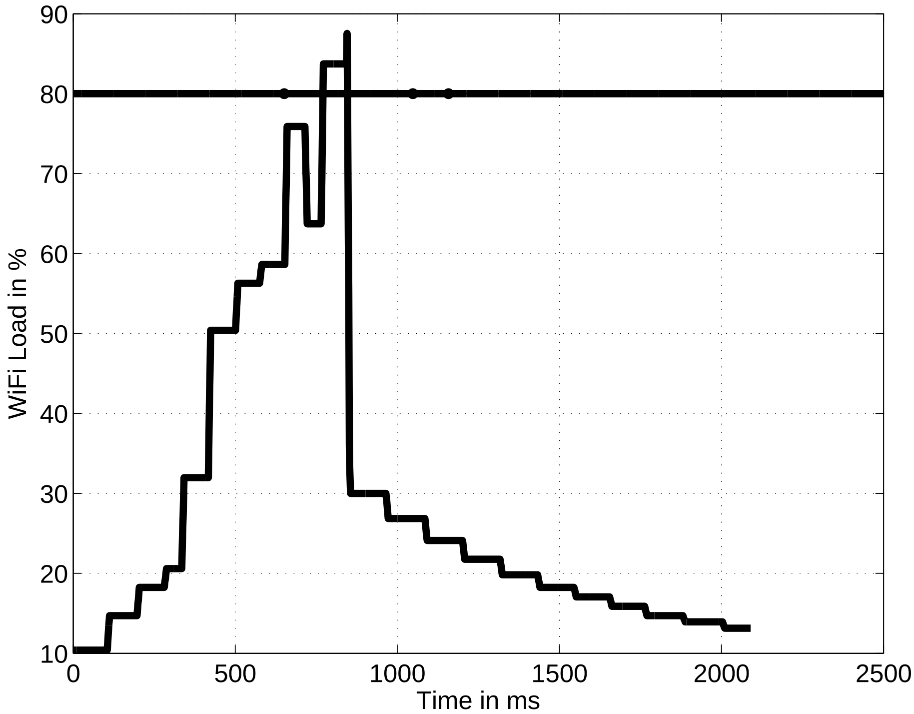


Figure 7.7: WLAN load profile during handoff

Figure 7.8 shows the simulated load of 3G network for scenario 1, when the 3G network is congested. It can be seen that there is a handoff at frame 70, where the load in the network starts to reduce. Figure 7.9 shows the downlink throughput of the 3G network when the carried load in the network is at 80%. When the load of any network exceeds the threshold, a trigger will be sent to LS module, which is shown in Figure 7.10.

7.3 RoF Manager

FUTON distributes the management of the hybrid optical-radio infrastructure existing between the fibre-optic interface of the CU and the RAUs in two main systems: the RoF network manager, and the middleware or CRRM, the latter being responsible for the management of the radio part of the network. FUTON brings a new paradigm by moving the UMTS Node B and the RNC functionalities together into the CU.

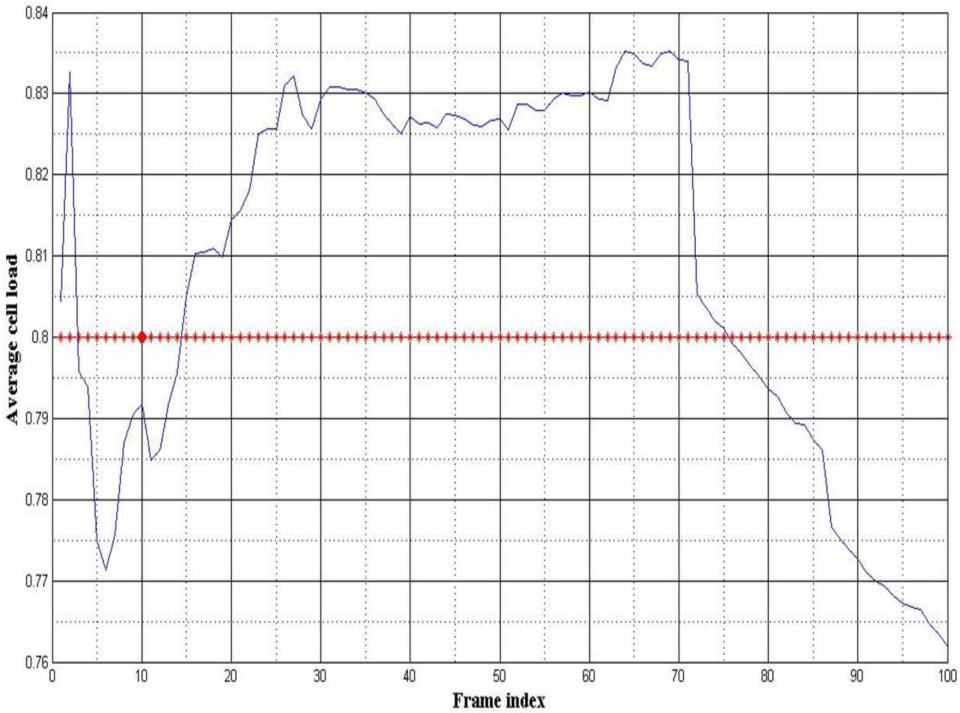


Figure 7.8: 3G load profile during handoff

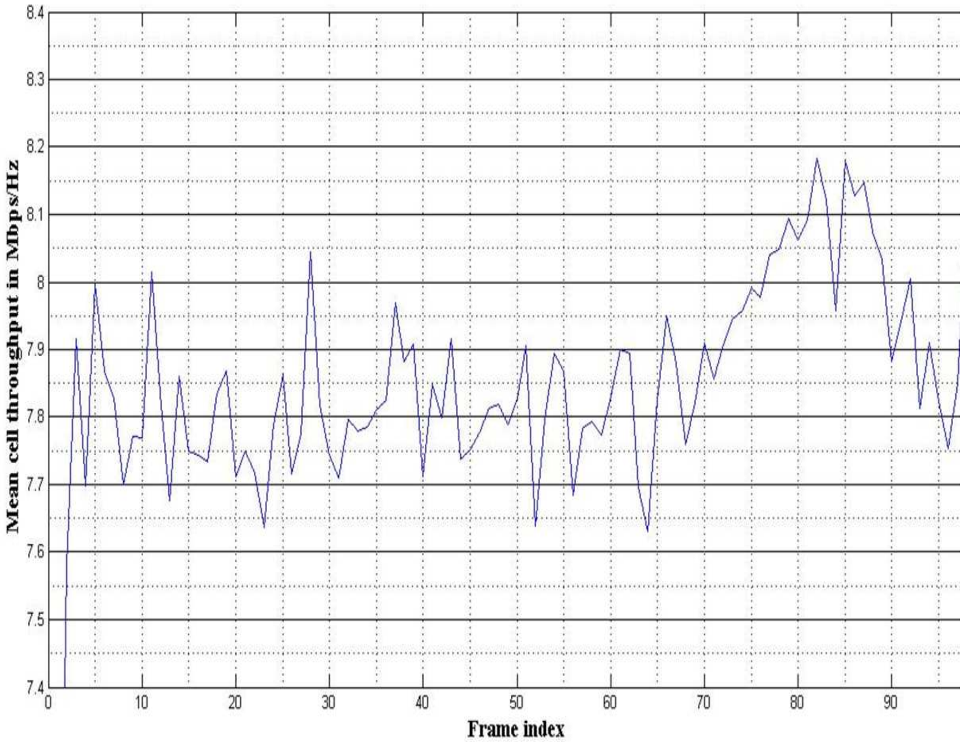


Figure 7.9: Mean cell throughput of UMTS

<timestamp>time</timestamp>
<WLAN>
<Ipaddress>10.110.12.93</Ipaddress>
<Rank>85</Rank>
</WLAN>
<UMTS>
<Ipaddress>127.0.0.1</Ipaddress>
<Rank>30</Rank>
</UMTS>
<WLAN>
<Ipaddress>10.110.12.100</Ipaddress>
<Rank>75</Rank>
</WLAN>

Figure 7.10: Trigger from CRRM to LS

The RoF manager [3] [12], shown in Figure 7.11 manages the network equipment on the optical front haul between the CU and RAUs connected by it, as well as the communication links. It provides updated information to the middleware database concerning the status of communication links and the occurrence of any eventual fault, thus enabling end-to-end service problem resolution and service quality management by the FUTON middleware.

The most crucial functions of the RoF manager are:

- Configure the optical components of the RAUs
- Monitor the optical components and RF signal power of the RAUs
- Manage the optical link status
- Update the middleware database when the status of the optical links have changed

The RoF manager presented here follows the basic characterization of network management functions "FCAPS" - Fault [13], Configuration, Accounting, Performance [14] and Security. The RoF manager introduces some key requirements and capabilities for the support of a FUTON-like RoF infrastructure, but only concentrated towards managing network performance, network faults and configurations for convergent networks.

The functionalities of the RoF manager are described below. The RoF manager monitors all network elements (RAUs) from the CU up to the antenna while providing and updating information concerning network topology and optical link status. The later information will be available for middleware and CRRM. The operating logic of the NEs is not interfered with by the control messages. RoF manager operates exclusively in the signal control plane and thus has no access to the user specific data or flows. The RoF manager also has exclusive access to NE data. All other components of the CU that need to have access to NE data have to make a request for them directly to the RoF manager.

Figure 7.11 shows the RoF manager architecture explaining the interworking within the different blocks [12]. The main components of RoF manager that controls the network and network equipments are listed here: Configuration management (CM); Fault management (FM); Performance management (PM); Security management. These sub-systems constitute the base platform and are integrated with other FUTON modules, namely middleware, CRRM and the RAUs.

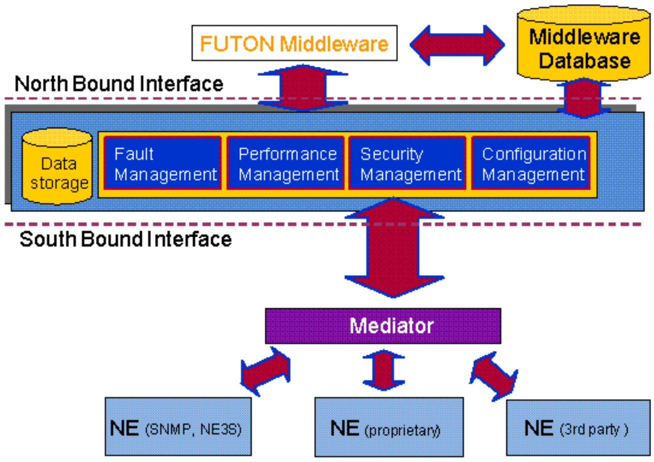


Figure 7.11: RoF Manager Architecture

Figure 7.12 shows the interfaces of RoF manager with other modules of FUTON architecture [12]. It can be seen that RoF manager interfaces with optical infrastructure at RAU and CU and between them.

7.3.1 Implementation of Authentication

The Security of network elements is very important in Futon infrastructure due to the existence of heterogeneous RATs in a single geographical area sharing a common infrastructure. Also the infrastructure contains different NEs like RAUs, optical components and other subsystems which gives rise to security issues. The goal of security management in RoF manager is to provide security to the Network elements.

The Triple-A based protocol is considered to provide authentication to NEs. All the NEs like RAU, databases, servers, and switches/routers etc., are authenticated by the RoF manager before giving access to CU. The RoF manager includes Triple-A client, which takes the requests from the NEs and forwards the requests to the Triple-A server, which authenticates the NEs. In this section the implementation of authentication in security management module is discussed.

The implementation of security management in Futon [15] is based on RADIUS protocol [16]. The scenario here is, a certain NE with username and password tries to gain access to the CU. The message flow of authentication process, starting from the request by NE until it is authenticated by the server

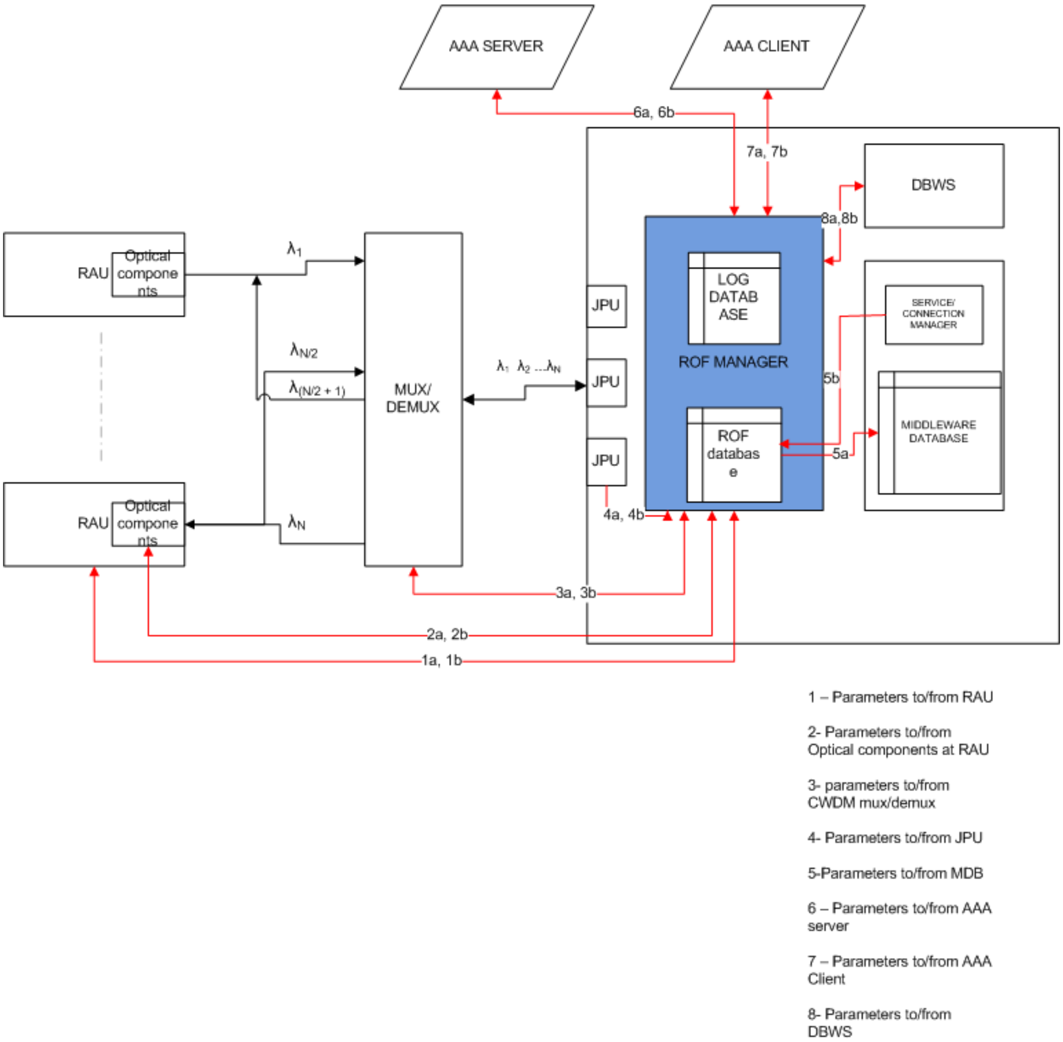


Figure 7.12: RoF Manager interface diagram

is shown and explained in this section. Table 7.4 shows parameters used in the validation security implementation.

The user in this case a NE with username *futon*, which sends the access request message with the password encrypted with password authentication protocol (PAP) protocol using MD5 hash function. The access request message is converted in to a radius packet with attributes in type-length-value (TLV) format. Figure 7.13 shows some of the attributes in access-request message before being sent to the triple-A server. We can see that the attributeType(s) follow

Table 7.4: Parameters used in security validation

Parameter	Value
Protocol for security	RADIUS
Username	futon
User password	test
Authentication protocol	PAP
Shared password	testing123
Hash function	MD5

the RADIUS specification where attributeType=1 is user-Name, attribute-Type=32 is NAS-identifier, and attributeType=1 or user-Name attribute contains bytes of data called "*futon*" in ASCII format as user name.

Figure 7.14 shows access request message by client after password has been encrypted with PAP protocol using MD5 hash function.

The access request message is sent to the triple-A server in TLV format as an array of bytes, according to Radius specification. The access request Message sent in TLV format is received by triple-A server, which reconstructs the message from the attributes mentioned in the packet. The password in reconstructed message is decrypted using the shared password, which was used by the user in encryption. The user password matches with the decrypted password, and access accept message has been sent to the user. Figure 7.15 shows an access accept message that is received by the user, which means that the provided username and password in access-request message was correct and leads to a successful authentication.

7.4 Conclusions

In this chapter 7, a generic middleware architecture is presented to ensure seamless mobility and service continuity through IPv4/IPv6 MIP core. The modules of middleware, which are SCM,LS, CRRM, MIH etc are explained. One of the main functionality of middleware is vertical handoff, which is triggered by either CRRM or MIH, and the LS selects the best link, cell site and RAT for handoff. The middleware presented is generic and can be used in the next generation networks.

The implementation of CRRM, which is one the main modules of middleware is presented. The simulation/emulation details of WLAN and UMTS network are shown. The CRRM and LRRM modules are validated by integrating

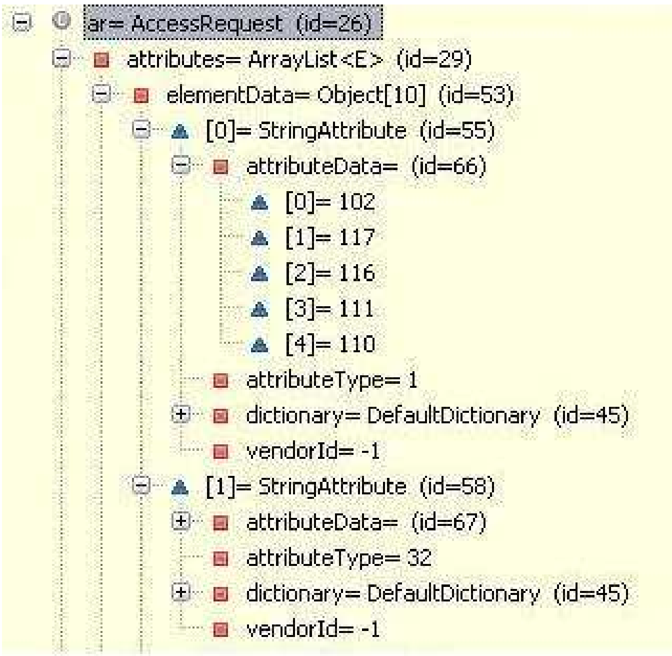


Figure 7.13: Message attributes of client before the authentication

Figure 7.14: Access request by client after password has been encrypted

with other modules of middleware. The results obtained for vertical handover scenario are presented for the case of WLAN and UMTS. The implementation

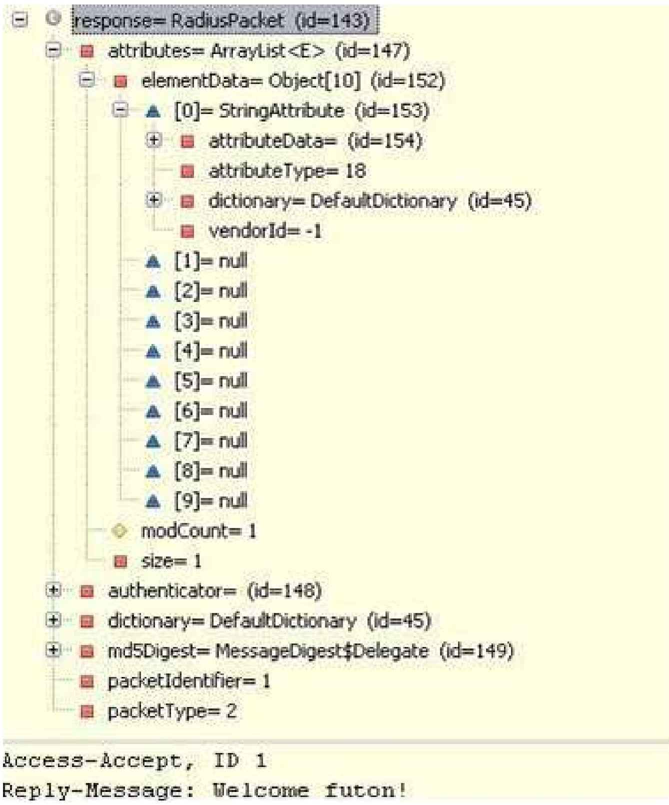


Figure 7.15: Authentication response received by user

of CRRM can be used to further extend and add more functionalities like AC ,interference control etc.

The RoF manager is an important module in any RoF network for managing the NEs. The main modules and framework of RoF manager is presented and the implementation details of authentication procedure is explained. The authentication is important in a network to check the identity of any NE before giving access to the other modules of the network. Here the Triple-A based RADIUS protocol is used for implementing the authentication in RoF manager.

References

[1] Sofoklis Kyriazakos,George Karetsos, "Practical Radio Resource Management in Wireless Systems", Artech House, ISBN 1580536328, April 2004 FP7 EU-funded ICT Project FUTON, at http: www.ict-futon.eu.

[2] FP7 EU-funded ICT Project FUTON, "Implementation of FUTON Middleware Module," Deliverable 4.4, Sep 2010

[3] Carlos Santiago, Bodhisattwa Gangopadhyay, Artur M. Arsenio, M.V.Ramkumar, Neeli R. Prasad, "Next Generation Radio over Fiber Network Management for a Distributed Antenna System," Wireless Vitae, pp. 182 - 186, May 2009.

[4] M. V. Ramkumar, Albena Mihovska, Neeli Prasad, "Management of Distributed and Extendible Heterogeneous Radio Architectures," NAEC 2009, Italy

[5] FP7 EU-funded ICT Project FUTON, "CRRM Functional Architecture and Interference Scenarios on FUTON Project," Deliverable 4.1, May 2009.

[6] M.V. Ramkumar, Albena Mihovska, Neeli R Prasad, Ramjee Prasad, "Fuzzy-logic based Call Admisison Control for a Heterogeneous Radio Environment," Proceedings of Wireless Personal Multimedia Communications, Japan, Sep 2009. Italy

[7] M. V. Ramkumar, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad, "QoS-Guaranteed Admission Control for OFDMA-based Systems," IEEE Globecom, pp. 606-610, Dec 2010.

[8] J.Rodriguez, M.Tsagaropoulos, I.Politis, S.Kotsopoulos, T.Dagiuklas, "A Middleware Architecture Supporting Seamless and Secure Multimedia Services across an Inter-technology Radio Access Network," IEEE Wireless Communications Magazine, vol. 16, no. 5, pp. 24-31, Oct 2009.

[9] FP7 EU-funded ICT Project FUTON, ”Final specification: Link configuration and context transfer, Network discovery, Cross-Layer Algorithms, Link selection policies.,” Deliverable 4.3, Jan 2010

[10] Harri Holma, WCDMA for UMTS, 2nd ed. , Wiley, Sep 2002.

[11] FP7 EU-funded ICT Project FUTON, ”Demonstration of FUTON Middleware Module,” Deliverable 4.5, Oct 2010

[12] FP7 EU-funded ICT Project FUTON, ”Specifications of the RoF management module,” Deliverable 6.1, July 2009.

[13] FP7 EU-funded ICT Project FUTON, ”Fault Management Sub-System,” Deliverable 6.3, March 2010.

[14] FP7 EU-funded ICT Project FUTON, ”Performance Management Sub-System,” Deliverable 6.4, June 2010.

[15] FP7 EU-funded ICT Project FUTON, ”Validation of the RoF Management Module,” Deliverable 6.5, September 2010.

[16] RFC 2865, ”Remote Authentication Dial In User Service (RADIUS),” June 2000

8
Self Healing

In this chapter we discuss the challenges in self healing and presents a 4-stage generic framework for self healing, which includes fault localization, KPI thresholding and monitoring, fault modeling and fault diagnosis. Fault localization deals with locating the fault in the network, which narrows down the problem and hence the number of symptoms to consider in fault modeling, and is helpful in a huge network with large volumes of data. The KPI thresholding and monitoring stage deals with monitoring various parameters and setting right threshold for each parameter and provide the symptoms to fault modeling stage. The fault modeling stage finds the root cause with available symptoms. Lastly, the fault diagnosis finds the solution to the root cause from the previous history of cases. The main focus of this chapter is the challenges in self-healing and the presentation of a generic framework for self-healing with proposal of existing methods applicable for each module in the framework. The framework addresses issues in autonomic management and improves QoS while lowering the operating expenses (OPEX).

8.1 Introduction

Due to a rapid increase of data rates and increase in the variety of applications with different quality of service (QoS) requirements there is a huge increase in the number of terminals and hence the size of the network and volumes of data to consider for self healing. In case of a next generation network (NGN), to support these diverse QoS requirements of the users, heterogeneous networks existing in the same geographical area is considered. The existence of several radio access technologies in NGN induces high complexity due to various network elements that support the heterogeneity of these RATs in a single infrastructure. In this context self-healing or automated diagnosis of faults is a real challenge due to heterogeneity in the network elements, services or applications and RAT to support the communication. With these challenges the need for self healing is highly essential. Manual fault diagnosis causes the total downtime of the system to increase and also the total cost of the diagnosis increase due to the human intervention. Hence self healing mechanisms are useful and very crucial to reduce the downtime caused by the failure and the human intervention. The reduction of OPEX and downtime improve the overall performance of the network, QoS to the users and hence increase overall revenue for an operator.

One of the main components of a network management system, responsible for running a network at optimum level and concerned with the automatic detection, isolation and recovery from faults is self-healing. A network manager requires self-healing to detect, log, and notify users of problems and automatically fix network problems, keeping the network running effectively, since faults can cause downtime or unacceptable network degradation. While designing self-healing system, it should be taken into account that there will be large number of nodes, heterogeneity exist in hardware, software, vendors and network will be dynamic in behavior.

A Self organizing network (SON) [1]-[4] deals with planning, configuration, deployment, optimization, and maintenance of the network and self healing play major role in the maintenance part of the SON framework. In [1]-[4], the architecture of SON is proposed, which specifies the interfaces between various blocks inside the SON framework and the interfaces between terminal and wireless network. The self healing module should communicate with other modules of the SON system like CRRM, network planning module, spectrum management module etc., with the main objective to diagnose the faults of the network with the available symptoms from the network and other management modules in SON.

In [5], a global fault management architecture is proposed for heterogeneous network scenario, which uses model-based diagnosis module. In [6], a hybrid diagnosis model is proposed using the belief networks [7] and case based reasoning. Also in [8], a hybrid model for automated network management is proposed using neural network and case-based reasoning (CBR). In [9] fault diagnosis system is proposed based on CBR. The problem is using the CBR approach alone in fault diagnosis is its inability to adapt to new faults, as this approach uses the past history. In all above works a single method based on neural networks or belief networks is used or a hybrid method with combination of two methods are proposed. In this chapter we propose a generic self healing framework with four stages, fault localization, KPI thresholding and monitoring, fault modeling and fault diagnosis. Each stage of the self healing framework performs a specific functionality which can be implemented using various methods. The main focus of this chapter is to provide the challenges in self healing and give a survey of the available methods in the literature for each step of the proposed self healing architecture.

The chapter is organized as follows, in Section 8.2 the main challenges in the self healing are discussed. In Section 8.3 the generic self-healing architecture is proposed and each stage of the architecture is discussed. Sections 8.4 - 8.7 surveys the existing methods in the literature for fault localization, KPI thresholding and monitoring, fault modeling, and fault diagnosis respectively. In Section 8.8 the BMAP discretization technique is evaluated. In the last section, the chapter is concluded discussing the pros and cons of each method.

8.2 Problem Definition

A fault is a malfunction or misbehavior that has occurred in any part of the network, either in hardware or in software. The faults can be classified as secondary faults if these are triggered by other faults in the system, and primary faults [10] if they are not caused by other events. Furthermore, faults occurred in the system can be classified in to permanent, intermittent and transient [11]. A permanent fault exists in the system until it is corrected, where as intermittent faults occur for short periods of time and vanish and then reappear and they are sometimes early indicators of permanent failures. Transient faults appear and then vanish.

An alarm is an event triggered by a fault and alarms are raised by the management system when the situations cross the expected behavior. Alarms can be raised by a NE facing the problem or by network elements affected by the problem, for example neighbor network elements. Alarms can be classified as physical alarms which could be hardware errors like power failure etc., and logical alarms [8] which are statistical errors like excess load in the network etc. Also the alarms can be significant and candidate [12], where significant alarms will definitely appear when a fault occurs and candidate alarms may or may not occur. In a bipartite graph between faults and symptoms, the significant and candidate alarms are differentiated depending on the value of the conditional probability between that fault and symptom. Alarm may exist for certain period of time till the fault gets rectified and can be weighted as critical, high and low [13] depending on the severity level and the impact of the alarm on the overall behavior of the network. An instance of a fault may cause several alarms to be raised by the management system. Alarms raised may depend on dependencies among NEs, current configuration, services in use since fault occurrence, presence of other faults, values of other NEs. Due to this system knowledge may be inaccurate and inconsistent.

The faults can be identified from symptoms, which are raised as alarms by the management system when certain variable exceeds the threshold. Hence the thresholding mechanisms are very important for catching the right symptoms at the right time. The inefficiency of the thresholding mechanism may cause loss of symptoms or may generate spurious symptoms [14].

Hence with problems, faults, alarms, symptoms, a self-healing system tries to find the root cause of the problem, with the help of alarms and symptoms. From the root cause, the appropriate solution or diagnosis method is suggested by self healing system. During the process of finding the root cause and the solution to a problem, the challenges faced by the self-healing system are as follows:

- Fault evidence may be
  - ambiguous, for example same alarm may be generated as an indication of many faults [15] [10]
  - inconsistent, due to disagreement between NEs, one may think its fault other may not
  - incomplete, when there is alarm loss or delay
- Generation of Multiple alarms, which are due to the following reasons [15]
  - Due to transient faults
  - Duplicate alarm by same device

  - Duplicate alarms from multiple devices
  - Or it could be alarms by secondary devices due to primary fault

- Isolation of multiple simultaneous faults. In this case, the self healing system should identify and separate corresponding alarms for each fault [10]. The simultaneous faults could be dependent or independent.
- Scalability, able to cope with the increasing complexity in terms of size, applications, etc., [16][17]. Due to the size of the network it's difficult to process large volumes of information of the entire system; hence distributed management system can resolve the scalability problem.
- Temporal relationships between the events, all alarms may not be raised at the same instant [18][10]; the effect of the fault may be seen after sometime. Window based techniques where the alarms are collected over a time frame can resolve this issue to some extent.
- Alarms occur in bursts hence there is less time for the self healing system to react on each alarm [18]
- As the network grows, the addition of new network elements and new software may change the characteristics of the alarms and alarm sequences.
- For the case of a wireless network [15], there are more challenges apart from the mentioned above. Some of them are as below
  - Due to fading and environmental conditions, the probability of symptom loss is higher
  - Due to restricted BW and power, the amount of information transferred is less; hence self healing may have to deal with fewer symptoms or incomplete data.
  - Due to latency or delay in the information, the temporal effects are more in wireless networks
  - And due to noise and interference the data may not be accurate

8.3 Self-healing Architecture

Figure 8.1 shows the proposed model for self-healing which has four modules, fault localization, KPI thresholding and monitoring module, fault modeling (FM) and fault diagnosis (FD) module. All the four modules are connected to a database to store the relevant information, which can be reused by same module or by other modules, for example KPI thresholding module finds the conditional probability of a cause given a symptom, which may be used by FM module.

Fault localization (FL) module locates the problem, example the segment in which the fault occurred. The location of the problem could be a NE, or a software module, class or object or a software bugs inside the code [19], or kernel level issues like memory, CPU resources etc. Fault localization module receives the alarms from various segments, both primary and secondary segments where secondary segments are influenced due to primary segments. Hence FL module finds the primary segment where the fault initially occurred. The main purpose of fault localization is to reduce the number of symptoms and data to process in a huge network, such that the KPI thresholding can only process the symptoms related to the segment.

KPI thresholding and monitoring module take the parameters from the network or the system under monitoring and perform statistical analysis of the data and calculate the symptoms of the problem in the form of KPIs. If any KPI exceeds a threshold in case of a fault then it is treated as a symptom. This module finds all the affected parameters related to the segment received from the FL module, and the corresponding logical alarms are raised and sent to the next level. Discretization techniques are used in this module, which divide the continuous data in to intervals with proper selection of thresholds.

After receiving the logical alarms from the KPI thresholding module, fault modeling module finds the most suitable cause or causes by mapping logical alarms with possible causes by using Bayesian, neural network, fuzzy logic. etc., methods. The most suitable cause/causes from fault modeling module along with the KPIs/symptoms is sent to the next module, which is fault diagnosis module. FD module creates a new case for each cause with corresponding symptoms obtained, and tries to match with the existing cases stored in a database. This is done using CBR (case based reasoning) techniques or rule based techniques or any other event correlation techniques [20][21][12]. Once it finds the most suitable case close to the new case, it is then adapted to the present problem and tested on the segment where the fault has occurred to check if the problem is resolved with this solution. Once it finds the solution it is stored in the database and will be used in the future.

The next four sections discuss in detail about fault localization, KPI thresholding, modeling and diagnosis and various methods existing in the literature in performing these routines.

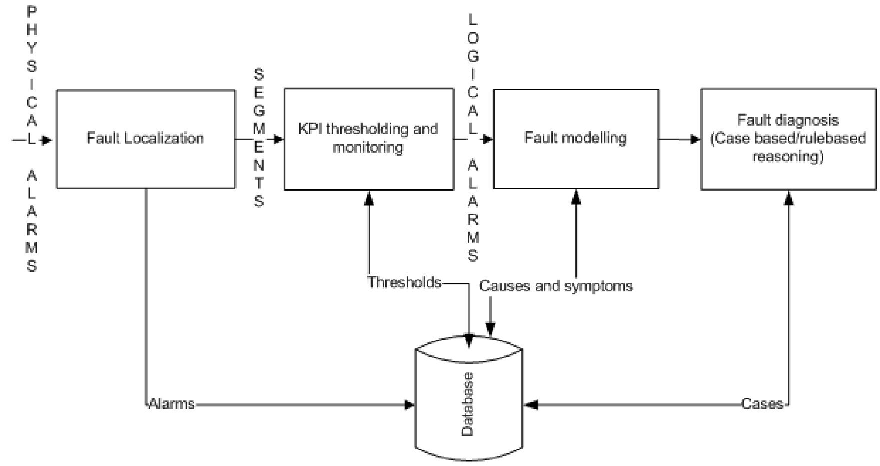


Figure 8.1: Self-healing Architecture

8.4 Fault Localization

Fault localization (FL) module finds the location of the problem or the source of the problem. The faults that can happen are broadly two types, hardware faults and software faults. Hardware faults, for example could be antenna tilt, power failure of specific NE etc., and software faults, for example could be excessive pilot power, excessive call drops, memory leaks, insufficient CPU resources, lack of throughput, etc., for a wireless network. Therefore the fault localization module detects the location or source of the problem. The fault occurred might cause degradation in the performance of a specific NE, which in turn might cause the degradation of few other network elements. In this case, the fault localization module finds the root segment by correlating the alarms raised by various segments. If the root alarm is solved, the other alarms will also get resolved.

Alarms related to a single fault may be received in an out-of-order manner at different time instants, where all the alarms must be considered to increase the accuracy of fault localization. In case a single fault results in a large number of alarms raised by primary and secondary NEs, the volume of alarms need to be reduced by using alarm filtering techniques. In general alarm filtering contains four processes: compression, counting, suppression and generalization. Compression removes multiple occurrences of the same alarm, whereas counting substitutes similar alarms with a new alarm, suppression stops a low-priority alarm in the presence of a high priority alarm and generalization refers an alarm by its super class.

FL module receives alarms raised by the network elements which are affected by a fault. Each alarm received by the fault localization module represents the fault from the point of view of the NE raising an alarm. Hence the alarm raised by a NE may contain partial information about the root fault. The FL module collects these partial views and correlates the information to find the root fault.

Fault localization module also updates the database whenever it receives an alarm. The alarms are received by the fault localization module from various NEs, and the function of FL module is to reduce the volume of alarms by applying alarm correlation techniques. Hence memory management is also one of the important functionality of this module. A number of methods for performing fault localization have been proposed in the literature, are discussed in this section.

8.4.1 Alarm filtering and correlation

A model based alarm correlation is proposed and implemented on intelligent management platform called IMPACT [18]. The model described here contains two components, which are structural component that describes network elements and their connectivity, and behavioral component that describes about the alarms and the correlation rules between the alarms. Here the structural component contains the network configuration model which describes NEs and their connectivity with other NEs and network-element class hierarchy which describes various classes of NEs and the relationship between them. The behavioral component contains three parts; message class hierarchy, correlation class hierarchy and correlation rules, where the message class hierarchy describes the relation between the messages raised by various NEs.

In [12], a hierarchical reasoning method for alarm correlation is proposed; where an automated fault diagnosis system called alarm correlation view (AC view) is used for efficient fault localization and identification in a multi-domain environment. The fault propagation is modeled as a bipartite causality graph where faults manifest themselves as aggregate alarms emitting from the network elements affected by those faults. In this model a segment based alarm is classified in to two groups as the significant and candidate alarm, and the fault localization model derives the belief values about the faults that trigger significant alarms and locating the most likely NE where the fault has occurred.

Alarm correlation engine (ACEngine) is implemented in AC view. The front end of ACEngine collects the significant alarms and deduces all the suspicious faults in each NE by using alarm to fault mapping. In this model the network is divided in to various levels and domains, and the domain manager for each level monitors the network elements in its domain, performs fault reasoning with the data from its own NEs and lower level NEs, and delivers the inferred results to the upper level. Using the hierarchical alarm correlation, the location of the problem can be found. This architecture can be used when there are simultaneous faults in the system. But this model depends on the accuracy of the probabilities of the alarm in a segment, when a fault happens in other segments.

In [23] the fault localization is done using a framework that identifies frequently occurring episodes from sequential data. A network alarm sequence analyzer is proposed, which discovers knowledge from an alarm databases, and its called KDD (Knowledge discovery in databases) process. The KDD process contains two steps, pattern discovery and post processing. The algorithms used for knowledge discovery in this framework are based on rule-based formalism.

8.4.2 Fuzzy logic based Fault localization

FL based on fuzzy logic [24] is performed from inference rules, which are usually obtained from the past history of data. In [25] fuzzy logic is used for the fault localization, by using the trouble ticket history record. In case of a fault, the monitoring system send the symptoms observed to the fuzzy-based management system, which determines the membership grade from the knowledge available and gives a priority list of suspicious devices and the degree to which these might be faulty.

Fuzzy based FL is not suitable to use for huge networks, for example with 1000 nodes as it needs a huge database that can maintain the symptom description and suspicious devices that could cause the problem. This method would be more useful, when implemented in a distributed fashion, where each network segment has its own fuzzy-based management system, and sends the priority list of faults to the centralized management system, which can take decision based on the received inputs from all lower-level managers.

8.4.3 Distributed Fault localization

In communications the networks are partitioned in to distinct management domains, based on geographical or technological or policy or organizational reasons. These domains may be disjoint, overlapping or nested. In many instances, faults and their effects are localized, and do not propagate throughout the network. In most cases the effect of a fault is seen in its own domain or in its neighboring domains. But the centralized management system has to process the information from all the domains, which will make the process cumbersome. Therefore if the fault localization process is distributed, then the information from one domain is shielded to many other domains [17].

In [16] and [17] the distributed management is discussed, where the case of a fault effecting the network elements in different management domains is considered. In this case, the management center builds the domain of an alarm from the database. The domain of alarm is the set of all independent NEs that could have caused the alarm or the NE's that might be at fault due to an alarm. If the domain of an alarm has NEs beyond the domain of management center then the fault has to be solved in a distributed fashion, with a mutual understanding between management centers. The cluster of alarms is the set of all alarms with intersecting management domains, and the best solution that explains the cluster of alarms should

1. Explains all alarms in the cluster
2. The solution should have the least number of faulty NEs

In the distributed management system, each domain estimates the cost it would incur for explaining the alarm, and communicates this information to the neighboring domain. For example, if domain one knows the cost of domain two, then it would represent the NEs of the alarm in domain two as proxy network elements. A probabilistic algorithm [16] is applied to all alarms that cross the boundary, by assuming that the number of faults in the network is less than k, a parameter. In the distributed approach there is a possibility that manager in the NE also fail, in which case the information cannot be communicated.

8.4.4 Neural Network approach

Neural networks [15][26]
[8][27][13] treat fault localization as an abductive problem to identify the set of devices that explains the alarms. Neural networks [5] contain interconnected nodes called neurons, which operate like a human brain. They learn from the input data and are resilient to the noise in the input data. The disadvantages of neural networks are that they need long training data [15]. Self-organizing map (SOM) based artificial neural network can be used for alarm correlation and fault identification [26].

8.5 KPI thresholding and monitoring

KPI thresholding and monitoring module monitors the KPIs and provides symptoms in case of a fault where one or more KPIs exceed the threshold. Another function of the KPI thresholding and monitoring is to aggregate the variables and KPIs, for example sum, average, extreme values, percentiles and histograms [28]. The type of aggregate used depends on the type of the variable. For example, network throughput, spectral efficiency highest measures are required, where as call blocking rate, call dropping rate, number of errors for different variables, etc., lowest measures are required. As this module contains statistical data concerned with the overall network, this can be used to improve the overall performance of the network or to take decisions from an overall network point of view.

The choice of KPIs that need to be monitored should be done carefully such that the chosen KPIs are relevant to the faults or problems occurring. The parameters should be selected in such a way, that the fault modeling has a high probability of finding the root cause. KPIs are particularly important in case based reasoning in order to select the best case based on the cause and the symptoms. KPIs are also used in adapting the existing case to the new case. Once the parameters to be monitored are chosen, the main functions of KPI thresholding and monitoring are to find thresholds for variables, modify thresholds within the confidence level, to raise alarms based on the thresholds and to update the database with the thresholds.

Once the parameters to be monitored are chosen then the main functions of KPI thresholding and monitoring module are

1. Find the most suitable threshold for each variable
2. Modify the threshold or continue step (1) till stopping criterion or confidence level is obtained.
3. Raise alarms whenever a parameter crosses a threshold and forward to next module.
4. Consistently update the database with thresholds

8.5.1 Discretization techniques

Discretization techniques are used to find the most suitable threshold for a given KPI. This section discusses various discretization techniques available in the literature. The goal of discretization is to find a set of cut points to partition the range in to a small number of intervals that have good class coherence, which is usually measured by an evaluation function. Usually discretization involves two tasks, first task is to select the number of intervals and second task is to find the width or the boundaries of each interval given the range of continuous attribute.

The discretization methods can be categorized as static and dynamic, top-down and bottom-up, etc. [29]. The difference between static and dynamic depends on whether the method takes feature interactions in to account. Static methods like entropy based partitioning, IR algorithm, etc., determine the number of partitions for each attribute independent of other features, whereas dynamic method capture interdependencies in feature discretization. In a top down approach, one big interval with all values corresponding to a feature is partitioned in to subintervals until stopping criteria is reached, while the bottom-up approach starts with few number of intervals and combines these intervals until the stopping criteria is reached.

In discretization, a wrong selection of threshold will result in spurious symptoms, which affects the performance of the fault modeling when locating the root cause. KPI thresholding and monitoring is the main component in the architecture to remove spurious symptoms and reduce the number of lost symptoms. Feature extraction methods are used to remove spurious symptoms and a wrong choice of feature selection algorithm may result in lost symptoms. Complexity is one of the issues in this module as it need to process huge amount of data from various parts of the network to find KPIs.

Equal-width and frequency method

The most obvious simple method, called equal-width interval method [30][31], is to divide the attribute into N intervals of equal size. Thus if A and B are low and high values respectively, then the intervals will have width $W = (B-A)/N$ and the interval boundaries will be at $A+W, A+2W, A+(N-1)W$. In Equal-frequency method the interval boundaries are chosen such that each interval contains equal number of values.

Chi-Square based method

Chi-square is a statistical measure that tests the relationship between the values of a feature and the class. The top-down method based on chi-square called Chisplit searches for the best split of an interval by maximizing the chi-square criterion applied to two sub-intervals. The interval is split if both subintervals differ statistically by a considerable amount. The bottom-up approach based on chi-square called Chimerge [31] searches for the best merge by minimizing the chisquare criterion of two intervals. The two intervals are merged if they are statistically similar.

While merging intervals in Chimerge approach, X2 value is computed for each interval and adjacent intervals are merged with lowest X2 value. Merging process continues until X2 value reaches a X2-threshold. Hence choosing higher X2-threshold results in fewer and larger intervals, whereas lower X2-threshold results in too many intervals. If any attribute results in to one interval then that attribute will be removed.

Chi2 [32] is an automated version of Chimerge. C4.5 is used to verify the effectiveness of Chi2. It is shown that the Chi2 algorithm effectively removes unwanted or noisy attributes, when the merging process ends up with single interval. In Khiops [33] method, discretization starts with a single interval and evaluates the merges between adjacent intervals and selects the best one according to the chi-square criterion applied to all intervals. The merging is stopped when the confidence level is reached and doesn't decrease anymore. Confidence level is computed with chi-square statistic.

8.5.2 Entropy based methods

Entropy Minimization Discretization (EMD)

Entropy minimization discretization methods [34] [30] partition the symptoms from a training set of cases by sorting the cases in increasing order of symptom value and at each cut point the data is divided into two subsets and the class entropy of the partition is computed. The threshold is selected for which the entropy is minimal among the entire candidate of cut points.

Let $|D|$ denote the number of cases in a subset D and the total number of causes C is divided in to the causes related to symptom j as C_r^j and causes not related to symptom j as C_n^j. Let $|D(C_r^j)|$ is the number of cases in D whose cause belongs to C_r^j and $|D(C_n^j)|$ is the number of cases in D whose cause belongs to C_n^j. The class entropy of the subset D is defined as:

$$Ent(D) = -\frac{|D(C_r^j|}{|D|}.\log_2(\frac{|D(C_r^j|}{|D|}) - \frac{|D(C_n^j|}{|D|}.\log_2(\frac{|D(C_n^j|}{|D|}) \tag{8.1}$$

EMD [35] has been adapted as follows. For each continuous symptom S_i:

- The cases in the training set D are first sorted by increasing value of the symptom S_i. The midpoint m_j between the symptom values in each successive pair of cases in the sorted sequence is considered as a potential threshold m_j.
- Each candidate threshold m_j partitions the data set of cases D into two subsets D_1 and D_2. The class entropy of the partition is evaluated as described below.

$$Ent(D, m_j, S_i^C) = \sum_{k=1}^{2} \frac{|D_k|}{D} Ent(D_k) \tag{8.2}$$

  where $Ent(D_k)$ is the entropy of subset D_k, which is calculated using eq (8.1).
- The best threshold t_i for S_i^C is the candidate threshold which minimizes the class entropy of the partition.

Thus, a binary discretization for D is determined by selecting the threshold t_i for which $Ent(D, m_j, S_i)$ is minimal among the entire candidate cut points m_j. Minimum description length (MDL) method is used as the stopping criteria.

EMD calculates class entropy of a subset over all the causes. However a symptom is related only to some causes. Hence SEMD [36] is the selective entropy minimization discretization, which differentiates the causes as, related to symptom(C_r) and not related to symptom (C_n). Hence the complexity of SEMD is 2/K times of EMD.

8.6 Fault Modeling

Fault modeling finds the root cause of the problem or possible list of causes, with a probability or rank for each cause, which helps during the testing process if the most probable cause does not solve the problem the next cause is tried to find a solution.

Most fault modeling approaches are based on probability and depend on the previous history to obtain the prior and conditional probabilities. The performance of these methods depends on the amount and accuracy of the available data and on the accuracy of the symptoms. Fault modeling methods should consider the effect of lost and spurious symptoms while finding the probability of the root cause. Symptoms obtained due to network disorder are called negative symptoms, where as symptoms that are not observed by the management applications are called positive symptoms and by including the effect of the positive symptoms, the accuracy will increase.

Another issue in fault modeling is the occurrence of simultaneous dependent or independent faults and while mapping symptoms with causes, the symptoms related to the fault should be considered. When a fault has occurred, the symptoms due to the fault may occur over a period of time, which depends on the nature of the fault and therefore fault modeling must consider the temporal effects of faults.

8.6.1 Bayesian Modeling

Bayesian modeling is one of the popular approaches of fault modeling for finding the root cause of the problem from the symptoms obtained from KPI thresholding module. Bayesian [7] [36] approach can infer possible causes even when the data is uncertain and imprecise [13]. In this approach, a directed acyclic graph (DAG) is used with nodes connected, indicating the probabilistic dependence of the cause to a particular symptom. The graph is built using the method of induction based on the frequency of occurrence of the symptoms from the information available in the database [13]. The fault modeling [36][35] using Bayesian approach contains two steps

- Knowledge acquisition, which represents the knowledge required to identify root cause, which includes selecting the right symptoms and causes.
- Applying Inference method algorithm that identifies the cause of the problem based on symptoms

The knowledge acquisition consists of two steps, qualitative and quantitative steps. In qualitative step the causes and symptoms for a diagnosis are identified. Here the causes are modeled as discrete random variables with two states: absent/present, and the symptoms are KPIs received from KPI thresholding and monitoring module. The KPIs are modeled as discrete random variables with any discrete number of states representing the range of a KPI e.g.: normal, high, medium, and low.

In quantitative model, probabilities related to causes and symptoms are specified. The Probabilities required are the following:

- Prior probabilities of causes: $P(C_i)$ is the probability of the problem being caused by cause C_i.
- Conditional probabilities of the symptoms given causes: $P(S_j/C_i)$ is the probability of the symptom S_j given the cause C_i.

Once the quantitative and qualitative models are defined, then inference method consists of calculating the probability of each possible cause, given the values of symptoms $S_1, S_2...S_M$ obtained as

$$P(C_i/S_1,...S_M) = \frac{P(C_i).\prod_{j=1}^{M} P(S_j/C_i)}{\sum_{n=1}^{K} P(C_n)\prod_{j=1}^{M} P(S_j/C_n)} \tag{8.3}$$

where $P(C_i)$ is the prior probabilities of the causes and $P(S_j/C_i)$ are the probabilities of the symptoms given the causes. The output of the system is a list of causes ranked by their probabilities.

The prior probabilities of the causes are obtained from experts or calculated from training data as the frequency of occurrence of each type of cause. The calculation of the probabilities from the training data can be done using maximum likelihood (ML) estimation or using m-estimate (MEST) method.

In maximum likelihood estimation (MLE), the probabilities are calculated based on the frequency of occurrence. In MEST, the prior probability of a cause C_i is estimated as the expectation of the probability of the cause happening in the next trial when there were $|D(C_i)|$ cases where the cause was C_i in the $|D|$ previous cases. And the conditional probability of symptoms given the causes is computed using the m-estimate.

The Bayesian method of fault modeling is easy and widely used. But the disadvantage of using this approach is that it depends on the previous history to obtain the prior and conditional probabilities. Hence the performance of this method depends on the amount and accuracy of the available data.

8.6.2 Incremental Hypothesis updating

Incremental hypothesis updating (IHU) [37] creates a set of most likely hypothesis, based on the information from arriving symptoms and constantly upgrades them. Each hypothesis is a subset of causes that explains all symptoms that are observed. The method of IHU algorithm is to find subset of causes that maximizes probability that all causes in the subset occur and each observed symptom is explained by at least one cause in the subset.

Each hypothesis explains a symptom, if it contains at least one cause that explains a symptom and each hypothesis is ranked with a belief metric b. If a new symptom is found then the set of hypothesis is updated. On occurrence of new symptom a belief metric for every hypothesis is calculated which explains the probability

- All causes belonging to hypothesis have occurred and
- Hypothesis explains every observed symptom

The symptoms obtained due to network disorder are called negative symptoms, where as the symptoms that are not observed by the management applications are called positive symptoms. The inclusion of positive symptoms increases the accuracy of fault localization. In the method of IHU, the effect of positive symptoms are included in calculating the total belief of the each hypothesis. Some symptoms though they are observed by the management system they are lost due to errors in the transmission. In this method, the symptom loss is also considered while calculating the belief metric.

The advantage of this model is that it takes in to account the effect of positive symptoms and also the symptom loss, which increase the effect of accuracy of finding the faults. Also it updates the hypothesis continuously and provides a list of hypothesis to the operator, so that if the first hypothesis doesn't work then there is an opportunity to test second hypothesis according to the rank based on belief metric.

8.6.3 Divide and Conquer approach

Divide and conquer [10] method uses a dependency graph and assume that one failure is allowed per object, where object is a part of the network, for example a node, a layer in protocol stack, software process hardware component, link, etc. This method produces a set of fault hypothesis with a confidence measure for each hypothesis. It considers a time window, during which the alarms are collected. Each object depends on the other objects and the faults occurred in one object propagate through other object or effects the other object.

Each node in the dependency graph represents the terminal objects represented by e_i with weight p_i, which is the probability that e_i fails independent of any other object. The directed edge (e_i, e_j) has a weight p_{ij}, which is the probability that e_j fails as a result of the failure in e_i. An object e_i can raise alarm a_i, and the domain of alarm $D(a_i)$, is defined as the set of objects that might have cause alarm a_i. For a given cluster of alarms denoted by A there exists a $S = \bigcup(D(a_i))$, $\forall a_i \in A$. For each e_i in S the probabilities p_i and p_{ij} are assumed to be known. The objective of divide and conquer algorithm is to find the minimum subset of S such that

1. The entities in the chosen subset explain all the alarms in A
2. Maximize the probability that at least one of the objects in S is a fault

This algorithm runs in two phases. In the first phase, the alarm cluster domain S is partitioned in to two subsets, with each set having the maximum mutual dependency, which is done based on the dependency weight. In the second phase, a set of symptoms that explains all the received alarms and has the maximum probability that at least one of the entities in the set is a fault is found. This algorithm solves the problem of the simultaneous independent faults and also captures all the symptoms occurring at different times as it uses a time window. This model can be further improved by using dynamic time window based on the type of alarms occurred, as the effect of faults on time scale depends on the type of faults and also on the situations. But this algorithm doesn't handle lost or spurious symptoms.

8.6.4 Codebook approach

The subset of symptoms selected to provide the desired level of distinction between the causes is called codebook [38]. The codebook approach uses the dependency graphs consisting of events and root causes as nodes and directed edges to represent the dependencies. The graph is transformed in to correlation matrix, as shown in Table 8.1, where the columns in the matrix represent root causes while rows represent events. In the codebook approach each problem is denoted by a vector of 1's and 0's, where the length of the vector represents the symptoms considered in the system. In the vector, a 1 represents occurrence of the symptom due to the problem and 0 represents non-occurrence of the symptom. Figure 8.2 shows the correlation graph between symptoms S_1, S_2, S_3 and causes C_1, C_2, C_3.

For example, if the first bit of the each code represents symptom S_1, the second bit represents S_2 and third bit represents S_3 and the code for C_1 is (1, 1, 0); code for C_2 and C_3 are (1,1,0) and (0,1,1). The problem of finding the root

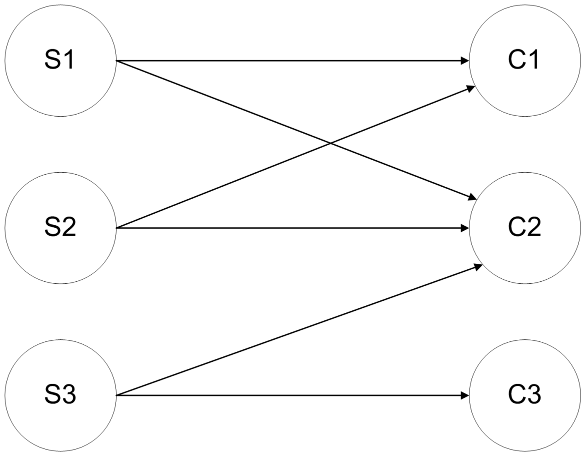


Figure 8.2: Correlation graph

cause is defined as finding the cause whose codes optimally match an observed symptom vector [38]. The complexity of correlation is a function of number of symptoms considered. The set of causes and symptoms as represented in the Table 8.1 represents the correlation matrix.

The distinction between the causes is measured by Hamming distance and radius of a codebook is defined as half the minimal distance between the codes. In the example shown in Table 8.1, the minimum hamming distance is 3 and the radius is 1.5. Hence it can guarantee the distinction by a single symptom. Using this method the lost or a spurious generation of symptoms can be corrected and detected. In general the observation errors in $k-1$ symptoms can be corrected and k symptoms can be detected as long as k is less than or equal to the radius of the codebook.

Table 8.1: Example of codebook with radius 1.5

	C_1	C_2	C_3	C_4	C_5	C_6
S_1	1	0	0	1	0	1
S_2	1	1	0	1	0	0
S_3	1	0	1	0	1	0
S_4	1	1	1	0	0	1
S_5	0	1	0	0	1	1
S_6	0	1	1	1	0	0

8.6.5 Model-based approach

In [39], a model based fault modeling is proposed for WLAN network, which consists of structural and behavioral model (SBM) constructed from structural causality and behavioral statistics from network measurements. The model's causality structure is built from the knowledge of protocol specifications and WLAN configuration. A logic based backward reasoning is proposed for finding the root fault by back tracing the causality structures whenever the network behavior deviates from the expected behavior. The SBM model is updated when the network configuration or network conditions changes.

Once the SBM is constructed, the fault modeling module determines the component in the SBM that is the root cause of the fault by recursively checking the upstream inputs for any mismatches. This method of finding the root cause is called the effect-mismatch algorithm (EMA). The architecture that leverages SBM and EMA is decoupled, hence easy to update the SBM due to changes in the network.

8.6.6 Context-free grammar

Context-free grammar [40] is a method to represent the dependencies between the NEs, using the expressions that are built from sub-expressions. Two algorithms were proposed in [40], which find the minimum set of faults that explains all the alarms. In the algorithms, the lost and spurious symptoms are considered. The advantages of this model, is that it can tackle scenarios with multiple faults and gives the solution by considering the spurious and lost symptoms.

8.7 Fault diagnosis

Fault diagnosis (FD) module finds the solution to the problem by using the past history of cases. It receives a list of possible causes from FM module with the probabilities of failure for each cause. The symptoms obtained for each cause are also mentioned by FM module and also updated in the database. FD module creates a new case for every cause with its symptoms and any other variables to be used in the process of solving the fault, and tries to find/match for a similar case in database. Once a similar case is found, the case is adapted such to the current situation by adjusting the parameters. For example if the root cause of a problem is low_throughput for a user and if the solution for the cause is to increase the BW to the user, then the variable BW need to be adapted such that the user gets expected throughput, which may depend on SINR etc.

8.7.1 Case-based reasoning

Case based reasoning can be used in fault diagnosis as it provides a good framework to use past solutions for present problems. The goals of CBR system are [9] [5] [6] [8]

1. To learn from experience
2. To offer solutions to new problems based on past experience
3. To avoid extensive maintenance

The main steps involved in CBR are to retrieve, adapt and update as shown in Figure 8.3. Retrieving involves finding the case from the database which is close or similar to the current situation. The next step is to adapt the retrieved case to the current situation by adjusting the parameters. The last step is to update the the database with the current solution for future use.

For example, the structure of a case is as follows $Case = case - id, situation, fault, action1, action2, ...frequency$. The situation is used to describe the effected segments in the network, from their device names, IP addresses, Class or object in a software etc. Given the situation, the fault is described with a list of causes and symptoms, the list of actions that could rectify the fault, and the frequency of this case is mentioned, so that the new case can start searching for more frequent cases.

A case is retrieved from the database which is similar to the current case. The current case is compared with existing set of cases using k-nearest neighborhood (K-NN) approach or using decision tress [41]. K-NN method uses various similarity metrics to retrieve the closest case, such as the Euclidean distance, geometric similarity metrics, probabilistic similarity metric etc. Eq. (8.4) measures the similarity between the cases, which is based on hamming distance with weights and partial similarities between values of parameters p [6].

$$sim(c, c') = \frac{\sum_{p \in P} w(p) sim_p(v_c(p), \hat{v_c'}(p))}{\sum_{p \in P} w(p)} \tag{8.4}$$

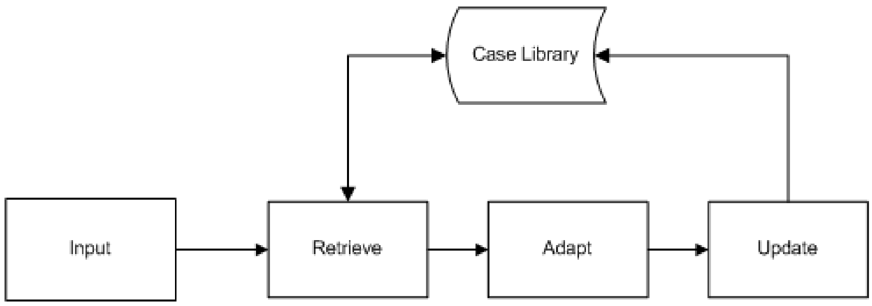


Figure 8.3: CBR Model

The case retrieved best suits as a solution when the parameters used to find the similarity are relevant to the fault and cause and larger the number of matches between parameters larger is the similarity.

Parameterized adaption [9] is one method of adaption in which the variables in the current case are adjusted based on the relation between solution and variables in the retrieved case. Below example explains the parameterized adaption, if the problem faced by a specific user in a network is "high BER" and the solution obtained from CBR is to increase the transmit power to the user, then the amount of increase should be adapted in such a way that BER received by the user is as expected, by taking the path loss and fading in to consideration.

Other way of adaption could be to avoid a particular solution which can be done by ranking the case to a low value, so that it is not used in future. Critic-based adaption [9] occurs when a critic repairs a retrieved case manually to fit the target case, example if the transmit power has to be increased by 1dB, then the function to adjust transmit power based on $BER\ tx\_power = f(BER)$ is done if the tx_power doesn't reach the maximum value. So the function changes to $tx\_power = f(BER, max\_power)$. Adaption can also be done using the rules based on similarities and differences [8].

8.7.2 Hybrid Rule-Based/Case-Based Reasoning

Rule based reasoning [20][21] contains a set of rules to match an event. If no appropriate rule exists, the problem is transferred to the rule based reasoning module, where the problem has to be solved by hand at first and then stored in the rule database. This new rule can be used as basis to update the rules and in case of reoccurrence of the same situation, it will be matched to the former situation stored in rule database.

In rule-based reasoning (RBR), a set of rules is used to actually perform the correlation. The rules have to form "conclusion if condition". The condition contains received events together with information about the state of the system, while the conclusion may consist of actions which lead to changes of the system and can be input to other rules. The rule based system contains knowledge base that contains the facts and rules, and inference engine [42]. The knowledge base is a repository about the system, with facts defining the objects, situations, etc., and the rules that recommends the type of action to be taken. The Inference engine is the central controlling component and its basic purpose is to determine the way rules are to be applied.

8.8 Experimental evaluation

In this section we considered the beta PDF method of discretization for finding symptoms in a LTE wireless network. The problem considered is too many dropped calls caused due to either "downlink interference", "uplink interference" or "lack of coverage". The symptoms considered are uplink SINR and downlink SINR. Due to lack of real network data the causes, interference and coverage, are created through simulations.

All the users are moving in a cell in random directions. A user is expected to face interference when he comes closer to a boundary, both in the uplink and downlink. The power control is turned off both in the uplink and downlink, so that the effect of interference is seen more clearly when the user comes closer to the boundary.

Another cause for call drops in a network is due to coverage problem in the downlink. The possible reasons for coverage problem are shadowing and fading occurred in the propagation path. Here the lack of coverage is created by adding more shadowing loss to a random user. The power control has been turned off, so that effect of shadowing losses is more visible. If the power control is not turned off, the shadowing losses are overcome with an increase of the power. When a user face the coverage problem due to more shadowing losses, then there is a degradation in the SINR.

Figure 8.4 shows the uplink SINR symptom denoted as UE_SINR_UL due to the cause of "uplink interference". It can be seen that uplink SINR reduces at around 200th frame. Similarly the downlink SINR symptom UE_SINR_DL due to the cause of "downlink interference" is seen at around 800th frame in Figure 8.5. Figure 8.6 shows the downlink SINR symptom UE_SINR_DL due to the coverage problem.

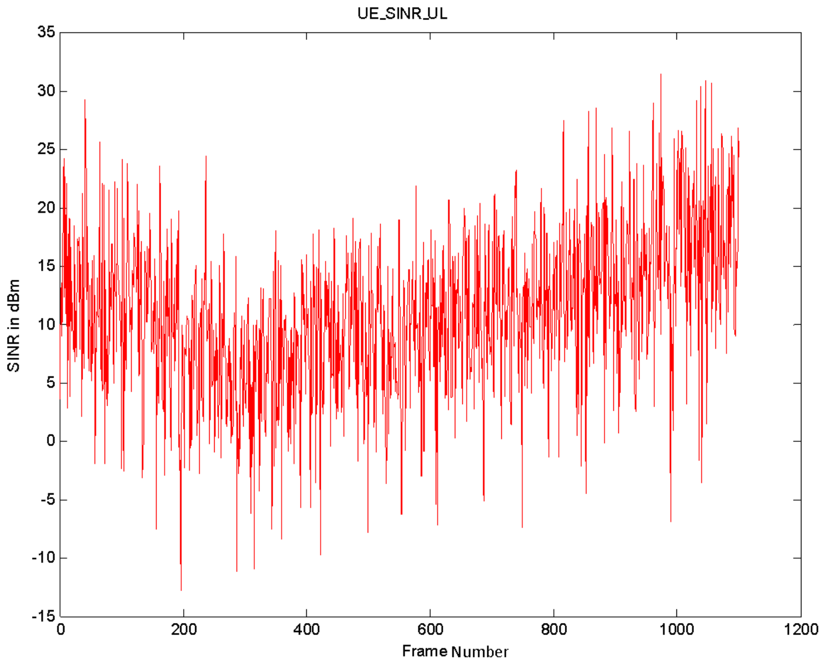


Figure 8.4: UE_SINR_UL due to uplink interference

For every user that experience interference or shadowing loss, the symptom values are approximated with a beta pdf. In order to find the threshold, we considered two sets of training data. The first set of training data is related to the cause, whereas the second training set of data is not related to cause i.e. the data which is free from interference as well as shadowing losses. The second set of training data is generated in which the boundary limits for movement of users have been reduced, and users don't see much interference or shadowing losses. Figure 8.7 shows the curve fitting of the both training data sets for downlink SINR, with and without interference. From the beta PDFs, the threshold due to downlink interference is calculated as 13.15dB.

Once the thresholds are calculated for every parameter, the symptoms are obtained when a parameter exceeds the threshold. The probabilities can be found from the training data. Once the probabilities and the symptoms are available, the root cause can be found using bayesian approach as mentioned in Section 8.6.1.

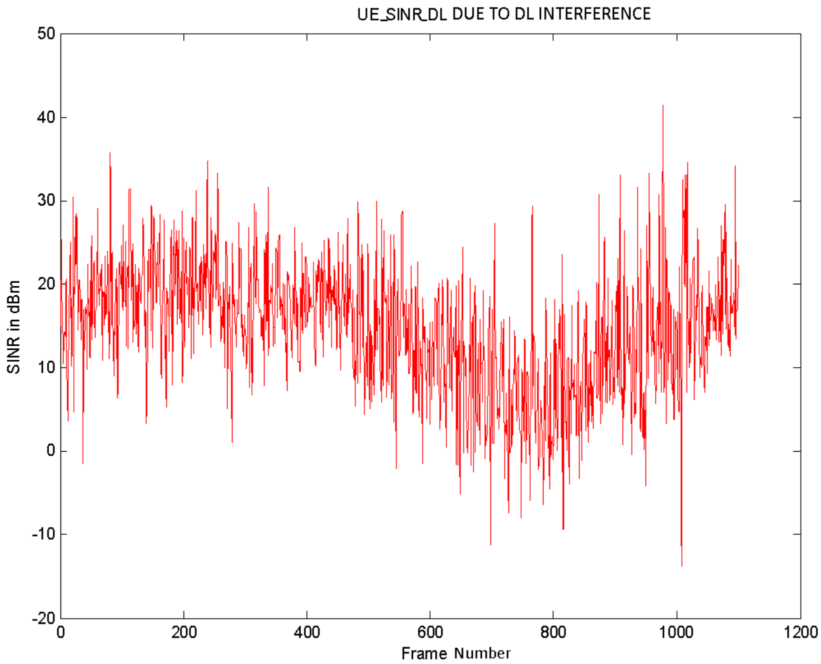


Figure 8.5: UE_SINR_DL due to downlink interference

8.9 Discussion

This chapter discusses a generic 4 stage self healing frame work applicable to any kind of network. For each stage various options existing in the literature are discussed. One of the issues in fault localization is memory management; any effective alarm correlation technique can be used to reduce the volume of alarms. The temporal effect of faults, which refers to the time instants at which each alarm related to a fault is raised and the order in which alarms are raised need to be addressed in FL. A time window based alarm correlation techniques can be used to reduce the temporal effects where the length of time window may be different for different faults. Another problem in fault localization is occurrence of simultaneous faults. In this case the faults should be isolated and alarms raised should be mapped separately for each fault which is done in hierarchical method for AC [38]. FL using fuzzy based reasoning method needs list of inference rules, which are difficult to obtain and also to maintain for a huge network. Hence distributed FL is the best approach in a huge network to reduce the amount of data to deal with, but there is a possibility that manager in the NE also fail.

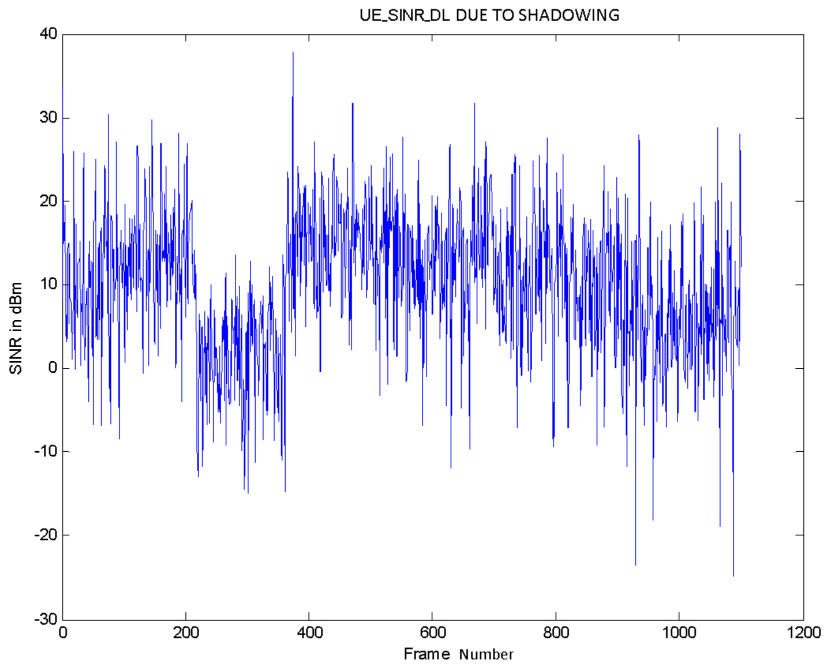


Figure 8.6: UE_SINR_DL due to lack of coverage

In KPI thresholding techniques the main issue is a wrong selection of threshold, which results in spurious symptoms and hence the accuracy of fault modeling and diagnois will be affected. Hence this module is the key in the architecture. The spurious symptoms should be removed by this module and also number of lost symptoms should be reduced. The feature extraction methods are used to remove spurious symptoms, and a wrong choice of feature selection algorithm may result in lost symptoms. The fault modeling should consider the effect of lost symptoms and also positive symptoms to improve the accuracy in finding the root cause [37]. Complexity is also one of the issue in this module. SEMD is one alternative to reduce the number of operations and decrease the complexity.

In Fault modeling, bayesian approach depends on the previous history to obtain the prior and conditional probabilities. Hence the performance of this method depends on the amount and accuracy of the available data. The advantage of incremental hypothesis updating method is it updates hypothesis whenever it receives a new symptom and provides a belief metric to rank the possible causes. Also this method takes the effect of positive and lost symptoms in finding belief metric, which increase the accuracy of finding fault. Divide and conquer algorithm takes the effect of simultaneous independent faults but it

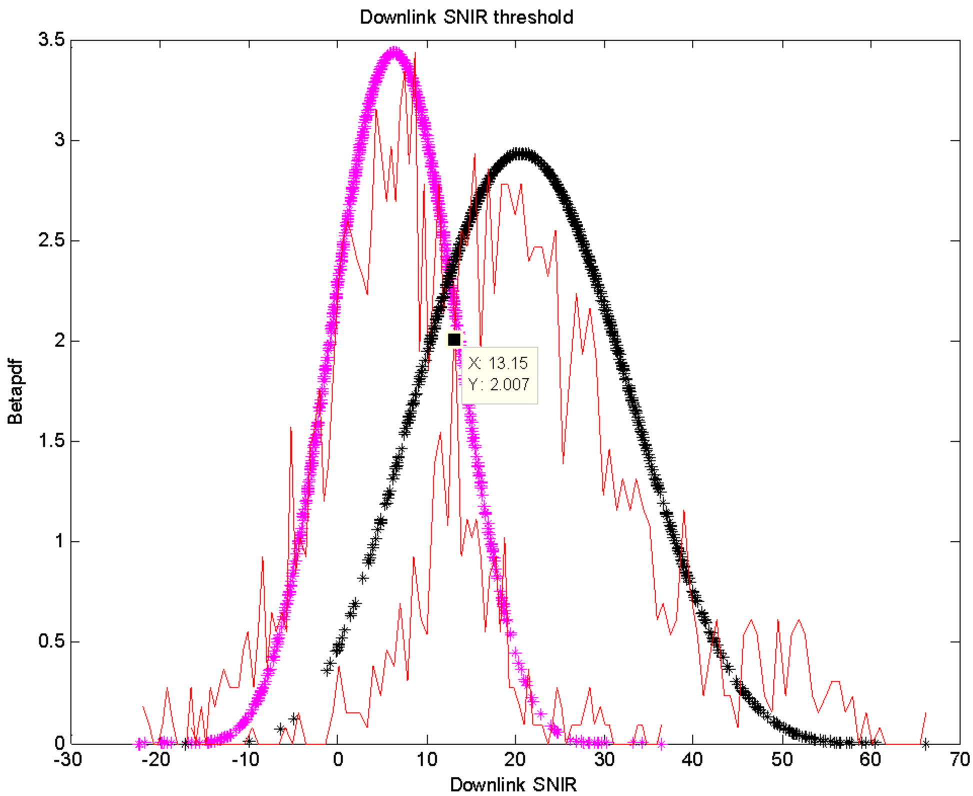


Figure 8.7: CBR Model

doesn't consider lost or spurious symptoms. Whereas the codebook approach can correct up to $k - 1$ errors and detect up to k errors in the symptoms, but its not suitable for simultaneous faults. Whereas the context-free grammar model can handle multiple faults and provides solution by considering the spurious and lost symptoms. The advantage of model based SBM model is it can be updated whenever the network configuration or network conditions change [39].

In Fault diagnosis the advantage of rule based or case based approach is the use of previous experience for present problem. But the problem is maintenance of rules/cases in the database, for example removing old cases, ranking each solution, removing duplicate cases etc.

The choice of the methods in each stage of the self healing framework should consider the pros and cons discussed above. By choosing a suitable method for each stage, the self healing framework provides a end-to-end framework to remove the faults in a network by locating the problem, identifying the symptoms of the problem, finding the root cause and providing the solution for the problem. The database helps in the use of data provided by one module by other module and also can be used as an interface to provide statistics to the operator for manual diagnosis, in case automated diagnosis cannot resolve the problem. Once the problem is solved manually the solution can be updated in the database, to help FD module for resolving similar problem in the future. Hence by using this framework the overall downtime of the network can be reduced, the amount of human intervention can be reduced which improves the overall performance of the network and decrease the OPEX for the operator.

8.10 Conclusions

Self-healing is an important feature for SONs in the NGN as the number of users, services and applications with various QoS requirements, sizes of networks with various technologies and NEs are growing rapidly. We proposed an end-to-end framework consisting of four stages: fault localization, KPI thresholding and monitoring, fault modeling and fault diagnosis. The framework is able to remove faults in a network by locating the problem, identifying the symptoms of the problem, finding the root cause and providing the solution for the problem. In the description of the individual modules, suitable methods from the literature to address issues like reducing volume of alarms, temporal effect of faults, occurrence of simultaneous faults, large volumes of data in huge network, spurious, lost and positive symptoms, etc are given. Hence the overall performance of the network, QoS for users will be increased and OPEX will be decreased due to reduced human intervention and reduced downtime of the system.

The future work includes the implementation of the framework in a generic manner where the proposed methods can be seamlessly integrated and benchmarks performed for any combination of methods across the different blocks.

References

[1] U. Barth, E. Kuehn, " Self-organization in 4G mobile networks: Motivation and vision," 7th International Symposium on Wireless Communication Systems (ISWCS), pp. 731-735, Sep 2010.

[2] NGMN, Next generation mobile networks, http://www.ngmn.org

[3] European Commission, "End-to-End Efficiency (E3)," Integrating Project (IP), FP7-ICT-2007-1.1, Jan 2008-Dec 2009

[4] European commission, "Self-Optimisation and Self Configuration in Wireless Networks (SOKRATES)," FP7 project, Jan 2008-Dec 2010, http://www.fp7-socrates.org

[5] Shanliang Jiang et al, "An Intelligent and Integrated System of Network Fault Management: Artificial Intelligence Technologies and Hybrid Architectures" Shanliang Jiang, D. Siboni, A.A.Rhissa,G. Beuchot" An intelligent and integrated system of network fault management: artificial intelligence technologies and hybrid architectures," Proceedings of IEEE Singapore International Conference on Information Engineering, pp. 265-268, Jul. 1995.

[6] Essam S. Ali, M.G. Darwish, "Diagnosing Network Faults Using Bayesian and Case-based Reasoning Techniques," International conference on Computer Engineering and Systems, ICCES, pp. 145 - 150, Nov. 2007.

[7] Judea Perl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, 1st ed. , Morgan Kaufmann, Sep 1988.

[8] D. Gürer, I. Khan, R. Ogier, R. Keffer, " An Artificial Intelligence Approach to Network Fault Management," Network, vol. 31, no. 415, pp. 1-10, 1998.

[9] L. Lewis "A case-based reasoning approach to the management of faults in communications networks," Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications, pp. 114-120, Mar. 1993.

[10] Irene Katzela, M. Schwartz, "Schemes for Fault Identification in Communication Networks," IEEE transactions on Networking, vol. 3, no. 6, pp. 753-764, Dec 1995.

[11] Shubu Mukherjee, "Architecture design of soft errors" S. Mukherjee, Architeture Design for Soft Errors, 1st ed. , Morgan Kaufmann, Mar 2008.

[12] C.S. Chao, D.L. Yang, A.C. Liu, " An Automated Fault Diagnosis System Using Hierarchical Reasoning and Alarm Correlation," IEEE Workshop on Internet Applications, pp. 120-127, Aug 1999.

[13] Roy sterritt, David W Bustard, "Fusing hard and soft computing for fault management in telecommunication systems," IEEE transactions on Systems, MAN, and Cybernetics, Part C: Applications and Reviews, vol. 32, no. 2, pp. 92-98, May 2002.

[14] M. Steinder, A.S. Sethi, " Probabilistic event-driven fault diagnosis through incremental hypothesis updating," Eighth International Symposium on Integrated Network Management, IFIP/IEEE , pp. 635-648, Mar 2003.

[15] M. Steinder, A.S. Sethi, "A survey of fault localization techniques in computer networks," Science of Computer Programming, vol. 53, no. 2, pp. 165-194, 2004.

[16] A. T. Bouloutas, S. B. Calo, A. Finkel, Irene Katzela, "Distributed Fault Identification in Telecommunication Networks," Journal of Network and Systems Management, vol. 3, no. 3, pp. 295-312, 1995.

[17] I. Katzela, A. T. Bouloutas, S. B. Calo, "Centralized vs Distributed Fault localization," Proceedings of the fourth international symposium on Integrated Network Management, pp. 250-261, 1995.

[18] Gabriel Jakobson, M. Weissman, "Alarm Correlation," IEEE Network, vol. 7, no. 6, pp. 52-59, Nov 1993.

[19] Yuan shun dai, Hinchey Michael, Hu Qingpei, "Consequence Oriented Self.Healing and Hybrid Diagnosis Integrating Decision Diagram, Fuzzy Logic and Neural Network," Fourth IEEE International Workshop on Engineering of Autonomic and Autonomous Systems, pp. 127-136, Mar 2007.

[20] Andreas Hanemann, "A Hybrid Rule-Based/Case-Based Reasoning Approach for Service Fault Diagnosis," 20th International Conference on Advanced Information Networking and Applications, AINA, vol.2, pp. 1-5, Apr 2006.

[21] Andreas Hanemann, M. Salier, "A Framework for Service Quality Assurance using Event Correlation Techniques," Proceedings of the Advanced Industrial conference on Telecommunications/E-learning on Telecommunications workshop, pp. 428-433, Jul 2005.

[22] Keith Harrison, " A novel approach to event correlation," Hewlett Packard, Jul 1994

[23] K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen, H.Toivonen, "Knowledge Discovery from Telecommunication Network Alarm Databases," Proceedings of the Twelfth International Conference on Data Engineering, pp. 115-122, Feb-Mar 1996.

[24] M.V. Ramkumar, Albena Mihovska, Neeli R Prasad, Ramjee Prasad, "Fuzzy-logic based Call Admisison Control for a Heterogeneous Radio Environment," WPMC 2009, Japan, Sep 7 -9

[25] Jiann-liang Chen, Pei-Hwa Huang, "A Fuzzy Expert System for Network Fault Management," IEEE International Conference on Systems, Man, and Cybernetics, vol. 1, pp. 328-331, Oct 1996.

[26] Robert D. Gardner, D. A. Harle, "Alarm Correlation and Network Fault Resolution using the Kohonen Self-Organizing Map," IEEE GLOBECOM, vol. 3, pp. 1398 - 1402, Nov 1997.

[27] A. Goel, J. Ramanujam, P. Sadayappan, "Towards a neural architecture for abductive reasoning," IEEE International Conference on Neural Networks, vol. 1, pp. 681 - 688, Jul 1988.

[28] A. Pras, J. Schonwalder, M. Burgess, O. Festor, G. M. Perez, R. Stadler, B. Stiller,"Key research challenges in Network management," IEEE communications magazine, vol. 45, no. 10, pp. 104-110, Oct 2007.

[29] Sotiris Kotsiantis, Dimitris Kanellopoulos, "Discretization Techniques, A recent survey," GESTS International Transactions on Computer science and Engineering, vol. 32, no. 1, pp. 47-58, 2006.

[30] Ying Yang, Geoffrey I. Webb, Xindong Wu, Discretization methods, Chapter 6 in "Data Mining and Knowledge Discovery Handbook," Springer, pp. 113-130, 2005.

[31] R. Kerber, "ChiMerge: Discretization of Numeric attributes," Proceedings of the tenth national conference on Artificial intelligence AAAI, pp. 123-128, 1992.

[32] Liu, H, R. Setiono, ”Chi2 : Feature selection and discretization of numeric attributes,” 7th International conference on Tools with Artificial intelligence (ICTAI95), pp. 388-391, Nov 1995.

[33] Marc boulle, ”Khiops: A statistical discretization method of continuous attributes,” Machine learning, vol. 55, no. 1, pp. 53-69, Apr 2004.

[34] Usama M, Fayyad, Keki B. Irani, ”Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning,” Proceedings of the International Joint Conference on Uncertainty in AI, vol. 2, no. 1, pp. 1022-1027, 1993.

[35] C.S. Hood, Chuanyi Jit, ”Proactive Network Fault detection,” IEEE Transactions on Reliability, vol. 46, no. 3, pp. 333-341, 2002.

[36] R.Barco, P.Lazaro, L.Diez, V. Wille, ”Continuous versus Discrete Model in Auto diagnosis Systems for Wireless Networks,” IEEE Transactions on Mobile Computing, vol. 7, no. 6, pp. 673-681, June 2008.

[37] M Steinder, A.S. Sethi, ”Increasing robustness of fault localization through analysis of lost, spurious, and positive symptoms,” Proceedings of IEEE 21st Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 1, pp. 322-331, Nov 2002.

[38] S.A. Yemini, S. Kilger, E. Mozes, Y. Yemini, D. Ohsie, ”High speed and Robust event correlation,” IEEE communications magazine, vol. 34, no. 5, pp. 82-90, May 1996.

[39] Bo Yan, Guanling Chen, ”Model-based fault diagnosis for IEEE 802.11 Wireless LANs,” 6th Annual International on Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, pp. 1-10, Jul 2009.

[40] A. T. Bouloutas, S. Calo, A. Finkel,”Alarm Correlation and Fault Identification in Communication Networks,” IEEE transactions on Communications, vol. 42, no. 2/3/4, pp. 523-533, Feb/Mar/Apr 1994.

[41] Malik Jahan Khan, M. M. Awais, S. Shamail, ”Enabling Self-configuration in Autonomic Systems using Case-based reasoning with improved efficiency,” Fourth International conference on Autonomic and Autonomous systems, pp. 112-117, Mar 2008.

[42] R.D. Gardner, D.A. Harle, ”Methods and systems for alarm correlation,” IEEE Globecom, vol.1, pp. 136-140, Nov 1996.

9

Conclusions and future work

This chapter concludes the thesis and propose the future work, which can be build based on the ideas proposed. In brief, this thesis discuss network architectures that suits the needs of next generation networks, propose a management framework and the methods to perform radio resource management to improve the performance of the network and QoS of the user. Then a framework for fault management and methods to be used as a part of the framework are discussed. Throughout the thesis, either the simulation results or discussions or implemented prototypes are presented to validate the proposed ideas.

In this chapter the summary and conclusions of this thesis is presented. Then the future work that can be build upon this thesis is proposed. This work has identified some of the important problems in network management for the case of cellular networks. For the identified problems, the existing methods in the literature are investigated and the new methods are proposed that can give better results and performance or can give a different outlook.

In the first chapter, we considered the network architectures based on RoF or integrated radio-fiber for the next generation networks. The basic idea to choose RoF or integrated radio-fiber architectures is to integrate different BSs, different technologies using a common infrastructure. This architecture also reduces the deployment costs for an operator and also help for efficient implementation of network management algorithms. The availability of data at the CU, helps to develop better network management algorithms and helps to take better decisions from overall network point of view.

As a part of this, we consider OCDMA as better choice compared to other technologies for integrated optical and fiber network, due to its asynchronous operation, inherent security etc. Due to the soft blocking feature of OCDMA, the networks can be scalable and also survivable in case of network failures. Also OCDMA allows transmission of traffic at variable bitrates and variable BER by changing the length of the codes. The RoF manager provides performance, fault, security and configuration management of RoF networks.

Secondly, when a mobile user want to initiate a call with a BS the user has to be admitted or rejected by the system. In this thesis, a fuzzy logic based AC is proposed, which can be used for any next generation heterogeneous wireless network with multiple applications. The proposed method consider the application requirements, user conditions, network conditions and user preferences and admits or rejects a new or handoff user, while assigning best cell and best RAT. The disadvantage of using fuzzy base method is the framing of inference rules, which normally depends on past experience. Hence the accuracy of fuzzy based AC depends on the accuracy of framing the inference rules. Hence a cross layer AC based on markov modeling of the queue and AMC in the physical layer is proposed. With this method, the number of OFDMA slots required by a new user is estimated, hence the QoS of a new user and existing users is guaranteed. The future work is to extend the current framework for heterogeneous wireless networks, where the common AC selects the most suitable cell and RAT for the user that can cater the best QoS to the user while maximizing the overall capacity of the heterogeneous network. This can be done by combining the fuzzy based cell and RAT selection with Markov based estimation.

The proposed AC algorithms, reside in LRRM/CRRM of middleware. As a part of FUTON architecture, a middleware is presented which ensures seamless mobility and service continuity through IPv4/IPv6 Mobile IP (MIP) core. The main module of middleware is CRRM, LS, MIH etc. In this thesis, the implementation details of CRRM and LRRM are presented. The results obtained from CRRM for vertical handover scenario is presented for the case of WLAN and UMTS. This implementation of CRRM can be used to further extend and add more functionalities like AC, congestion control etc. The middleware presented is generic and can be used in the next generation networks.

Thirdly, the coverage and capacity optimization of an LTE network as a part of self optimization is considered. As a part of this, a AAA framework is proposed and the methods for dynamic resource allocation, AC and assignment are proposed. For dynamic resource allocation, a heuristic approach is chosen which predicts the SINR of each user in the next super frame and allocates resources to the users based on the predicted SINR. As a part of admission control, a mean resource based approach is proposed, which estimated mean number of resources used by all the users in the system based on their buffer conditions. The results of resource allocation are compared with state of art and observed a 10% increase in the overall throughput and cell edge throughput. Also this framework guarantees the QoS of new users and increase the total number of users in the system by considering multi user diversity. This framework can be extended to any next generation wireless network. The future work can be to extend this framework for heterogeneous network scenario from common radio resource management perspective, by including RAT selection and vertical handoff.

Lastly, for self healing a 4-stage approach is proposed, that provides end-to-end framework to remove the faults in a network by locating the problem, identifying the symptoms of the problem, finding the root cause and providing the solution for the problem. The common database helps in use of data provided by one module by other module and also can be used as an interface to provide statistics to the operator for manual diagnosis, in case automated diagnosis cannot resolve the problem. Hence by appropriate choice of methods at each stage of the framework the overall downtime of the network can be reduced, the amount of human intervention can be reduced which improves the overall performance of the network and decrease the OPEX for the operator.

Hence this thesis proposed various network management techniques by considering different user conditions like QoS, traffic conditions, buffer conditions, environmental and geographical conditions of the user and various network conditions like load, throughput, dropping/blocking probability, total number of users in the system, OPEX/CAPEX costs, etc.

9.1 Future work

The future work is to implement the an LTE or any other next generation wireless network with a fiber optic backhaul based on OCDMA. There could be many interesting challenges when the wireless system is integrated with a fiber optic backhaul. The future work could be to identify the challenges like joint resource allocation of the integrated network, impact on QoS due to the delay caused in the fiber optic backhaul, improvement of the security using OCDMA codes, effect of MAI on the overall performance of the network etc.

As a part of AAA framework, it is interesting to see the performance of the framework with different allocation methods and other admission control methods proposed in the thesis. Also the AAA framework could be implemented for a heterogeneous network scenario. In case of integrated optical and radio architectures the AAA framework could be extended to joint AAA framework, which performs joint allocation, assignment and admission control to the users in both wireless and optical domains.

The CRRM implementation discussed in the thesis can be extended to implement the proposed methods and validate the performance in a real network or an emulated network. As a part of FUTON project, the CRRM is validated on a RoF network, which could be extended to a RoF network based on OCDMA.

In case of self healing, the future work would be to implement the 4-stage self healing model and measure the accuracy of diagnosis with various methods in different scenarios. Apart from that, to combine the self-healing with self-optimization and propose a framework and design for self-organizing networks is a real challenge.