University of Coimbra
Faculty of Sciences and Technology
Department of Informatics Engineering

Cooperative Routing Management in Wireless Mesh Networks

Vinicius da Cunha Martins Borges

Coimbra
2012

University of Coimbra
Faculty of Sciences and Technology
Department of Informatics Engineering

Cooperative Routing Management in Wireless Mesh Networks

Vinicius da Cunha Martins Borges

Thesis submitted as partial fulfillment of the requirements of the Doctoral Program on Information Science and Technology in the University of Coimbra, for the award of Doctor in the domain of Telematics Networking and Communications under care of advisor Prof. Dr. Edmundo Heitor da Silva Monteiro and co-advisor Prof. Dra. Marilia Pascoal Curado.

Coimbra
2012

Not everything that can be counted counts and not everything that counts can be counted.
Albert Einstein

With love to my grandmother, Cessão. I will never forget what you did for us.

Acknowledgment

First and foremost, I would like to thank God for giving me the opportunity to carry out this research and the patience and strength to complete it. Secondly, I would like to express my sincere gratitude to my supervisor Dra. Marilia Curado for leading me into this exciting research area and giving me advice whenever I needed it. I am grateful for her constant encouragement, support, supervision and guidance during my PhD at the University of Coimbra. Her intellectual rigour, wisdom and kindness are attributes that will always represent a model for me to follow. I would also like to thank my second supervisor Dr. Edmundo Monteiro for his asistance, valuable suggestions and comments he made throughout the period of my PhD.

I am very grateful to have had the opportunity to work with so many wonderful people at the Department of Informatics Engineering, who deserve credit for providing such a congenial working environment. Special thanks too, to the members of the Laboratory of Communications and Telematics, Eduardo Cerqueira, David Rodrigues, Adriano Fiorese, Fernando Matos, David Palma, Vitor Bernado, Luiz Henrique, Saulo Queiroz and Roger Immich, and in particular, the co-authors of the publications included in this thesis, Daniel Pereira and Erik Dimitrov, whose assistance was of great value. I would also like to thank the other members of the university staff, Jorge Ávila, Dona Manoela, Lucinda, Diana, Márcia, Ricardo Ruivo, Carlos and many others during my study for all their help. It has been a privilege to work with all of you.

I would like to thank my friends at the Department of Informatics Engineering for their warmth and friendship, in particular, Dirson Campos, Alexandre Malhão, Ivano Irrera, Merkebe Getachew, Dien Than, Dinesh Kumar, Ricardo Mendão and André Gomes.

This work was sponsored by the Portuguese Ministry of Science (scholarship contract SFRH/BD/44378/2008), and partially supported by the FCT UBIQUIMESH project (PTDC/EEATEL/105472/2008) and the FCT User Centric Routing (UCR) project (PTDC/EEA-TEL/103637/2008).

Finally, my special thanks to my parents and sisters for their constant support and understanding.

Resumo

Redes em malha sem-fios (WMN) fornecem um *backbone* sem fios flexível para acesso ubíquo à Internet, e são desafiadas a melhorar a sua gestão para suportar vários tipos de requisitos, tais como escalabilidade de aplicações multimédia e integração com diferentes tecnologias sem fios. A estrutura *multi-hop* e soluções de baixo custo utilizadas nas WMN facilitam a extensão dos seus serviços para a cobertura de grandes áreas. Por esta razão, a escalabilidade é uma questão de gestão fundamental para WMN, sendo necessário que as WMN consigam lidar com quantidades crescentes de tráfego e de nós de uma forma eficiente. Neste cenário, o processo de encaminhamento pode servir como um mecanismo útil para a gestão deste tipo de redes e atender as exigências de aplicações multimédia de uma forma mais escalável. Este processo emprega métodos com soluções distribuídas, tais como algoritmos, protocolos de encaminhamento e métricas que em conjunto permitem a seleção das melhores rotas permitindo uma otimização de desempenho. No entanto, vários fatores devem ser levados em consideração pelas abordagens de encaminhamento para melhorar a escalabilidade nas WMN, tais como informações imprecisas de encaminhamento, altos níveis de *overhead* dos protocolos de encaminhamento em redes de larga escala e as áreas de congestionamento próximas dos *gateways*.

O argumento central desta tese é que embora tenha havido várias propostas de encaminhamento para melhorar o desempenho das WMN, as soluções atuais não conseguiram adotar uma abordagem que seja capaz de lidar com os três principais aspectos do processo de encaminhamento numa mesma abordagem, nomeadamente a imprecisão das métricas para medir a qualidade do link sem fio, o *overhead* dos protocolos de encaminhamento e a ocorrência de *gateways* sobrecarregados. Na verdade, todo o esforço de investigação anterior foi centrado num único aspecto. O objectivo do trabalho de investigação apresentado nesta tese foi demonstrar que é possível criar uma abordagem de encaminhamento que permite melhorar a escalabilidade das WMN de uma forma eficaz. Para atingir este objetivo, este trabalho propõe uma arquitetura, chamada *Architecture of Routing Management* (ACRoMa), que é apresentada através de uma abordagem *top-down* em que os principais componentes e sinergias são descritos de uma forma detalhada. ACRoMa foi projetada para fornecer um algoritmo de encaminhamento para balanceamento de carga *inter-cluster*, chamado *Routing Algorithm for Inter-cluster Load Balancing* (RAILoB), que reduz o *overhead* de encaminhamento e evita as situações de sobrecarga nos *gateways* e uma métrica *cross-layer* de encaminhamento, chamada *Metric for Interference and channel Diversity* (MIND), para melhorar a precisão da decisão de

encaminhamento através do uso de medidas mais exatas para descrever interferência e carga de tráfego. RAILoB acelera o processo de balanceamento de carga entre gateways (*inter-cluster*). Além disso, existe uma interação entre RAILoB e a métrica MIND que permite realizar uma decisão de encaminhamento *intra-cluster*. Portanto, RAILoB representa a arquitetura ACRoMa conceptualmente através da combinação de todos os componentes em uma forma incremental.

Modelos de simulação foram desenvolvidos para validar ACRoMa através de uma avaliação extensa, e estes levam em conta os principais fatores que influenciam o desempenho do tráfego (por exemplo, topologia, aplicações e tamanho da rede). Em primeiro lugar, a avaliação de MIND mostra que ela supera várias métricas de encaminhamento *cross-layer* em configurações diferentes, o que é uma evidência de que os mecanismos mais precisos empregados em MIND têm impacto sobre a decisão de encaminhamento. Na sequência, houve uma avaliação em diferentes cenários e aplicações da abordagem RAILoB. Os resultados mostraram que RAILoB consegue um desempenho de tráfego melhor do que as abordagens mais relevantes de balanceamento de carga usando também *clustering*, uma vez que fornece uma solução mais ágil para a balanceamento de carga *inter-cluster*. Assim, a arquitetura ACRoMa alcançou seus objetivos iniciais, mostrando que é possível melhorar a escalabilidade das WMN sem a necessidade de acrescentar novos equipamentos ou tecnologias de redes sem fios, combinando soluções que cooperam entre si na mesma abordagem.

Abstract

Wireless Mesh Networks (WMN) provide a flexible wireless backbone for ubiquitous Internet access, and are being challenged to improve their management to support various kinds of requirements, such as scalable multimedia applications and integration with different wireless technologies. The multi-hop structure and low-cost solutions used by WMN make it easier to extend their services to cover larger areas. For this reason, scalability is a critical management issue for WMN and therefore, it is required that WMN are enabled to handle growing amounts of traffic load and nodes in a skilful manner. In this scenario, the routing process can serve as one of the most useful mechanisms for managing this kind of network and meeting the requirements of multimedia applications in a more scalable way. This process employs methods for distributed solutions, such as routing algorithms, protocols and metrics that work together to select the best routes to enable a performance optimization. However, several factors should be taken into consideration by the routing approaches adopted to improve the WMN scalability, such as inaccurate routing information, high routing overhead in large networks and the congestion areas around the gateways.

The central argument of this thesis is that although there have been several routing proposals to improve the WMN performance, the current solutions have failed to adopt an approach that is able to handle the three main aspects of the routing process, i.e. inaccuracy in routing information, overloaded gateways and high overhead. In fact, all the development has been centered on a single aspect that does not combine different solutions that tackle each aspect of the routing process. The aim of this study was to demonstrate that it is possible to create a routing approach that allows the WMN scalability to be leveraged in an effective way. In order to achieve this goal, this work employed an architecture, called Architecture of Routing Management (ACRoMa), that is presented using a top-down approach in which the main components and synergies are outlined through a detailed description. ACRoMa has been designed to provide a routing algorithm for inter-cluster load balancing, called Routing Algorithm for Inter-cluster Load Balancing (RAILoB), which reduces the routing overhead and avoids overload situations in gateways and a cross-layer routing metric, called Metric for INterference and channel Diversity (MIND), to improve the accuracy of the routing decision through the use of precise measures to depict interference and traffic load. RAILoB speeds up the process of load balancing between gateways (inter-cluster). Moreover, there is an interaction between MIND and RAILoB that enables to perform intra-cluster routing decisions.

Hence, RAILoB represents the ACRoMa architecture conceptually by combining all the components in an incremental way.

Simulation models have been carried out to validate the soundness of ACRoMa through an in-depth evaluation, and these take into account the main factors that influence the traffic performance (e.g. topology, applications and network size). First of all, the evaluation of MIND shows that it outperforms several cross-layer routing metrics in different configuration matrices, which is evidence that the most accurate mechanisms employed in MIND have a beneficial influence on the routing decision. Following this, there was an assessment of RAILoB in different scenarios and applications. The results showed that RAILoB achieves higher traffic performance than the most relevant clustering load balancing routing approach in WMN, since it provides a more flexible and agile solution for inter-cluster routing load balancing. Thus, the ACRoMa architecture fulfilled its original goals, by showing that it is possible to enhance the WMN scalability by combining solutions in the same approach which cooperate each other.

Foreword

The work described in this thesis was conducted at the Laboratory of Communication and Telematics of the Centre for Informatics and Systems of the University of Coimbra within the context of the following projects:

- **Project UBIQUIMESH** - Project funded by the Science and Technology Foundation (FCT), from September 2009 to July 2012. The design and validation of the main mechanisms proposed in this thesis was performed in the context of this project.

- **Project User Centric Routing (UCR)** - Project funded by the Science and Technology Foundation (FCT), from September 2009 to August 2012. This project allowed for the dissemination and discussion of the concepts used in the proposed architecture presented in this thesis.

The work done during this thesis resulted in the following publications:

Borges, V. C. M., Vieira, D., Curado, M., and Monteiro, E.. Routing Metric for Interference and Channel Diversity in Multi Radio Wireless Mesh Network. 8th ADHOCNOW 09, Springer LNCS, 5793, 2009, pp 55-68.

Borges, V. C. M., Curado, M., and Monteiro, E.. A Cross-Layer Routing Scheme for Scalable Triple Play Service in Wireless Mesh Networks. In Proceedings of the 19th IEEE ICCCN 2010, Zurich, Switzerland, August 2-5, 2010, pp 1-6.

Borges, V. C., Curado, M., and Monteiro, E.. Cross-Layer Routing Metrics for Mesh Networks Current Status and Research Directions. Computer Communications, 34(6), 2011, pages 681-703.

Borges, V. C., Curado, M., and Monteiro, E.. The Impact of Interference-aware Routing Metrics on Video Streaming in Wireless Mesh Networks. Ad Hoc Networks, 9(4), 2011, pages 652-661.

Borges, V. C. M., Dimitrov, E., Curado, M., and Monteiro, E.. Performance Assessment of Cluster Load Balancing Routing Methods for Triple Play Services in Wireless Mesh Networks. In Proceedings of the 13th IEEE/IFIP NOMS 2012, pages

974-980.

Borges, V. C. M., Dimitrov, E., Curado, M., and Monteiro, E.. RAILoB - A Routing Algorithm for Inter-cluster Load Balancing in Wireless Mesh Networks. In Proceedings of the 9th IEEE CCNC 2012, pages 904-909.

Borges, V. C. M., Curado, M., and Monteiro, E.. ACRoMa Architecture for Cooperative Routing Management in Wireless Mesh Networks. Submitted Paper to the Springer Wireless Networks, 2012.

Contents

List of Figures

List of Tables

List of Acronyms

ACKs ACKnowledgements

ACRoMa Architecture of Cooperative Routing Management

AIL Average Interfering Load

AODV Ad hoc On-Demand Distance Vector routing

AODV-ST AODV-Spanning Tree

BATMAN Better Approach to Mobile Ad-hoc Networking

BER Bit Error Rate

BFS BreadthFirst Search

CAN Configurable Access Network

CARMA Clustering Approach for Routing MAnagement

CARMEN CARrier grade MEsh Networks

CBT Channel Busy Time

CBT Channel Busy Time

CH Channel

CL Cumulative Load

CoCLuS Collaborative CLustering Scheme

COTS Common-Off-The-Shelf

CSC Channel Switching Cost

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance

CS-threshold Carrier-Sensing Thresholds

CWB Contention Window Based

DCA Distributed Clustering Algorithm

DCF Distribute Coordination Function

DCSPT	Dual-Carrier Sense with Parallel Transmission-awareness
DHW	Dynamic History Window
DIFS	InterFrame Space
DS	Dominating Set
DSDV	Destination-Sequenced Distance Vector
DSP	Digital Signal Processing
DSR	Dynamic Source Routing
DYMO	DYnamic MANET On Demand
EETT	Exclusive Expected Transmission Time
EFT	Expected Forwarding Time
EIFS	Extended InterFrame Space
ELP	Expected Link Performance
ENT	Effective Number of Transmissions
ERM	Efficient Route Maintenance
ETT	Expected Transmission Time
ETX	Expected Transmission Count
EWMA	Exponential Weighting Moving Average
FDD	Frequency Division Duplexing
FER	Frame Error Rate
FHW	Fixed History Window
FPS	Frames per Second
FSO	Free-Space Optical
GCMR	Gateway-Centralized Multi-hop Routing
GDTSP	Greedy Dominating Tree Set Partitioning
GHz	Gigahertz
GPO	Gateway Placement Optimization
GSM/EDGE	Global System for Mobile Communications/Enhanced Data rates for GSM Evolution

HWMP	Hybrid Wireless Mesh Protocol
IAR	Interference-Aware Routing
iAWARE	interference AWARE
ID	Imbalance Difference
IDAR	Interference and Delay Aware Routing
IDLB	Inner Domain Load Balancing
iETT	Improve Expected Transmission Time
IGDS	Iterative Greedy Dominating Set
IL	Interfering Load
ILA	Interference-Load Aware
ILP	Integer Linear Problem
INSTC	INterference Survivable Topology Control
INTER-LOAD	INTER-flow interference and LOAD awareness
INX	Interferer Neighbors Count
IP	Internet Protocol
IR	Interference Ratio
IRMA	Integrated Routing and MAC scheduling
IRU	Interference-aware Resource Usage
LAC	Load-Adaptive Clustering phase
LAETT	Load Aware ETT
LBA	Load Balancing Approach
LBD	Link Breakage Decision
LDQ	Load-Distributive QoS
LQA	Link Quality Assessment
LTE	Long Term Evolution
LUNAR	Lightweight Underlay Network Ad-Hoc Routing
MAC	Media Access Control

MANET	Mobile Ad hoc NETworks
MCLSR	Multi-Channel Link-State Routing
MCMR	Multi-Channel Multi-Radio
MCR	Multi-Channel Routing
MD	Minimum Delay
mETX	Modified Expected Number of Transmissions
MIC	Metric of Interference and Channel-switching
MIMO	Multiple-Input and Multiple-Output
MIND	Metric for INterference and channel Diversity
ML	Migrated Load
MN	Migrated Nodes
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MPR	Multi-Point Relays
MTI	Metric of Traffic Interference
NLOS	Non-Line Of Sight
NOC	Network Operation Center
ODLB	Outer Domain Load Balancing phase
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLSR	Optimized Link State Routing
OSC	Optimal Static Clustering
OSI	Open Systems Interconnection
P2P	Peer-to-Peer
PC	Personal Computers
PDA	Personal Digital Assistants
PHY	PHYsical

PLB	Partition-based Load Balancing
POS	Probability Of Success
PPRCH	Clear-To-Send (CTS) Handshake
QL	Queue Length
QoE	Quality of Experience
QoS	Quality of Service
QUORUM	QUality Of service RoUting
RAILoB	Routing Algorithm for Inter-cluster Load Balancing
RANN	Root ANNouncement message
RARE	Resource Aware Routing for mEsh
RC	Remaining Capacity
RER	ROUTE ERROR
RFMon	Radio Frequency Monitor
RREP	ROUTE REPLY
RREQ	ROUTE REQUEST
RRU	Radio Resource Utilization
RSSI	Received Signal Strength Indication
RTS	Packet Probing with Request-To-Send
RTT	Round Trip Time
SIFS	Short InterFrame Space
SINR	Interference-plus-Noise Ratio
SSIM	Structural SIMilarity Index
TCP	Transmission Control Protocol
TDD	Time-Division Duplexing
TDMA	Time Division Multiple Access
TSL	Top Sub-Links
TSN	Top Sub-Nodes

TTL	Time To Live
UDP	User Datagram Protocol
UMTS/HSPA	Universal Mobile Telecommunications System/High Speed Packet Access
UPT	Update Propagation Threshold
VBR	Variable Bit Rate
VMC	Virtual Micro-Clusters
WCETT	Weighted Cumulative ETT
WCETT-LB	Weighted Cumulative ETT-Load Balancing
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Inter-operability for Microwave Access
WMN	Wireless Mesh Networks
WSN	Wireless Sensor Networks
ZRP	Zone Routing Protocol

Chapter 1

Introduction

This Thesis deals with the problematic area of the routing process in Wireless Mesh Networks (WMN). In this chapter, the motivating factors behind the research study are discussed in Section 1.1. The objectives of the work, the methodology employed and the contribution it can make are described in Section 1.2. Section 1.3 outlines the way it is structured.

1.1 Motivation and Problem Statement

WMN provide a low cost investment environment for next generation networks based on a multi-hop wireless backbone which extends the Internet access in an ubiquitous way. These networks have been emerging as a new communications paradigm that is not subject to the traditional restrictions of ad hoc networks (e.g. energy and processing capacity) [Akyldiz et al., 2005]. Moreover, WMN aims to integrate diverse kinds of wireless networks. Despite this, these networks have generally employed the IEEE 802.11a/b/g standards as a wireless technology because this is a low cost solution, which results in a wireless link with restricted capacity (e.g. with a limited number of non-overlapping channels).

In this environment, the support of triple play services [Ekling et al., 2007], i.e. voice, video and data applications are all provided in a single access subscription (service providers), which requires suitable Quality of Service (QoS) levels. It is a challenging task in this environment, since it is difficult to manage the scarce bandwidth to support the service assurance of these kinds of services. In addition, the wireless backbone comprises gateways and mesh routers, where a set of mesh routers offers the multi-hop backbone to reach the gateways which have a wired connection with the Internet. Gateways are potential bottlenecks in this scenario, since most of the traffic load in WMN travels to or from them. In this context, WMN configure a loosely cooperative service network, where gateways and mesh routers must collaborate with each other to take advantage of the overall network capacity.

The multi-hop structure and low-cost solutions taken by WMN, make it easier for them to expand and cover larger areas. For this reason, scalability is a critical management issue of WMN since it seeks to handle growing amounts of traffic load and nodes in a dexterous manner. In view of this, a number of network technologies has been proposed for the PHY (PHYsical) and Media Access Control (MAC) layers to increase the performance of the wireless networks, such as Multiple-Input and Multiple-Output (MIMO) [Chu and Wang, 2010] and cognitive radio [Akyildiz et al., 2009]. Although these technologies improve the network performance significantly by increasing the capacity of the wireless link, it is not yet possible to gauge the potential value of these technologies, since they entail an increase in the cost and signalling overhead, as well the complexity of their deployment, which makes it difficult to manage the WMN. For these reasons, these solutions need more work before they can be consolidated in the industrial and academic environment. In light of this, the routing process has become one of the most useful mechanisms to complement these network technologies in order to support the requirements of multimedia applications in a more scalable way. This process employs distributed solutions, such as routing algorithms, protocols and metrics that can compute the best routes and enable a more complete performance optimization of the wireless medium without additional cost or deployment.

However, routing protocols, such as Optimized Link State Routing (OLSR) [Jacquet et al., 2001] and Hybrid Wireless Mesh Protocol (HWMP) [Bahr, 2007] have serious limitations with regard to scalability. In particular, they cause a large overhead in proactive routing strategy and a considerable delay in reactive routing strategy for large wireless networks [Baumann et al., 2008]. For example, setting up a routing path in a very large wireless network may take a long time, and the end-to-end delay can make it even longer. This identified problem can be tackled by solutions of clustering which have been employed in WMN to improve the management of the routing decision-making process, since they increase the scalability of the current routing protocols in large wireless networks by reducing the routing overhead [Yu and Chong, 2005][Ros and Ruiz, 2007].

By employing these schemes, the WMN can be divided into different virtual groups, where the nodes are allocated geographically so that they are adjacent to the same cluster and conform to specific rules. A cluster usually consists of a gateway (i.e. clusterhead) and a set of mesh routers in WMN. The use of clustering has some advantages, since it allows a smaller and more stable structure to be produced. In this scenario, if a mesh router fails, only the routers that are in the corresponding cluster need to update their information. As a result, local changes do not spread out and have to be updated across the entire network, which reduces the amount of information processed and stored by each node. Thus, clustering stands out as an efficient means of reducing the processing and propagation of routing information.

Despite the fact that clustering improves the performance of routing protocols in WMN, clustering is not sufficient to achieve a truly scalable solution when the traffic load increases in the network. This means that intelligent routing decisions that aim at load balancing at both intra-cluster and inter-cluster levels, play an important role in WMN, since gateways towards wired networks are potential bottlenecks. Intra-cluster load balancing schemes [Hsiao et al., 2001][Dai and Han, 2003] handle the load balancing issue inside a single cluster (i.e. they provide a local perspective), by distributing the traffic load among the routing sub-trees in which the gateway is the root.

Although intra-cluster routing load balancing can improve the traffic performance locally, it fails to distribute the traffic load uniformly throughout the whole network, since the intra-cluster load balancing is restricted by the capacity of the gateway. The inter-cluster load-balancing deals with this issue by reducing the cluster congestion in a holistic perspective, and directing the mesh router traffic towards lightly-loaded gateways. Hence, inter-cluster load balancing routing between multiple gateways is a necessary mechanism to manage the traffic load in WMN [Bejerano et al., 2007] in a scalable way. It thus improves the overall capacity of the network by avoiding congested gateways.

The wireless medium is shared by several nodes in WMN. A wireless link in WMN does not have dedicated bandwidth and consequently, neighboring node transmissions may also compete for the same bandwidth, so that a transmission in one wireless link interferes with transmissions in neighboring links. Past research studies into the routing process have usually recommended reducing the interference through channel assignment in multi-channel multi-radio WMN [Marina et al., 2010], time slot scheduling [Wu et al., 2006] and MIMO [Chu and Wang, 2010]. However, the restricted number of available channels in the physical specification does not allow one channel to be assigned to each wireless link in the WMN (e.g., the simultaneous operation of three non-overlapping channels in the 2.4 gigahertz (GHz) band and 12 non-overlapping channels in the 5 GHz band). As a result, the channels have to be assigned in a repetitive way between the links, which still causes interference in some of them.

There are similar constraints in the time slot scheduling and MIMO. Hence, these solutions do not scale well, due to the very restricted resources of the wireless medium and the fact that most decisions are centralized. A further point is that the routing process must also be aware of the link quality to enable it to pick up the interference between the neighboring nodes and thus improve the traffic performance.

The neighboring links with high traffic load should also be depicted by the routing metric, since the traffic load also causes interference (i.e. self-interference) and increases the congestion in the wireless links. As a result of this, there is a decline in the traffic performance. Hence, paths with a high level of interference and traffic load can be avoided by using the local information to improve the traffic performance. In view of this, routing metrics play a key role in picking up interference levels and traffic load by using local information to make a routing decision in a distributed way, while avoiding excessive overhead caused by the measurement and dissemination of this information. In addition, the cross-layer design has been employed in WMN [Akyildiz and Wang, 2008] to gather this kind of information in a precise way, since the information about interference and traffic load is not available in the routing layer of the traditional Open Systems Interconnection (OSI) network model [ITU, 1994]. Furthermore, routing protocols usually employ routing metrics that do not display the link quality by combining information from different layers. Through this approach, the routing scheme becomes aware of the link capacity, number of interfaces, interference and traffic load in neighboring nodes, and is then able to select the paths that will provide an enhanced traffic performance.

The routing approaches which have been proposed in WMN combine a routing process with MAC scheduling [Wu et al., 2006], spectrum management [Alsarhan and Agarwal, 2009] and high availability [Ashraf et al., 2011]. However, there has been not very much discussion about ways to improve the scalability of WMN through the routing process. To fill this gap, consideration has been given to integrating a clustering scheme, a load balancing routing and a cross-layer routing metric in this thesis since these are the most significant means of improving the traffic performance for WMN in a scalable way. This integration improves the overall network performance through the routing process by achieving a greater degree of traffic performance, and hence, enabling paths to be selected that can satisfy the requirements of application demands such as VoIP and video.

1.2 Objectives and Contributions

The main goal of this thesis is to propose an architecture for cooperative routing management that is able to improve the scalability of WMN, which is the main limitation of the existing routing approaches. This architecture integrates the most significant components, to manage the routing process in a way that allows a higher degree of traffic performance to be achieved. In other words, it combines a clustering scheme, load balancing routing algorithms, and a cross-layer routing metric. The specific goals of this thesis are as follows:

- To enable the best paths to be selected by depicting accurate measurements of the link quality through a cross-layer routing metric.
- To reduce the routing overhead, which is usually caused in the flat routing structure of the traditional routing protocols, by using a clustering routing approach.
- The means of avoiding the overload situation is based on an inter-cluster load balancing routing algorithm.

The relevant contributions of this research study are summarized in the following sub-sections.

Architecture for Cooperative Routing Management

The problematic area of the traffic scalability provision through the routing process requires the use of some solutions. In view of this, an architectural model, called Architecture of Cooperative Routing Management (ACRoMa), is proposed to manage the routing process which is designed to employ a top-down approach. Hence, ACRoMa integrates three components to improve the traffic scalability in WMN, which are a clustering approach, load balancing routing algorithms, and a cross-layer routing metric. Furthermore, these components contain synergies that help to ensure that the main goal is achieved [Borges et al., 2012a].

Taxonomy of Cross-layer Routing Metric

Several cross-layer routing metrics have been proposed in WMN. Due to the wide range of metrics, as well as the complex mechanisms and measurements used, it is difficult to identify the open issues of this research field. For this reason, a taxonomy was proposed to classify and characterize the most relevant and recent cross-layer routing metrics. This taxonomy offers a consistent view of the way that this research field has evolved and comprises the following categories: information-gathering methods, measurements and stability mechanisms [Borges et al., 2011a].

Cross-layer Routing Metric

The next contribution is the conception of a cross-layer routing metric, called Metric for INterference and channel Diversity (MIND). The development of the metric draws on the previous characterization of open issues in the cross-layer routing metric. As a result of this, this metric obtains precise measurements concerning the main factors that influence the link quality. In other words, the proposed metric combines accurate measurements of interference and traffic load. In addition, the purpose of the metric is to take into account the different characteristics of the applications such as those of triple play services which use three distinct data rates and transport protocols [Borges et al., 2009].

Performance Assessment of Cross-layer Routing Metric

The performance evaluation of the proposed metric was performed by simulation and varying configurations that can influence the traffic performance when the cross-layer routing metrics is used, such as traffic load (i.e. number of flows), transmission and interference ranges. This evaluation takes into account the main performance parameters which are as follows: Quality of Service (QoS) and Quality of Experience (QoE) parameters, routing overhead and routing stability. Moreover, the cross-layer routing metric can be evaluated with different applications (i.e. data, video, voice and a combination of all of them), as well as in outdoor and indoor environments. Furthermore, the different routing schemes (link state routing with the proposed cross-layer routing metric and load balancing routing) are compared so that an analysis can be conducted of their impact on the traffic performance [Borges et al., 2011a, Borges et al., 2011b].

Clustering Approach for Traffic Migration

A clustering approach, called as Clustering Approach for Routing MAnagement (CARMA), has been proposed to provide an enhanced and scalable solution for WMN, which supports the traffic distribution between the gateways for load balancing purpose. First of all, an analysis of the existing clustering approaches in WMN was carried out in order to identify the open issues and limitations of these approaches. After that, CARMA was specified so that these limitations could be overcome. CARMA comprises a traffic migration method (i.e. mesh traffic migration), a clustering routing scheme, called as Collaborative CLustering Scheme (CoCLuS), and an inter-cluster load balancing routing algorithm, named Routing Algorithm for Inter-cluster Load Balancing (RAILoB) [Borges et al., 2012c]. Furthermore, RAILoB and MIND have a synergy in order to make intra-cluster routing decision.

Performance Assessment of the Clustering Approach

The performance evaluation of the proposal was also conducted by simulation and assessing varying configurations that influence the traffic performance in a clustering approach, such as traffic load, number of nodes, number of gateways and topologies. It takes into account the main performance parameters which are as follows: QoS and QoE parameters, routing overhead and number of migration events. Moreover, the clustering approach is also assessed with different applications [Borges et al., 2012c, Borges et al., 2012b, Borges et al., 2012a].

1.3 Outline of the Thesis

The remainder of the thesis is divided in six chapters structured as follows.

Chapter 2 provides background information about WMN such as elements, features, architecture and wireless technologies. In addition, the main concepts of routing protocols and strategies employed in WMN are also included in this chapter.

Chapter 3 conducts a survey of the most relevant and recent routing approaches as a means of highlighting the open issues. As a result of this, the inherent difficulty of employing these approaches to provide a scalable solution for WMN is identified as their main drawback. To overcome this limitation, the architectural model is proposed, together with a general description of each of its components.

Chapter 4 includes a taxonomy and survey of the most recent and relevant routing metrics as well as a proposed cross-layer routing metric. Furthermore, a performance evaluation of some cross-layer routing metrics (including the proposed metric) is carried out with distinct applications, environments and performance parameters (i.e., QoS, QoE, overhead and stability).

Chapter 5 contains a taxonomy and survey of the most recent and relevant clustering approaches. In addition, the proposed clustering approach is described together with a performance evaluation based on distinct applications, topologies, network configurations (i.e. number of nodes and gateways) and performance parameters (i.e., QoS, QoE, overhead and number of migration events).

The conclusions that emerged from the research work described in this thesis, are outlined in Chapter 6. The experience gained from this work is then used to define the open issues that need to be addressed in future work.

Chapter 2
Wireless Mesh Networks

Wireless Mesh Networks (WMN) are a potentially valuable type of wireless network that have the capacity to provide ubiquitous Internet access as well as broadband wireless coverage, to large areas with minimal up-front investment and infrastructure requirements. WMN are self-organized and self-configured in a dynamic way, and able to undertake the addition and removal of nodes as the situation requires. Thus, they can provide very good flexibility and scalability. There are several application scenarios where WMN are employed, such as broadband home networking, metropolitan area networks, transportation systems, security surveillance systems, and the deployment of rural community networks [Akyldiz et al., 2005, Campista et al., 2008, Ishmael et al., 2008]. In this context, the routing process is a mechanism of great importance that can take advantage of the WMN capacity in a scalable way and thus plays a key role in supporting all these application scenarios. For these reasons, some background information about WMN (Section 2.1) and the routing process (Section 2.2) is provided in the next sections.

2.1 Overview

WMN act as a common backhaul network which allows inter-operability between several heterogenous wireless technologies as well as providing an interconnection with the wired networks. In addition, WMN can also be defined as a special case of wireless multi-hop networks that combine various characteristics of other wireless networks, such as Mobile Ad hoc NETworks (MANET) [Corson and Macker, 1999], Worldwide Inter-operability for Microwave Access (WiMAX) [IEEE, 2005], and Wireless Sensor Networks (WSN) [Yick et al., 2008]. Furthermore, WMN have been defined as follows [Held, 2005]:

"A wireless mesh network is a packet-switched network with a static wireless

backbone."

The main elements of WMN are described in Sub-section 2.1.1. The architectures are presented in Sub-section 2.1.2. Sub-section 2.1.3 discusses the characteristics. Finally, the main tecnologies wireless which are usually employed in WMN in Sub-section 2.1.4.

2.1.1 Elements

WMN comprise three network elements, gateways, mesh routers, and mesh clients that can be defined as follows:

- ***Gateways***. These elements have bridging functionalities that allow the WMN to be integrated with other wireless networks. Moreover, they make a direct connection of WMN with the Internet or the wired network.

- ***Mesh Routers***. These elements form the backbone of the wireless network and provide connectivity to a range of services for the mesh clients (including the Internet). A mesh router can achieve the same coverage as a conventional wireless router with much lower transmission power, through a multi-hop routing strategy. In this way, they serve as a relay system to forward the traffic among the mesh clients and gateways.

- ***Mesh clients***. These include a wide range of devices, such as Personal Digital Assistants (PDA), laptops, desktops and cell phones, with varying degrees of mobility and the aid of network technologies. Mesh clients also have routing functionality insofar as they are able to participate in multi-hop routing.

Table 2.1 shows an outline of the specific characteristics of each component, according to the number of radios, mobility and power consumption constraints involved.

Table 2.1: Characteristics of the WMN Elements

Characteristics	**Gateways**	**Mesh Routers**	**Mesh Clients**
Number of Radios	Multiple	Multiple	Single
Mobility	Low	Low	High
Power-Consumption Constraints	No	No	Yes

2.1.2 Architectures

This sub-section sets out the main types of WMN architectures, namely infrastructure, client and hybrid [Akyldiz et al., 2005].

- **Infrastructure or Backbone WMN**. This type of WMN includes mesh routers and forms a kind of infrastructure that mesh clients can connect to. The WMN infrastructure or backbone can be built by means of various types of wireless technologies, in addition to the most widely used IEEE 802.11 technology [Wi-Fi, 2007]. Mesh clients can connect directly with the wireless backhaul if they have the same technology; otherwise, they must connect to base stations that have wired or wireless links with the wireless backhaul. Figure 2.1 shows this kind of architecture.

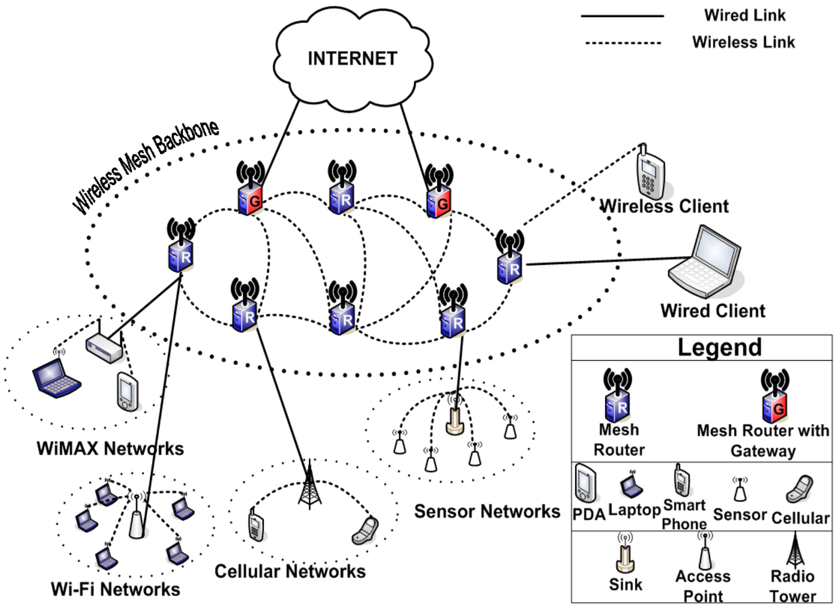

Figure 2.1: Infrastructure or backbone WMN - Adapted from [Akyldiz et al., 2005]

The mesh routers usually use multiple radio interfaces for the wireless backbone and a radio interface to connect with the mesh clients and different networks. Moreover, the mesh routers can also employ directional antennas to reach longer distances. This architecture is what is currently the most commonly used [Islam et al., 2010][Ashraf et al., 2011][Dely et al., 2010]. For these reasons, the proposed solutions in this thesis takes into account this network architecture.

- **Clients WMN**. In this type of architecture, mesh clients form a peer-to-peer network with routing and configuration functionalities, as well as provid-

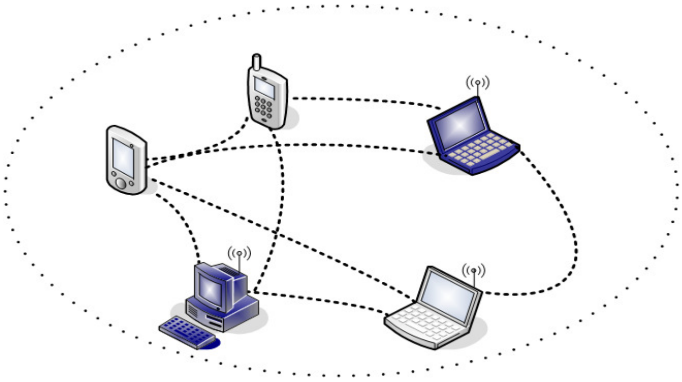

Figure 2.2: Client WMN - Adapted from [Akyldiz et al., 2005]

ing/consuming content. Moreover, mesh clients usually employ a type of radio that depends on devices, that are very similar to those of MANET. Figure 2.2 illustrates the Client WMN architecture.

- **Hybrid WMN**. This architecture is a combination of the infrastructure and the client's architecture, as it can be seen in Figure 2.3. While the wireless backbone provides connectivity with other networks.

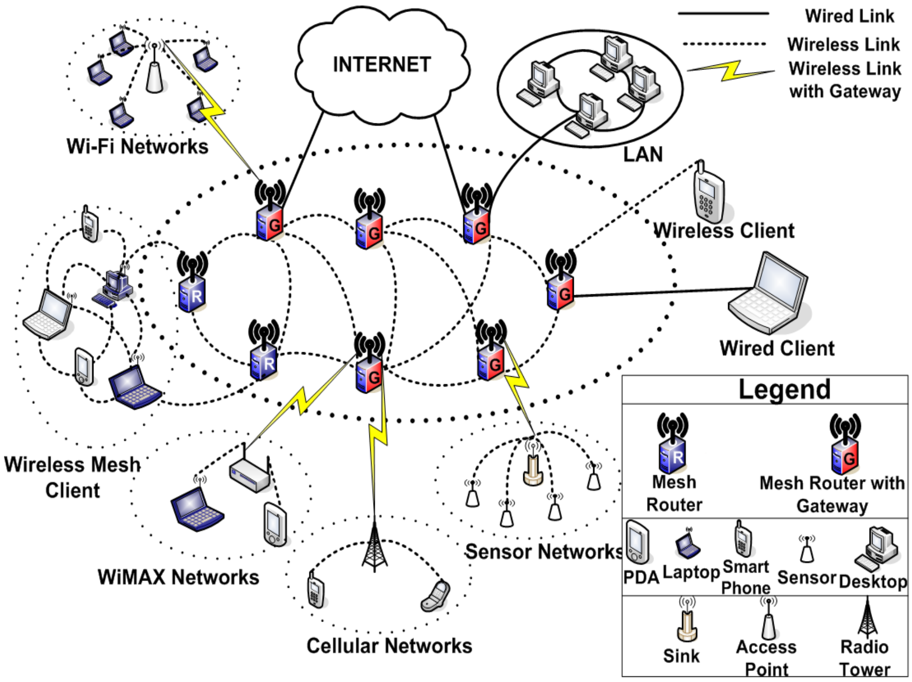

Figure 2.3: Hybrid WMN - Adapted from [Akyldiz et al., 2005]

The mesh clients can access the network through mesh routers as well as directly meshing with other clients, and thus improve the connectivity and coverage inside the WMN.

2.1.3 Characteristics

The main characteristics of WMN can be outlined as follows [Akyldiz et al., 2005, Held, 2005, Waharte et al., 2006]:

- **Multi-hop Wireless Network**. WMN extend the coverage range of current wireless networks by using multi-hop paths through shorter link distances and offering a more efficient frequency re-use. Thus, nodes tend to connect with each other through intermediate nodes rather than making a direct connection. This means that the data is forwarded from one device to another until it reaches its destination. As a result, the WMN can achieve a higher throughput without sacrificing the effective radio range, and can cover the same area with less transmission power than a traditional wireless router thus experience less interference between the nodes [Akyldiz et al., 2005, Waharte et al., 2006].

- **Self-organization Properties**. The features of the multi-hop wireless network give rise to self-configuration and self-healing properties. With regard to the capacity for self-configuration, nodes can be added and removed from the network when needed without the intervention of any centralized administrative infrastructure. For example, novice users can set up their own mesh node when they need to connect to the WMN quickly, especially if these devices use omni-directional antennas. The mesh routing protocol enables nodes to learn about their neighbours (i.e., connection failures or new users) and actively convey the data between themselves as the nodes enter and leave the network. As regards the self-healing property, if a node disappears from the network, due for instance to hardware failure, the network can still operate without the need for any special administrative intervention, because, as each node is connected to other nodes, the neighbours can find alternative routes to their destination. The self-healing capability depends on the number of alternative routes available. However, there tends to be an increase of interference when more alternative routes are added. Hence, it is necessary to obtain an equilibrium between the number of alternative routes and the contention levels to maintain network performance at an acceptable level, while providing sufficient self-healing capability. It is thus clear that WMN can provide connectivity and easy deployment through these self-organizing properties.

- **Low Up-front Cost**. There are several factors that explain this characteristic. These networks are able to employ Common-Off-The-Shelf (COTS) products to form their backbone. For instance, Personal Computers (PC) can be used as mesh routers and gateways. Since WMN have effective routing and configuration equipment (i.e. gateways and mesh routers), it is possible to make a significant reduction in the mesh clients' load. As a result, the mesh-clients' requirements are reduced, and this lowers the cost of the devices. Moreover, there are some application scenarios, such as building, businesses, or, home and neighborhood networks, where a wired Internet connection is not needed at every access point, since some access points can be replaced by mesh routers; this serves to reduce the deployment costs.

- **Network Structure and Traffic Patterns**. The positions of the gateways and mesh clients usually comply with a certain rule, i.e. the gateways are located at the opposite side of the mesh clients. As a result, most WMN traffic is usually between the mesh clients and gateways. This means that the mesh routers that are in close proximity to the gateway are much more likely to become congested. Nonetheless, there is also traffic between the mesh clients and mesh routers (i.e., intra-mesh traffic), which means that the WMN can support Peer-to-Peer (P2P) traffic model [Castro, 2011].

- **Interoperability with existing wireless networks**. The mesh router can play a role of bridge between different wireless technologies and thus, WMN can comprise of different wireless networks. For example, Wireless Sensor Networks, Wi-Fi, WiMAX, ZigBee and cellular networks.

2.1.4 Wireless Communication Technologies

Since the routing process complements the wireless technologies to improve the traffic performance in a scalable way and WMN can provide interoperability with diferent wireless networks, it is necessary to analyse the main wireless technologies to explain why the particular one was selected; as a result of this, the routing architecture can be specified in accordance with the chosen wireless technology. For these reasons, the most common wireless technologies used in WMN are discussed in this subsection, including some wireless technologies that have been employed in WMN. The main wireless technologies can be described in the following sub-sections.

2.1.4.1 Wireless Fidelity (Wi-Fi)

Wi-Fi [Wi-Fi, 2007] is a trademark which is used for the IEEE 802.11 family of standards and is the most popular wireless technology employed for WMN. This can be attributed to several reasons such as the fact that it operates in the license-free zone (unlicensed band) and is composed of commodity-off-the-shelf hardware which is readily available. Among the IEEE 802.11 specifications, IEEE 802.11 a/b/g are the most widely used for communications in backbone wireless, in outdoor and indoor city-wide scenarios [Barraca et al., 2008] (neighborhood and university areas) and operate in data rate of 54, 11 and 54 Mbit/s, respectively. However, there are some drawbacks that make it difficult to optimize the network performance when using this technology. For example, there is a high level of interference in this standard as a result of inconsistent definition of spectrum assignments and the operational limitations caused by its use of an unlicensed band, since the wireless medium can be shared with a large group of users.

It should be stressed that both IEEE 802.11e and IEEE 802.11n have also been employed in WMN [Xiao, 2005, Chu and Wang, 2010] to improve the performance optimization of the WMN. The former supports transmission with differentiation for some traffic classes, while the latter offers higher data rates (i.e. from 65 Mbit/s to 300 Mbit/s) but increases the cost. Since background information is provided in WMN, this chapter would not be complete without a discussion about the IEEE 802.11s standard [Bahr, 2007]. This standard will determine how the mesh elements can interconnect to create a WLAN mesh network, which can be used for static topologies and ad-hoc networks. To achieve this, this standard is concerned with MAC and aspects of routing layers. Nevertheless, the IEEE 802.11s standard still depends on either 802.11a, 802.11b, 802.11g, or 802.11n to carry out the application traffic.

2.1.4.2 Worldwide Interoperability for Microwave Access (WiMAX)

WiMAX [WiMAX, 2007] is also a trademark which is used for the IEEE 802.16 family of standards [WiMAX, 2011]. WiMAX supports a metropolitan area network and has a signal radius of between 30km and 50km. In addition, WiMAX offers data rates of up to 100 Mbit/s for mobile users and 1 Gbits/s for fixed users [802.16m, 2011]. For this reason, this technology has often been used as the communication technology for a backhaul WMN. It can also be combined with Wi-Fi technology, which has been used in the local area networks, and WiMAX, which has been used to enable the delivery of last mile wireless broadband access in metropolitan area networks. There are several versions of IEEE 802.16, IEEE 802.16d (fixed WiMAX [802.16d, 2002]) and IEEE 802.16e (mobile WiMAX [Etemad, 2008]) standards are usually employed. The fixed WiMAX uses Orthogonal Frequency Division Multiplexing (OFDM) technique of PHY layer which divides the wireless medium into 256 sub-carriers. This allows OFDM to mitigate the multipath interference and improve signal propagation, especially in Non-Line Of Sight (NLOS) coverage areas.

The fixed WiMAX makes possible to prioritize traffic, for instance the time sensitive traffic (voice and video) is given priority over non-time sensitive traffic (data). However, the fixed WiMAX does not support mobility management and handoffs. The mobile WiMAX was recommended as a means of overcoming this limitation and employs an Orthogonal Frequency Division Multiple Access (OFDMA) technique that divides the carrier into more sub-carriers (up to 2048) than OFDM, as well as supporting MIMO. Thus, with the mobile WiMAX the interference is even further reduced. The IEEE 802.16e standard also enables a trade-off to be made between throughput and the coverage area. For example, the base station can reduce the number of channels while increasing the gain of the signal to each channel and thus reach users farther away. As well as this, when a user gets closer to a cell site, the number of channels will increase and the modulation can also change to increase bandwidth. However, the mobile WiMAX is more expensive than the fixed WiMAX.

IEEE 802.16n [802.16n, 2011] is a WiMAX standard that extends the OFDMA technique to support multi-mode operations (mobile, relay and base station modes) with radio path redundancy that improves robustness in degraded network conditions (i.e. node failures or network connectivity) and in this way, it provides a self-healing property for WiMAX. The IEEE 802.16n standard is still in an incomplete stage. Even though WiMAX provides ways to reduce interference through OFDMA, the wireless medium is also shared between users in a given radio sector and this means that there may be a high level of interference which can reduce the level of performance if there are many active users in a single sector. WiMAX operates in licensed bands which makes it difficult to install for political/administrative reasons. In addition, the cost and power consumption of WiMAX equipment are higher than those of Wi-Fi. For example, [Kuran et al., 2007, Kong et al., 2009, Liu et al., 2009] employ WiMAX as wireless technology.

2.1.4.3 Free-Space Optical (FSO)

FSO [Garlington et al., 2005] is a wireless technology that has been used for short and long-haul space communications and is based on laser. However, FSO has attracted attention as an effective means of communication in backhaul WMN [Smadi et al., 2009, Kashyap et al., 2007b, Moradi et al., 2010]. FSO can provide high data rates (up to 4 Gbit/s) and medium distances (up to 4kms) in terrestrial networks. In addition, FSO operates in an unlicensed spectrum as well as entailing lower costs and power consumption than Wi-Fi and WiMAX. Nonetheless, both performance and reliability are severely curtailed when FSO is used, since it is very sensitive to atmospheric phenomena such as fog, dust, sand, and heat which can cause high level of interference in this technology and degrade or interrupt the performance of FSO links. The transmission range of FSO has been reduced to increase the reliability of the FSO links; for instance, 200 to 500 meters has proved to be sufficient to support acceptable grades of performance. Hence, this technology has shown it has significant potential and can be applied to temporary networks, since it is is able to become established or relocate quickly and hence be able to deal with unreliability issues in the network.

2.1.4.4 Cognitive Radio

The radio spectrum is statically allocated, divided and scheduled to its own type of band (i.e. licensed or unlicensed) in Wi-Fi and FSO technologies (which were discussed earlier). For example, Wi-Fi users only transmit data in the unlicensed band that has been pre-defined for this wireless technology. As a result of this, some frequency bands are increasingly experiencing a scarcity of radio spectrum, while large parts of the entire radio spectrum remain unused, regardless of time and location. The problem of the spectrum scarcity of Wi-Fi is addressed by cognitive radio [Akyildiz et al., 2009] which is a network paradigm that seeks to ensure that the radio spectrum is used in a more flexible and efficient way [Alsarhan and Agarwal, 2009, Marinho and Monteiro, 2011].

It does this by allowing wireless devices to opportunistically access parts of the idle radio spectrum of licensed band without causing any degree of interference to the licensed users However, this concept has not reached an advanced stage yet and, there remain a lot of open issues such as, how to prevent secondary users from interfering with primary users, and how to provide accurate information to make a spectrum decision while reducing the overhead of signalling. The IEEE 802.22 standard [Stevenson et al., 2009] is the specification that enables cognitive wireless regional area networks to be employed.

2.1.4.5 Long Term Evolution (LTE)

LTE [Motorola, 2011] is a wireless technology for high-speed data which has evolved from the Global System for Mobile Communications/Enhanced Data rates for GSM Evolution (GSM/EDGE) and Universal Mobile Telecommunications System/High Speed Packet Access (UMTS/HSPA) network technologies. For example, this tecnology enables the wireless devices to achieve a high data rate (up to 300 Mbit/s), high performance mobile data (up to 350 km/h), a large coverage area (up to 100 km), carrier bandwidth (i.e. from 1.4 MHz to 20 MHz) and low latency (10ms), through leading-edge hardware and Digital Signal Processing (DSP) techniques that have recently been developed. Although LTE is based on a simple network architecture, it offers good support for Frequency Division Duplexing (FDD) and Time-Division Duplexing (TDD) in the same platform. As a result of these improvements, LTE offers a cost-effective value proposition.

It is supposed that LTE comprises the 2G and 3G networks, however the wireless interfaces of 2G and 3G devices are usually incompatible with LTE wireless interfaces, since they operate in different wireless spectrum. To overcome this limitation, there are companies (e.g. Nortel and LG) that have already provided products that enable handoff between LTE and the 2G/3G networks. Thus, existing 2G and 3G spectrum can gradually be re-farmed to LTE. This technology is an ongoing technology that has not been fully tested yet, although there are some research projects regarding this technology in WMN, for instance in [Ouni et al., 2011].

2.1.4.6 Discussion of the Wireless Technologies

Table 2.2 displays a comparison chart of the wireless technologies depicted in this sub-section.

Table 2.2: Comparison of the Wireless Technologies

Factors/ Tecnologies	**Family**	**Radio Technology**	**Coverage Area**	**Cost**	**Data rate**
Wi-Fi	IEEE 802.11(n)	OFDM/MIMO	1km	Low	600 Mb/s
WiMAX	IEEE 802.16(m)	MIMO-OFDMA	50km	High	1 Gb/s
FSO	Optical	Laser	4km	Low	4 Gb/s
Cognitive Radio	IEEE 802.22	OFDMA	100km	Low	22Mb/s
LTE	3GPP	MIMO-OFDMA/SC-FDMA	100km	High	1 Gb/s

There is still no consensus about which technology will be able to provide broadband wireless access to large areas. The reasons for this ongoing debate is that there are conflicting objectives regarding what the selected technology should support. On the one hand, technologies which increase the capacity of wireless links significantly are usually expensive. On the other hand, it is hard to provide performance guarantees when adopting a cheap wireless technology. In view of this, the Wi-Fi standard was chosen to be employed as the wireless technology, specifically the standards 802.11a/b/g, since it is the most widely used wireless technology for WMN.

The main reasons for this are that this technology is simple to deploy and incurs low costs and thus makes it easier to increase the network size and enlarge the coverage area in a multi-hop WMN. Moreover, bureaucratic delays are avoided when installing it, since Wi-Fi uses unlicensed bands and COTS can be reused to form the WMN, with minimal up-front investments, which allows the network to grow in size. However, when using this technology, there is still a need for a solution that offers more scalability for WMN to accommodate this growth. Furthermore, there are several kinds of application scenarios that can be supported with the aid of this technology in WMN, such as broadband home networking, community and neighborhood networks, enterprise networking, metropolitan area networks, and health and medical systems.

2.2 Routing Process

Since the proposal of Wi-Fi technology, a great deal of effort has been devoted to improving the MAC and physical layers to make it possible to take advantage of innovative techniques (such as MIMO and Cognitive Radio) and to meet the performance requirements of the WMN. Due to the high cost and complexity of the current means of addressing the problems of the MAC and physical layers, considerable research is still needed to consolidate these technologies in WMN. The problem remains that, although these technologies improve network performance by increasing its capacity, interference still persists and this can reduce the potential capacity of these technologies. In addition, these technologies are unable to scale in a way that can ensure free-collision transmissions for every user, every time, and this situation is worse in large-scale WMN. Hence, regardless of what wireless technology is adopted, a routing process that is aware of interference, has the potential to provide a more scalable solution for WMN.

The routing process [Doyle and Carroll, 2005] has three main components, namely the routing protocol, algorithm and metric. The routing metric provides information support to calculate the routes. The routing protocol specifies how the metrics should be disseminated in the network to report changes, and allow dynamic adjustments to the network conditions, as depicted by the routing metrics. The routing algorithm uses the metric to select the network paths.

A large number of routing protocols for WMN have been proposed in the literature, although most existing mesh routing protocols are based on routing protocols for MANET. For this reason, this section details the main unicast routing protocols of MANET that can be extended for use in WMN. The most recent and relevant mesh routing protocols will be explained in the next chapter. Basically, the routing protocols are classified in line with the routing strategy. There are three routing strategies in MANET and WMN, namely, proactive, reactive and hybrid. In view of this, the routing protocols are grouped in these routing strategies, as will be outlined in the following sub-sections.

2.2.1 Proactive Routing Strategy

Proactive protocols are based on a table-driven approach in which every node maintains updated information about the whole network topology. This is possible because of a constant exchange of routing information, which is transmitted by flooding. As a result of this, routes are immediately available at anytime, even though the overhead associated with the periodic exchange of messages can impose an excessive burden on the network. There are two kinds of proactive routing protocols which can be defined as follows: distance vector or link-state. On the one hand, nodes share their routing table with their neighbors in distance vector routing protocols and thus can calculate the routes by consulting the neighbors' routing table [Chen and Heinzelman, 2007]. On the other hand, in link state routing protocols, all the nodes has full information of the network connectivity which is used to calculate their own routing table. Examples of each kind of proactive routing protocol are given in next pages.

2.2.1.1 Destination-Sequenced Distance Vector (DSDV)

DSDV [Perkins and Bhagwat, 1994] is a distance vector routing protocol based on the Bellman-Ford algorithm [Bellman, 1958] which solves the routing loop problem through the sequence number in each entry of the routing table, i.e. each entry is labelled with a sequence number that has been generated by the destination. This number is incremented by the destination whenever it sends its reachability information. The route labeled with the highest sequence number is always used as this ensures the freshness of the routing information and in this way, loops are avoided. There are two types of packets that minimize the overhead generated in

DSDV. The former is known as a full dump packet which is a packet that is aware of all the information about the changes and for this reason, these full dump packets are not often sent. The latter is known as the incremental packet and this is only used to propagate occasional changes in the topology. The incremental packets are sent periodically. However, DSDV is unsuitable for highly dynamic networks because the network can take a long time to re-converge when there is a change, since a new sequence number is necessary. Furthermore, this protocol incurs high routing overhead in large networks caused by the flooding process.

2.2.1.2 Babel

Babel [Chroboczek, 2011] is an extension of the DSDV protocol that can lead to fast convergence. In view of this, according to the authors, Babel uses a history-based routing selection to reduce the routing oscillation which is caused by alternating different routes for the same source-destination pair. Hence, the routing process in Babel would rather use the previously selected route than the newest one. Babel performs triggered updates and explicit requests for routing information when there are link failures in the preferred route. However, although Babel speeds up the routing convergence, it does not propose any specific mechanism to prevent the routing overhead caused by the flooding procedure, in a similar way to DSDV. This means that Babel can also result in a poor performance when the size of the network increases its size.

2.2.1.3 Optimized Link State Routing Protocol (OLSR)

OLSR [Jacquet et al., 2001] is a link state routing protocol where the routes are always immediately available. Every node uses the routing information to compute next hop destination for all the nodes in the network using a shortest path routing algorithm (e.g. Dijkstra algorithm [Dijkstra, 1959]). Hence, OLSR constantly stores routes to all destinations in the network. This protocol uses `hello` and `Topology Control (TC)` messages to discover and propagate the link state information throughout the WMN. As a result of the `hello` message, the OLSR protocol discovers 2-hop neighbor information at each node and performs a distributed election of a set of Multi-Point relays (MPRs). This creates a path to each of its 2-hop neighbors via a node selected as an MPR which originates, aggregates and forwards the `TC` messages that contain the MPR selectors.

There are many benefits that the MPR concept provides to this protocol - for example, the fact that all the routing information is not shared among all the nodes, but only for a subset of nodes. In other words, only the links that represent the MPR selections are advertised. In this way, MPRs help to reduce the routing overhead, while keeping the nodes with updated routing information. OLSR benefits the networks where most communication is concentrated between a large number of nodes which characterizes a dense network. Hence, it is suitable for the infrastructure architecture of WMN, in which most of the mesh routers communicate with the gateway most of the time. Even though the MPRs decrease the routing overhead, this protocol still provides a high routing overhead in large networks which require a reasonably large amount of bandwidth. However, since it is a proactive protocol, the increase of routing overhead might be disproportional to the increase in the number of nodes and thus, the scalability of the OLSR protocol is constrained.

2.2.1.4 Better Approach to Mobile Ad-hoc Networking (BATMAN)

BATMAN [Johnson et al., 2008] is a proactive routing protocol that provides an original approach to the routing strategy. The main goal of this protocol is to prevent flooding by distributing the routing information in an intelligent way. Hence, no node has complete routing information about the network topology, like OLSR. To achieve this, each node only processes information about the message direction which received the data from (i.e. originator-messages). By counting these messages, the packet only takes a first step in the right direction. Routes are created dynamically when the data travels through the whole network. This process is repeated until the data reaches its destination. Every node periodically sends out broadcast messages to inform its neighbours of its existence. Although this protocol improves scalability by reducing routing overhead, the originator-message metric used by BATMAN is not accurate enough to measure the link quality in the wireless network, i.e. it favours asymmetric links, since it only reflects the link quality in a backward direction.

2.2.2 Reactive Routing Strategy

Reactive routing strategy acts on the principle of on-demand routing, i.e. a node does not keep routing information or disseminating routing messages if there is no data communication. If a node needs to send a packet to another node, the routing protocol starts the route discovery to establish the connection to the destination. The route discovery usually occurs by flooding the route request packets throughout the network. By means of this process, a reactive strategy is able to reduce the routing overhead significantly. However, it introduces delays in data traffic, since it takes more time to compute the routes. Although the reactive strategy has also been employed in WMN, this strategy was preferred in MANET and WSN. The reason for this is that these networks have dynamic topologies and strict resource constraints (e.g. data rate, battery and CPU processing), and this means that a routing strategy which consumes more resources cannot be suitable for these networks [Waharte et al., 2006].

2.2.2.1 Dynamic Source Routing (DSR)

DSR [Johnson and Maltz, 1996] is a source routing protocol which works entirely on demand with no periodic routing messages. There are two mechanisms which are as follows: route discovery and route maintenance. In the route discovery mechanism, a `ROUTE REQUEST (RREQ)` message is broadcast by the source node. This message adds destination, source, and a route record, together with an empty list of addresses which will be stored at all intermediate nodes and a unique request id. These messages are forwarded in a hop-by-hop way. After receiving a request, an intermediate node stores the route record (i.e. own address id); if it is not the destination, it appends its address to the route record (i.e. list of addresses) and broadcasts the `RREQ`. If the message arrives in the destination node, this node replies with a `ROUTE REPLY (RREP)` message. The destination can receive multiple `RREQ` messages through different paths, but the algorithm will pick up the shortest one the basis of the hop count metric. The destination appends its address to the list header and sends a reply packet on the reversed route. In addition, the destination node also stores the route record which is used for the propagation of `ROUTE REPLY (RREP)` back to the source node. Thus, the source node knows the complete route to the destination.

DSR prefers to use the single request id for each message to reduce the overhead. In the route maintenance mechanism, DSR does not employ any periodic messages from the nodes which fail to send packets to its next hop. It uses two types of packets for route maintenance: `ROUTE ERROR (RER)` and `ACKnowledgements (ACKs)` packets. A `ROUTE ERROR (RER)` message is sent to the source node. Upon the receipt of a `RER` the source node concludes that the path is no longer valid and sends a `RREQ` again. `ACKs` packets are used to check the correct operation of the route links. Moreover, the broken link is removed from the route cache of the source node. As a result of this, DSR reduces the routing overhead. However, it can take a long time to establish a route in DSR in large-scale networks and this may not be acceptable to multimedia applications.

2.2.2.2 Ad hoc On-Demand Distance Vector routing (AODV)

AODV [Perkins, 1999] is a distance vector routing protocol that works in a similar way to DSR. AODV obtains and maintains routes only as long as data packets are sent along the route. AODV uses sequence numbers to ensure the freshness of routes as well as propagating a `RREQ` to the whole network. All the nodes which receive this broadcast message store backward pointers to the source node. The destination responds with a `RREP` message. In this phase, the intermediate nodes set up forward pointers to the destination node. Thus, the routing information is stored locally at each node in contrast to DSR and this explains why the addresses of the intermediate nodes are not included in the routing information of the message.

An intermediate node that receives a `RREQ` message only replies to the source node using a `RREP` message if it has a route to the destination node whose corresponding destination sequence number is greater or equal to the one contained in the `RREQ` message. The `RREQ` message also contains the most recent sequence number for the destination which the source node is aware of. Once the source node receives the `RREP` message, it may start to forward data packets to the destination. Later, the source node can update its routing information for that destination, if it receives a `RREP` message containing a greater sequence number or the same sequence number with a smaller hop count, and in this way, it can begin to use the better route. Although AODV speeds up the route discovery process, it still takes a long time to discover a route in large-scale networks.

2.2.2.3 DYnamic MANET On Demand (DYMO)

DYMO [Chakeres and Perkins, 2007] is a reactive routing protocol which takes into account previous experiences of reactive routing strategy (e.g. AODV). In light of this, it seeks to simplify the protocol implementation by reducing the system requirements of participating nodes. On the one hand, it employs mechanisms from previous reactive routing protocols, for instance the use of sequence numbers to enforce loop avoidance. On the other hand, DYMO provides enhanced features, such as the following: implementing a path of accumulation in which each intermediate node records a route to the source in the `RREQ` as well as to the destination in the `RREP` during this hop-by-hop discovery process; nodes that extend route lifetimes upon successfully forwarding a packet to preserve the routes in use; and nodes that monitor links over which traffic is moving so that they can react more quickly to changes in the network topology. As a result of these improvements, this protocol can achieve better traffic performance in small and medium-sized networks. In large networks, DYMO can be more suitable when the nodes only communicate with a few of the neighboring nodes. However, when the communication involves many nodes in large networks, the traffic performance can decrease significantly.

2.2.3 Hybrid Routing Strategy

The hybrid routing strategy consists of both proactive and reactive strategies which attempt to exploit their advantages to optimize the routing process. Nevertheless, this strategy is complex to develop, since it depends on a trade-off between the proactive and reactive strategies. For this reason, there are few routing protocols that employ this strategy in MANET and WSN. It has attracted more attention in WMN, because of the heterogeneous nature of this kind of networks (e.g., different networks architectures and traffic patterns). Some examples of routing protocols which use this strategy are given below.

2.2.3.1 Zone Routing Protocol (ZRP)

ZRP [Haas et al., 2002] is the first hybrid routing protocol in MANET. It proposes the concept of zone that consists of the k-neighbourhood of the nodes. In other words, all the nodes within the k hops of the node are inside the node's zone. The proactive component can be defined as intra routing zones for which a route to a destination can be established from the source proactively. It is worth noting that the proactive component is only valid if the source and destination of a packet are in the same zone. Otherwise, the reactive component that should be used is that which works in a similar way to previous reactive routing protocols. In other words, the source node sends a `RREQ` message to the border nodes which are nodes that are exactly k hops away from the source of its zone. This message has the source and destination addresses as well as a single sequence number.

Each border node checks its local zone for the destination. If the destination is not a member of this local zone, the border node includes its own address to the route request packet and forwards the packet to its own border nodes. Otherwise, it sends a `RREP` message on the reverse path, back to the source node which uses the path in the route reply packet to send data packets to the destination. The main disadvantage of this protocol is that it increases complexity. For instance, it is difficult to define the zone radius to minimize the overhead and delay in the intra-zone and inter-zone routing respectively, since the zone radius is the parameter which influences the efficiency of ZRP.

2.2.3.2 AODV-Spanning Tree (AODV-ST)

AODV-ST [Ramach et al., 2005] is a hybrid routing protocol which is mainly based on the AODV protocol. It aims to improve the AODV in some respects. First, AODV attempts to discover a route that is no longer the optimal route due to network congestion or the fluctuating characteristics of wireless links. Since it is a reactive routing protocol, AODV can result in a high route discovery latency in large-scale networks. To improve these features, AODV-ST employs the proactive maintenance of spanning trees whose roots are the gateway nodes, thus reducing route discover latency while avoiding loops. Furthermore, it employs IP-in-IP tunnels to route data traffic from mesh routers to the gateways. As a result of this, AODV-ST eliminates unnecessary route discovery overhead for external destinations that are only reachable via the gateway, which also reduces the route table size at each mesh router to the sum total of mesh routers and gateways.

2.2.3.3 Hybrid Wireless Mesh Protocol (HWMP)

HWMP [Bahr, 2007] was chosen as the default routing protocol in the IEEE 802.11s standard. The HWMP protocol employs a radio-aware routing metric which is used by the mesh routers to select the best routes to the gateway, although there is no routing metric to choose the best gateway. This routing metric will be discussed in Chapter 4 in greater detail. HWMP employs a hybrid routing strategy in which the chosen strategy (i.e., reactive or proactive) depends on the network archictecture which currently exists in WMN.

On the one hand, the reactive strategy of this protocol assists in the route discovery between the mesh nodes (i.e. mesh routers and mesh clients) which is equivalent to the WMN client and is largely based on the AODV protocol. On the other hand, the proactive strategy aims to discover routes to the gateway in a hybrid or WMN infrastructure. The proactive strategy has some special features, such as the fact that it is based on a routing tree which is rooted on the gateway node and there are two kinds of proactive tree building mechanisms. The former mechanism uses a `RREQ` message to build the tree, including gateway and all the mesh nodes, which is periodically sent by the gateway. When a node receives this message, it stores the route information at the gateway and sends a `RREP` message back to build an updated route between this node and the gateway. The latter mechanism periodically uses the `Root ANNouncement message (RANN)` which periodically notifies the mesh nodes about the existence of the gateway. Every mesh node creates or updates a route to the gateway whenever it receives a `RANN` message. It then sends unicast

RREQ along the reversed path. Once the gateway receives this request message, it is able to send a RREP message to the source node so that it can build a route to the gateway.

It is worth pointing out that the gateway keeps routes to all the mesh nodes when each mesh node keeps a route to the gateway for both mechanisms. However, the hybrid routing strategy of this protocol is more concerned about how to adapt the routing strategy to the network architecture than with discovery routes that can improve the traffic performance in different areas of the WMN network architecture. In other words, HWMP raises scalability performance issues similar to those of OLSR and AODV. For example, hybrid WMN use a proactive strategy (i.e. they may have several gateways) and thus, high levels of routing overhead are still being generated. In the client WMN architecture, the reactive strategy can also result in a long delay in obtaining information about the route between the source and destination.

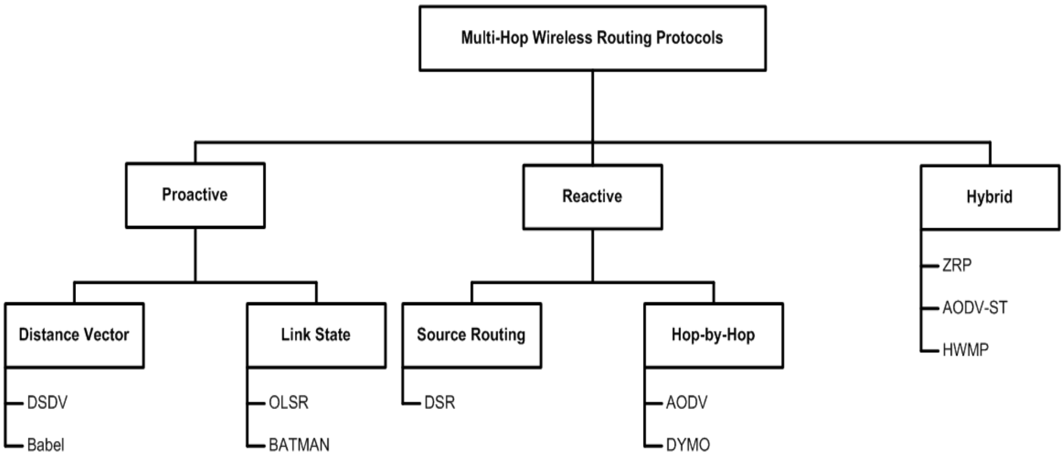

Figure 2.4: Taxonomy of routing protocols for wireless multi-hop networks - Adapted from [Ashraf, 2010]

Figure 2.4 illustrates a taxonomy for the routing protocols which were discussed in this sub-section. Most of the routing protocols that are displayed rely on the hop count routing metric which does not fully capture the link quality; a further drawback is that, they have poor performance when the network size increases. They also have a higher routing overhead in proactive strategy and it takes a long time for them to discover the route in reactive strategy. Hence, these discussed protocols fail to take full advantage of the potential scalability of WMN. For these reasons, many other routing approaches have been proposed for WMN to provide an enhanced method for dealing with the routing problem; these will be discussed in the next chapter.

2.3 Summary

This chapter involved conducting a general discussion of WMN. It detailed the characteristics, elements, network architectures and wireless technologies that are most used in these networks. It is worth noting that some recent wireless technologies have increased the performance of wireless links, although interference can still disturb the wireless communication even in the most recent technologies. Thus, the routing process is still necessary to optimize the performance of traffic applications in a scalable way. In light of this, the chapter also introduced some basic principles with regard to the routing process in WMN, as well as describing and analyzing the WMN routing protocols which are based on the routing protocols from MANETs and WSN. The proactive routing protocols show characteristics that enable them to provide a more efficient method of dealing with multimedia applications than reactive protocols, such as those available for routing information at any time. It also enables them to make more precise routing decisions. However, the high routing overhead of this strategy damages the performance of WMN, since this strategy can consume a considerable amount of wireless resources in large networks. Although reactive protocols reduce the routing overhead significantly, they are not suitable for multimedia applications due to the high latency of the route discovery mechanism. Furthermore, the existing hybrid protocols still fail to achieve an efficient trade-off between reactive and proactive strategies which are needed to provide improvements in traffic performance while keeping control of the routing overhead. Hence, these strategies fail to employ methods that can make possible a routing procedure to improve scalability for WMN.

Chapter 3

Architecture for Cooperative Routing Management (ACRoMa)

Owing to its special features, scalability is a challenging research issue in Wireless Mesh Networks (WMN). For instance, the shared wireless medium is usually unpredictable due to interference. Moreover, WMN enables a last-mile wireless hop to access the Internet which has many applications requiring certain performance criteria. From this standpoint, the routing process is crucial for scalable solutions to WMN. Certain factors have to be taken into consideration to provide a more scalable solution for WMN when using a routing process. First, interference caused by the large number of nodes which share the wireless medium significantly reduces the network performance. Next, most traffic applications in WMN follow a specific pattern, i.e. they are derived from the mesh clients and then forwarded by the mesh routers towards the gateway. Thus, congested regions can exist close to the gateways. Finally, the traditional routing protocols from Mobile Ad hoc NETworks (MANET) and Wireless Sensor Networks (WSN) which are used in WMN, result in large overhead and high levels of delay in large-scale networks. To address the involved issues, the most recent and relevant routing approaches in WMN are analysed in Section 3.1. Section 3.2 outlines an architectural model for routing management which is proposed to offer a more complete solution for improving scalability in WMN. An additional discussion about the advantages of the architectural model is presented in Section 3.3. This chapter is summarized in Section 3.4.

3.1 Related Work on Routing Approaches

The most recent and important approaches to routing for WMN are examined in this section through a survey which groups the routing approaches into three main categories that represent the issues involved in providing a more general form of scalability in a more complete way. These categories are as follows: interference (Sub-section 3.1.1), routing overhead (Sub-section 3.1.2) and congestion at the gateways (Sub-section 3.1.3).

3.1.1 Interference

Although recent radio technologies have mitigated interference, this phenomenon still influences performance in wireless networks. For this reason, several routing approaches have to be taken into account.

3.1.1.1 Interference-Aware Topology Control and QoS Routing in Multi-Channel Wireless Mesh Networks

Tang et al. define the minimum INterference Survivable Topology Control (INSTC) problem [Tang et al., 2005] which seeks a channel assignment algorithm for a given network to ensure that the induced network topology is a minimum interference channel set among all the K-connected topologies [Penrose, 1999]. Hence, INSTC also considers connectivity, as K-connectivity is required for survivability and load-balancing purposes. In this way, the authors exploit the influence of contentions for the multi-hop routing. K-connected topology is a concept from graph theory that defines the connectivity of a topology, i.e. a K-connected graph requires the removal of the k links to disconnect it.

A routing algorithm was proposed to solve the formulated Bandwidth-Aware Routing problem based on the the K-connected topology. The purpose of this algorithm is to ensure that the bandwidth allocated for existing connections is not affected by new requests. However, the scarce bandwidth of the wireless medium does not allow this solution to scale the performance with an increase of traffic load and nodes. Furthermore, it is also a centralized approach and thus causes performance problems when the network increases in size.

3.1.1.2 Integrated Routing and MAC Scheduling in Multi-hop Wireless Mesh Networks

Integrated Routing and MAC scheduling (IRMA) aims at avoiding interference contentions by creating a conflict-free schedule based on traffic demand across all the end-to-end routed paths [Wu et al., 2006]. Global optimality can be achieved by allocating schedules and paths simultaneously for each of the source-destination traffic pairs in the network, taking into account transmissions which do not interfere with each other. The problems of hidden and exposed nodes are avoided by arranging the conflicting transmissions at different time slots.

This solution is based on the available bandwidth metric. In the proposed Link Scheduling - Bandwidth Aware Routing algorithm, the local information about the potential MAC bandwidth is obtained before selecting a route for each flow. The available bandwidth is measured by the number of free slots. The metric of a link is the number of occupied and scheduled slots in a given Time Division Multiple Access (TDMA) frame. It is important to highlight that the scheduling of time slots is not a scalable solution. Furthermore, the IRMA algorithms are performed in a centralized manner.

3.1.1.3 Available Bandwidth Estimation and Admission Control for QoS Routing in Wireless Mesh Networks

Ergin et al. propose an admission control mechanism that is integrated with Lightweight Underlay Network Ad-Hoc Routing (LUNAR) protocol [Tschudin et al., 2004]. Since these wireless networks can provide a multiple range of paths between a source/destination pair, integrating admission control into the routing process enables alternative routes to be found if the shortest path is congested. The main contributions to this study consist of two methods for available bandwidth estimation, an admission control mechanism and a routing protocol extension [Ergin et al., 2008].

Dual-Carrier Sense with Parallel Transmission-awareness (DCSPT) is the first proposed method of available bandwidth estimation which makes use of the opportunities for parallel transmissions, while attempting to avoid flows that might violate the QoS requirements of its neighbours. It achieves this by exploiting the adjustable Carrier-Sensing Thresholds (CS-threshold) of the wireless transceivers. As a result, a node can be aware of its surrounding transmissions by changing the CS-threshold. However, DCSPT requires the hardware of the network interface to offer full support for the changing carrier-sense functions, which cannot be provided by some vendors. In view of this, a method was proposed which can be used in off-the shelf mesh networking equipment, called Packet Probing with Request-To-Send (RTS) and Clear-To-Send (CTS) Handshake (PPRCH). The RTS/CTS mechanism helps to determine the wireless contention caused by the hidden nodes around the measuring station.

This method sends two small back-to-back probe packets which are used for depicting the dispersion on the channel due to traffic load. If there is a large time interval between these two packets, the traffic load in the sensing area can be high. The authors point out the need for a scalable deployment strategy, since the parallel transmissions cannot be scalable. In addition, it is difficult to estimate the bandwidth that each flow will require or the available bandwidth in a wireless link. For this reason, there is a risk that the control admission mechanism of DCSPT might underestimate/overestimate the available bandwidth when based on the reserved bandwidth of the data rate of each flow. Furthermore, DCSPT assumes that a routing mechanism is concerned with interference in a static way (e.g., available channel time). For instance, it does not take into account the degree of instability of the wireless networks in a dynamic way (i.e., the level of interference caused by each interfering node on a link change over a period of time).

3.1.1.4 Supporting Carrier Grade Services over Wireless Mesh Networks: The Approach of the European FP-7 Strep Carmen

The aim of the CARrier grade MEsh Networks (CARMEN) project [Azcorra et al., 2009] is to design a WMN architecture to support triple-play services at a significantly low cost. In the light of this, CARMEN will employ self-configuration and management techniques in all phases, from planning to deployment, as well as in operations. In addition, it is based on an abstract interface which can support heterogeneous wireless technologies. CARMEN architecture comprises three main components which are resource management, spectrum management and self-configuration.

Resource management combines admission control and resource aware routing, in which admission control takes note of information from the mesh clients and the available estimated bandwidth, whereas the routing approach is based on a multipath solution to provide multiple connections to the backbone, as well as a cross-layer design between the network and MAC layers so that the network layer can be aware of link-quality measurements. However, the routing approaches still fail to offer any mechanism to reduce the routing overhead. The spectrum management enables the cognitive radio and channel assignment to improve the capacity of WMN. The self-configuration component is concerned with mobility management and the deployment of WMN. However, this is still an on-going research project. For this reason, it is difficult to assess how the synergies of CARMEN's components will improve the scalability of WMN.

3.1.1.5 Route Maintenance in IEEE 802.11 Wireless Mesh Networks

The Efficient Route Maintenance (ERM) [Ashraf et al., 2011] is proposed for improving link breakage detection for the on-demand routing protocols. If any intermediate link in the route fails to send a packet to the next hop (i.e. the number of retransmissions is greater than the transmission failure threshold), the physical layer reports a transmission problem to the network layer. This means that there is a route breakage and so the source node has to start the route discovery procedure to find a new route. Frequent route breakages result in an increase of routing overhead (e.g. route error and request), poor performance and higher route instability. Route breakages often occur in the WMN IEEE 802.11 infrastructure. The authors point out that the transmission failure threshold is not accurate, since the link breakage may have been caused by transient phenomena in the wireless networks (e.g. interference or noise). For this reason, good links are sometimes considered to have been broken by the routing protocol when they have been observed to become active immediately afterwards. As a result, ERM attempts to distinguish good links from broken links more precisely through a cross-layer solution which uses a MAC and physical layer information to make the link breakage decision.

ERM has two main components which are the Link Quality Assessment (LQA) and the Link Breakage Decision (LBD). LQA periodically estimates the long-term link quality on the basis of the physical and channel (i.e. MAC) link quality. When the MAC layer fails to transmit a data packet to the next hop node, it reports the transmission problem to LBD at the network layer. The main purpose of LBD is to decide whether the link is broken, on the basis of the information provided by the LQA. For example, a link which has few transmission failures and good physical and channel link quality, represents a good link, whereas a link with continuous transmission failures and poor physical and channel link quality over a long period of time, represents a broken link. The combination of physical and channel link quality information in a single measurement, provides more accuracy for the link breakage decision than the total number of retransmissions.

Although ERM improves the traffic performance and reduces the routing overhead for medium networks in high traffic loads, it does not propose a mechanism to speed up the route discovery in a large scale WMN or reduce overload situation in the gateways. Furthermore, this approach is specifically focused on the reactive routing protocols and thus, it cannot attain the same level of improvement as when proactive routing protocols are employed.

3.1.1.6 A Quality Based Routing Protocol for Wireless Mesh Networks

The Interference and Delay Aware Routing (IDAR) reactive routing protocol [Pal and Nasipuri, 2011] is proposed, based on a cross-layer routing metric that improves the ratio of the end-to-end delay, called Probability Of Success (POS). The authors argue that some routing approaches employ control packets to estimate the link quality, but the control packets consist of different schemes from the data packets. For example, they are broadcasted (whereas data packets use unicast or multicast schemes), they are smaller in size and are sent at a lower transmission rate than the data packets. To overcome this problem, they predict that the link quality can be obtained through a cross-layer routing metric that employs offline measurements for the current data packet transmissions.

IDAR also assumes that all the traffic loads are directed towards the gateway which also plays the role of manager for all the routing decisions that are based on a global knowledge of node locations and activities. IDAR considers the IEEE 802.11 with and without RTS/CTS. The disabled RTS/CTS mechanism is used for the sake of simplicity and in this case, the probability of a successful packet transmission is only dependent on the probability of a successful reception of the data packet by the receiver, which is based on the Signal to Interference-plus-Noise Ratio (SINR) threshold. At the same time, an enabled RTS/CTS mechanism is the general case and a data packet is only transmitted when the RTS/CTS exchange has been successful.

The authors propose mathematical models for estimating key measures that influence the link quality over wireless networks, such as channel access probability and delay, which are combined in the cross-layer routing metric. For example, queuing delays (i.e. the time that a data packet has to wait in its transmission queue before actually reaching the head of the queue) and access delays (i.e. the time that a data packet at the head of the transmission queue has to wait before the contention in the channel is resolved by Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and then, is able to obtain access to the wireless channel and start transmission).

Even though IDAR employs cross-layer routing metrics (POS), this routing approach has some drawbacks, such as its failure to recognise the importance of load balancing between the gateways and mechanisms to reduce the routing overhead in large-scale WMN.

3.1.2 Routing Overhead

Traditional routing protocols and approaches usually cause a large overhead to provide information for the routing decision. In view of this, it is important to analyse the routing approaches from this perspective.

3.1.2.1 Quality of Service Routing in Wireless Mesh Networks

The QUality Of service RoUting (QUORUM) [Kone et al., 2007] is a routing protocol for WMN that deals with the problem of offering QoS guarantees to applications based on bandwidth, end-to-end delay constraints and robustness. This routing protocol enables the discovered routes to accept application requests for required bandwidth and delay bounds for a flow, and reject a flow if there is no route that satisfies these constraints. Thus, each intermediate node uses an admission control scheme during the route discovery phase to check whether the flow can be accepted or not.

QUORUM makes two contributions. First, it proposes a mechanism that predicts the end-to-end delay of a flow, since the routing control packets are significantly different from the normal data packets (size) and hence cannot provide predictions of delay as good as the data packets. As a result, the `DUMMY-RREP` packet is proposed for the route discovery; it is an imitation of real data packets, and has the same size, priority and data rate as a real data packet. Second, a robustness metric was defined to obtain the link quality and its utility in route selection was demonstrated. Each node in the network estimates the robustness of its links to its one-hop neighbors by measuring the number of control packets received during a rolling time window. Moreover, each node collects the bandwidth reserved at its one hop neighbors (piggybacked on periodic control packets) and stores it in its neighbour table.

The QUORUM protocol does not take into account interference to detect the link quality either. In addition, QUORUM employs a limited clustering solution to reduce the flooding of control routing messages, but the protocol did not show either the criteria for the cluster formation or the performance evaluation of the clustering that was implemented.

3.1.2.2 Routing in Large-Scale Wireless Mesh Networks Using Temperature Fields

The HEAT approach [Baumann et al., 2008] proposes an alternative routing paradigm to reduce the routing overheads. This paradigm is based on the physical laws that describe heat conduction. In this scheme, the field which represents the network is composed of the sources and surrounding particles which are equivalent to gateways and nodes, respectively. Nodes are assigned a temperature and heat conductor from the gateways to each other. The higher the temperature of a node, the closer it is to a gateway and the greater is the diversity of paths to this gateway.

Therefore, a route can be calculated as the path that follows the steepest gradient. In other words, packets are always forwarded to the neighboring node with the highest temperature and thus eventually reach a gateway. HEAT defines the routes by evaluating the temperature of its immediate neighbors in the network based on purely local information. Hence, flooding is not required. Nevertheless, the results from HEAT do not show a significant improvement of scalability when compared with the traditional routing protocols (AODV and OLSR). Furthermore, HEAT does not take into account interference and the congestion in the gateways.

3.1.2.3 Novel QoS-Aware Gateway Centralized Multi-hop Routing for Wireless Mesh Networks

The Gateway-Centralized Multi-hop Routing protocol (GCMR) [Zhao et al., 2010] proposes a central mechanism for routing management in which the gateway computes all the routing paths for mesh routers periodically. GCMR uses a number of techniques to extend the HWMP protocol and reduce its routing overhead. For example, the routing decision is centralized at the gateways and thus, it is not necessary to spread the routing information through the whole network. A leaf-to-gateway update mechanism assigns a limited number of leaf nodes that generate update packets and, mechanisms to change the interval between the routing messages and Time To Live (TTL) depending on the status of each node and the number of hops between leaf nodes and the gateway.

Although GCMR achieves a significant improvement in performance when it is compared to HWMP, it does not take into account the problem of interference. In addition, this approach does not include multiple gateways in the same WMN. Furthermore, there are some drawbacks when the gateway makes all the routing decisions. For example, it limits the communication between the mesh routers, since only the routes to reach the gateway are known. Moreover, a mesh router has to pass by the gateway if it needs to communicate with other mesh routers in the same WMN; as a result, the gateway tends to be congested and there is a lowering of performance, as in this case. For these reasons, the GCMR routing approach raises some serious performance difficulties.

3.1.2.4 Link-State Routing without Broadcast Storming for Multichannel Mesh Networks

The Multi-Channel Link-State Routing (MCLSR) [Kim et al., 2010] is a modified link-state routing protocol that is designed to lower broadcast overhead caused by a reduction of link-state propagation in a multi-channel WMN. The authors point out that the multi-channel communication requires high overhead in a link state routing strategy. MCLSR reduces the broadcast routing overheads (i.e. flooding) by employing a clustering scheme which selects a set of nodes (called clusterhead nodes) to aggregate the broadcast messages. There are two categories of nodes which are clusterheads and dependents. The main criterion to choose a node as a clusterhead is the node id. Some of the other nodes that are within one hop of this node become its dependents. Each dependent can have multiple neighboring clusterheads, but only one is identified as its master clusterhead. A clusterhead cannot have another clusterhead as its tight neighbor (if two clusterheads are tight neighbors of each other, one of them will lose the privilege of being a clusterhead according to node id). A clusterhead and its dependents form a cluster. A clusterhead collects and spreads the link-state information to its dependents.

The concepts of node link-state and cluster link-state are the internal control information units for control messages, such as hello messages (i.e. node link-state) and inter-clusterhead (i.e cluster link) unicast messages. A node link-state consists of node information that contains the state of the node, and a set of link qualities. A cluster link-state message consists of the link-states of all the dependents which are formed by concatenating the link-state structure of each dependent. The node and cluster link-state messages are periodically spread through the network. However, the cluster link-state message is expected to be less frequent than a node link-state message. Hence, it is similar to the Multi-Point Relays (MPR) proposed by the OLSR protocol, except for the fact that the MPRs only aggregate information from two hop-neighbors. Due to the fact that the performance evaluation compares MCLSR with a reactive routing protocol, it is difficult to assess if this routing approach is able to reduce the routing overhead of a link-state routing strategy while, at the same time, providing accurate routing information for the whole network.

Hence, the performance assessment of MCLSR should be compared with other link-state routing protocols, such as OLSR. Furthermore, the routing decision is unaware of interference and traffic load and it does not include any discussion of ways to avoid congested areas around the gateways in this routing protocol.

3.1.3 Congestion in the Gateways

Since gateways are potential bottlenecks in WMN, routing approaches should avoid the overload situations around this kind of node. Insight on solutions that address this problem are described next.

3.1.3.1 Efficient load-balancing routing for wireless mesh networks

The Configurable Access Network (CAN) [Bejerano et al., 2007] is an architecture for load-balancing routing in which the utilization of the network is maximized, while providing fairness and bandwidth guarantees. In this architecture, WMN are managed by a centralized and external station, called the Network Operation Center (NOC). One of the main tasks of a NOC is to map out the routes between the nodes and the gateway, while at the same time, allocating an appropriate bandwidth for each traffic flow. By performing a Breadth First Search (BFS) algorithm, the NOC discovers the network topology, layer by layer, and then communicates with a node by means of a source-routing scheme. CAN provides algorithms for single-path routing and bandwidth allocation that can achieve near-optimal fair bandwidth allocation without the drawbacks of multi-path routing.

Nevertheless, CAN does not consider interference, since it assumes that there are directional antennas and a sufficient number of wireless channels, which is not realistic in WMN. Hence, the use of NOC entities cause scalability problems in large-scale WMN due to its centralized nature.

3.1.3.2 A Load-Distributive QoS Routing Protocol for Multi-service Wireless Mesh Networks

The Load-Distributive QoS routing protocol (LDQ) [Khabazian and Aissa, 2010] is an extension of AODV and DSR protocols which proposes a load balancing approach in a distributed way. LDQ is mainly based on bandwidth reservation and bandwidth splitting mechanisms in the network layer, an enhanced distributed contention access mechanism in the MAC layer as well the integration of the mechanisms for the MAC and network layers. The bandwidth reservation is carried out by the control routing messages of this routing protocol according to the required data rate of each application session (i.e. data flow). The IEEE 802.11e is also combined with a bandwidth reservation mechanism to provide QoS in the MAC layer.

The bandwidth splitting mechanism is based on certain features which are combined to define the bandwidth that can be offered. This equation is applied for each intermediate mesh node in the candidate route. The main features are as follows: the priority of application traffic, history of traffic load of the intermediate mesh node as well as the total, and both the available and requested bandwidths. On the one hand, the proposed bandwidth splitting mechanism allows high-priority traffic (VoIP and video) to be in a privileged position to access network resources, which means that this kind of traffic has a better chance of reserving the total required bandwidth. On the other hand, low-priority traffic (FTP, HTTP, webmail) can be offered a lower fraction of the required bandwidth which increases the likelihood of it being split over distinct and large paths. In addition, LDQ forces the busy mesh routers to provide a lower fraction of their available bandwidth to the low-priority traffic as well as attaching less importance to higher requested bandwidths to avoid overloaded nodes.

Although, LDQ does not employ any mechanism to reduce routing overhead or delay in the route discovery when the number of nodes in the network increases, it adopts a load distributive approach based on the kind of application, this load balancing solution does not consider multiple gateways and thus, fails to prevent bottleneck zones around the gateways. Furthermore, the routing decision does not take into account interference, and it is based on the available bandwidth metric which cannot be quantified precisely because of the transient phenomena of the wireless networks.

3.1.4 Comparison of Related Works

Table 3.1 displays a comparison of the related works addressed in the previous sub-sections.

Most of the routing approaches, such as INSTC, IRMA and CAN, offer centralized solutions which may cause performance problems when the network size increases. Several of the routing approaches extend the traditional routing protocols from MANET, for instance DSCPT, ERM, IDAR, QUORUM, MCLSR and LQD. The routing overhead and congestion in the gateways are the factors which are least discussed in the routing solutions. Even the approaches which take the decrease of routing overhead into account do not achieve a significant reduction of

Table 3.1: Comparison among Related Work

Related Work	Interference	Routing Overhead	Congestion in the Gateways
INSTC	Yes	No	No
IRMA	Yes	No	No
DCSPT	Yes	No	No
CARMEN	Yes	No	No
ERM	Yes	No	No
IDAR	Yes	No	No
QUORUM	No	Yes	No
HEAT	No	Yes	No
GCMR	No	Yes	No
MCLSR	No	Yes	No
CAN	No	No	Yes
LDQ	No	No	Yes

overhead (e.g. HEAT and QUORUM). Furthermore, most of the routing approaches (e.g. INSTC, IRMA and DCSPT) which are concerned with the problem of interference are based on solutions that are not scalable, such as time-slot scheduling and channel assignment which are a scarce resource in the wireless networks. It should be stressed that not all the factors are dealt with simultaneously in the case of any of the analysed routing approaches in WMN. For example, although IDAR employs cross-layer information to make a routing decision, it does not propose any solution for overload situations at the gateways, while MCLSR handles the routing overhead through a modified flooding mechanism. However, MCLSR does not take into account interference when making the routing decision. For these reasons, in this thesis, an architectural model is proposed which comprises different solutions for these routing issues in WMN.

3.2 ***A***rchitecture of ***C***ooperative ***Ro***uting ***Ma***nagement - ***(ACRoMa)***

The overall objective of this thesis is to propose a modular architecture, called ACRoMa [Borges et al., 2012a], to improve scalability in a more complete way. This architecture employs the most effective means of managing the routing process in a way that allows a higher degree of traffic performance to be achieved.

3.2.1 Motivation

It should be pointed out that in attempting to enhance the routing process in WMN, the previous study which was discussed in Section 3.1 failed to combine all the open issues needed to improve scalability, in a single unified approach. In order to fill this gap, ACRoMa introduces three new components which are as follows: a clustering scheme, an inter-cluster load balancing routing algorithm, and a cross-layer routing metric. Figure 3.1 illustrates the interactions between the solutions (gray ring) employed in ACRoMa to resolve the issues (white ring) in the form of a ring.

Since the ACRoMa's components have complex synergies, an architectural model is employed to design an activity from the macro-level to the micro-level of this proposed solution.

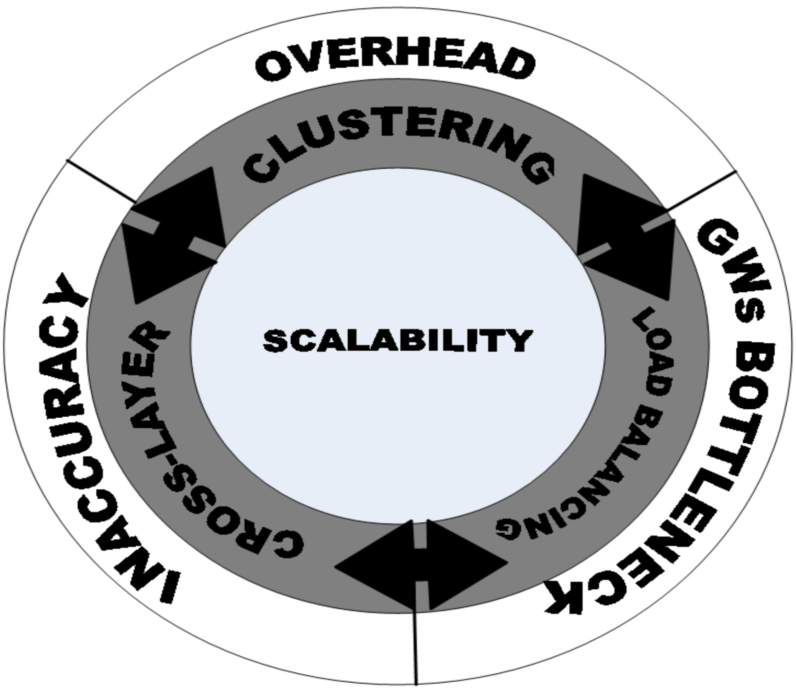

Figure 3.1: ACRoMa - Proposed Solutions and Open Issues

This integration improves the overall network performance through the routing process, by achieving a greater degree of traffic scalability and hence, an ability to handle growing amounts of traffic load and nodes in a skilful manner.

3.2.2 Architectural Model

Although a top-down approach is adopted to describe ACRoMa in this thesis, it was devised by means of a bottom-up approach which involves integration and testing; the components are integrated in an incremental way from the lowest to the highest level of components. In the light of this, each component was tested separately and then aggregated incrementally. The ACRoMa components are divided into two planes of routing management, which are: a) the topology and b) the process management. In the first plane, the clustering routing scheme belongs to the topology management plane. In the second plane, the process management employs the load balancing and link state routing algorithms and cross-layer routing metrics. Figure 3.2 illustrates the architectural model.

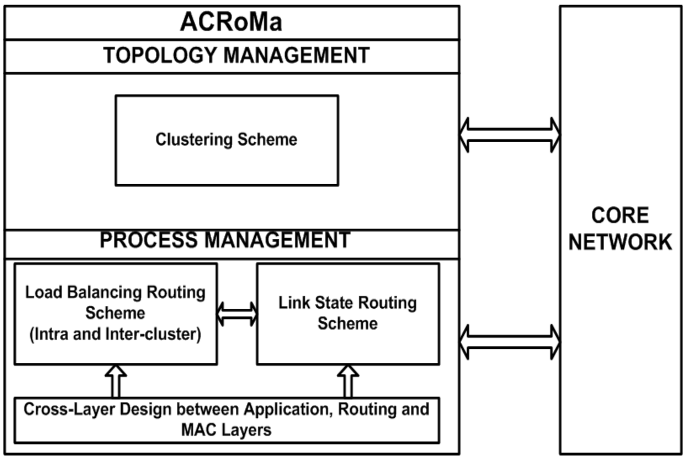

Figure 3.2: ACRoMa - Architectural Model

The main benefits of this architecture are as follows: the clustering scheme, called Collaborative CLustering Scheme (CoCLuS) [Borges et al., 2012c], which is employed in the network design phase in accordance with long-term requirements that are specified by a service provider. As a result of clustering, routing decisions become more accurate, due to the smaller scale of the region where cross-layer routing metrics are used. The main purpose of the CoCLuS is to provide a clustering structure that enables efficient inter-cluster load balancing routing. CoCLuS consists of new clustering elements and a hybrid routing scheme which is a combination of the load balancing routing and link-state routing schemes. With regard to the load balancing routing, the Routing Algorithm for Inter-cluster Load Balancing (RAILoB) [Borges et al., 2012c], which is used as the inter-cluster load balancing routing algorithm, spreads the traffic load between the multiple gateways in a distributed way. CoCLuS and RAILoB compose the clustering approach of the ACRoMa, which is called Clustering Approach for Routing MAnagement (CARMA) and it is used to describe the clustering solution proposed in this thesis more clearly. Metric for Interference and channel Diversity (MIND) [Borges et al., 2009] provides an accurate link quality to support the inter-cluster routing decision in the link state routing scheme, which depicts the interference and traffic load through the cross-layer design between the network, MAC and physical layers. A detailed description of the proposed clustering scheme, cross-layer routing metric and the inter-cluster load balancing routing algorithm will be given in Chapters 4 and 5.

Apart from the original components of ACRoMa, some mechanisms are employed from the related work to carry out the main objective of this architecture. For example, the C-OLSR routing protocol [Ros and Ruiz, 2007] was chosen as the routing protocol which is an extension of the OLSR protocol with support for the clustering. The OLSR routing protocol is commonly used in WMN, because it allows link state information to be disseminated efficiently [Genetzakis and Siris, 2008][Kim et al., 2010]. Furthermore, the Inner Domain Load Balancing (IDLB) algorithm [Choi and Han, 2008] is employed as the intra-cluster load balancing routing algorithm. This algorithm distributes the traffic load (i.e. by measuring the number of flows) among the routing sub-trees in which the gateway is the root.

3.3 Discussion

Conceptually, it is possible to compare ACRoMa against the other routing approaches in order to notice clearly how it overcomes them. In other words, ACRoMa improves the scalability of the WMN by providing the cooperation between three solutions that have taken into account the main factors that influence the scalability in WMN through the routing process, whereas the related work concerns only a part of these factors. The components considered cooperations to occur from different perspectives. For example, the load balancing routing algorithms are efficient solutions to provide a horizontal cooperation in the network layer between all the mesh nodes that improve the traffic scalability, where these nodes must have a collective awareness of the traffic load in the adjacent clusters (i.e. the nodes must share the information about the cluster traffic load with each other). In addition, a cross-layer design has been employed in WMN to exchange information between different layers; for instance interference and traffic load are picked up from the MAC and physical layers to support the routing decision. In this way, the cross-layer design allows a vertical cooperation in WMN where information from different layers is combined. Figure 3.3 shows these examples of cooperation.

The main synergies between the components are as follows: the cross-layer routing metric provides information which helps to make routing decisions, and the clustering approach provides a virtual routing structure for load balancing routing

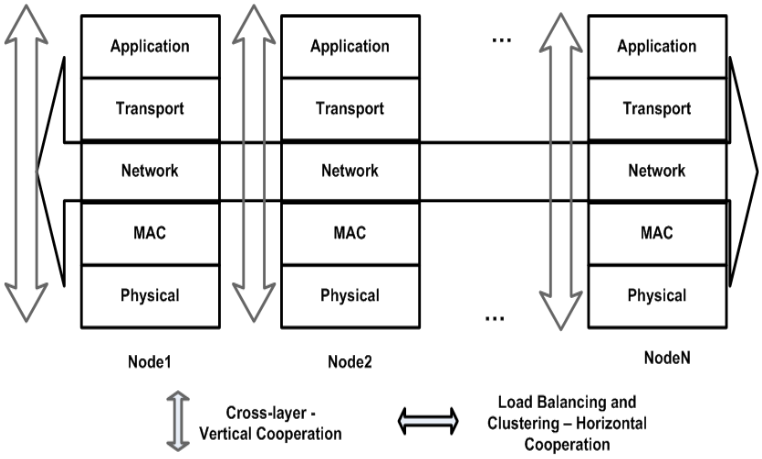

Figure 3.3: ACRoMa - Horizontal and Vertical Cooperation

algorithms, while reducing the routing overhead. Each component seeks to overcome the limitations found in its respective related work.

3.4 Summary

This chapter has conducted a survey of the most significant and recent routing approaches which have been found in the literature. The current routing approaches were analysed in the light of some of the most significant aspects to improving scalability in WMN. It was observed that none of the analysed routing approaches combines all these aspects at the same time. To fill this gap, an architectural model was examined which includes the described features required to provide a more scalable routing solution for WMN. This is the most important contribution made by this chapter.

Chapter 4

Cross-layer Routing Metrics

The traditional routing metrics which are employed in the MANET routing protocols, fail to depict the link quality in Wireless Mesh Networks (WMN), since these networks have different characteristics from MANET. In particular, there has been increasing interest in using information from MAC and PHY layers to make the routing decision more accurately. For example, interference and traffic load are important factors that influence the link quality in WMN. In the light of this, paths with a high interference level and traffic load must be avoided by using the local information to improve the network performance and thus, increase the scalability of WMN. However, the information about the interference and traffic load are not available in the network layer. In view of this, the solutions for WMN [Ashraf et al., 2011, Pal and Nasipuri, 2011] are not constrained by the same layer of traditional Open Systems Interconnection (OSI) specifications [ITU, 1994]. Thus, a design methodology is employed which proposes interactions between different layers of the protocol stack to achieve an overall performance optimization. This design methodology is called a cross-layer design and involves drawing on information which is shared between different layers to increase the adaptivity of the WMN approaches.

The cross-layer routing metrics will be described in this chapter which is structured as follows: an overview of cross-layer design is included in Section 4.1 with an emphasis on the main types of cross-layer approaches. Section 4.2 examines ideas regarding a proposed taxonomy for the classification of cross-layer routing metrics in WMN. A survey of several cross-layer routing metrics for WMN is conducted in Section 4.3. The MIND cross-layer routing metric which is designed for the architecture proposed in this thesis is described Section 4.4. Section 4.5 describes the simulation study that was performed and analyses the results obtained when the MIND metric is compared with the most relevant routing metrics. Section 4.6 concludes with a summary of this chapter.

4.1 Overview of Cross-layer Design

Due to the direct coupling between the different layers, the traditional Internet protocol stack does not provide a good performance in WMN, because of its inflexibility. It is necessary to consider different layers jointly to improve the overall performance, since they influence each other [Zhang and Zhang, 2008]. A new protocol layering has been devised that involves decomposing the overall network-performance optimization problem. This concept, called cross-layer design [Akyildiz and Wang, 2008], is based on an architecture in which different layers can exchange information to achieve an overall network performance. Thus, the cross-layer design enables a vertical cooperation in the protocol stack. Furthermore, recent studies have shown promising results that demonstrate the capacity of the cross-layer design to significantly improve the system performance in WMN [Tang et al., 2005, Wu et al., 2006, Anastasopoulos et al., 2007]. Nonetheless, it is important to be aware of a number of drawbacks in the cross-layer design, such as loss of protocol-layer abstraction, incompatibility with existing protocols, unforeseen effects on the future design of the network (i.e., difficult evolution capability), a more complex design and difficulties in maintenance and management. Thus, certain guidelines must be followed when carrying out the cross-layer design.

There are several typical examples of performance impact between the different layers, as shown in Figure 4.1. For example, congestion control can be carried out end-to-end in the transport layer or link-by-link in the MAC layer, while scheduling involves close interactions between MAC and physical layer. A scheduling algorithm determines the parameters for both MAC and the physical layers and depends on the congestion control to determine the best transmission rate. The interaction between congestion control and scheduling also involves the routing protocol. Hence, a well-defined joint optimization between congestion control and scheduling can enhance performance optimization in the layers, as well as application, transport, routing, MAC, and physical layers.

There are various combinations of cross-layer interactions between all the layers in the traditional protocol stack. Nevertheless, only the most significant interactions are discussed in this thesis. The main cross-layer interactions will be described in the following sub-sections.

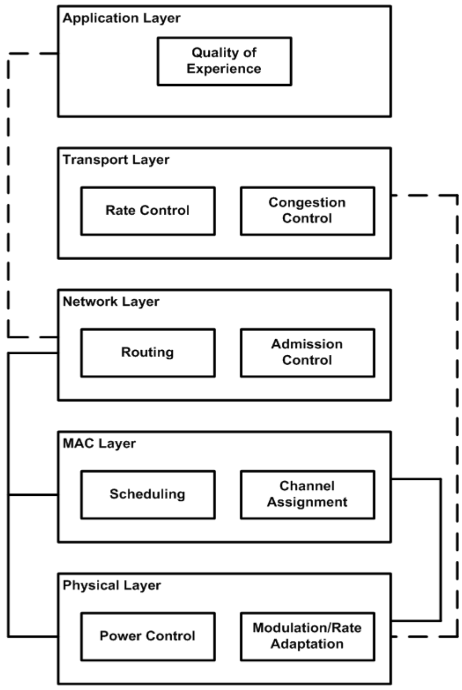

Figure 4.1: Cross-layer framework and interaction between the layers Adapted from [Zhang and Zhang, 2008]

4.1.1 Joint Optimization Algorithms Across Multiple-Protocol Layers

The design of the entire protocol stack can be formulated as an optimization problem, called full-optimization design and maps different protocol layers in the clean-slate protocol architecture (i.e., a protocol architecture that is quite different from the existing traditional protocol stack). However, it may not exactly match an existing protocol stack such as the Internet, because of compatibility problems. This difficulty can be overcome by formulating an optimization solution that considers the existing protocol architecture and is called a suboptimization design or optimization across multiple-protocol layers. This solution provides interactions between all the protocol layers ranging from the application to the physical layer. However, a cross-layer design with multiple layers is very complex and difficult to achieve and furthermore, the interaction between two or three different layers can be enough to provide the performance required [Akyildiz and Wang, 2008]. For this reason, the main cross-layer design approaches between two or three layers will be described in the next sections.

4.1.2 Interaction between Transport and Physical Layers

Due to the variable capacity of a wireless link [Lacage et al., 2004], the transport-layer protocols (e.g., Transmission Control Protocol (TCP) [Postel, 1981] and User Datagram Protocol (UDP) [Postel, 1980]) need to be optimized to achieve a better performance and result in a cross-layer design between the transport and physical-layer protocols. On the one hand, in the UDP protocol, a source node does not take into account the intermediate nodes and wireless links from itself to the destination node when it employs a transmission rate. Hence, the source rate must be regulated by other mechanisms (e.g., connection admission control or end-to-end rate control) to improve its performance. On the other hand, in the TCP protocol, the changes in the transmission rate of the source node depend on the status of congestion in the wireless link.

The interaction between TCP and physical layers can be classified in two categories. In the first category, the TCP congestion-control algorithm uses the information from the physical layer such as packet loss, delay in the queues and signals, to adjust the parameters of the congestion-control algorithm. For instance, TCP Vegas [Brakmo and Peterson, 1995] uses delay as a signal of congestion. In the second category, the physical layer and TCP interaction can be jointly needed in both directions [Chiang, 2005]. In other words, it is not only to change the TCP parameters, but the physical layer also modifies its parameters such as transmit power, coding and modulation, to avoid congestion.

4.1.3 Interaction between MAC and Physical Layers

The cross-layer design between MAC and physical layers is usually employed in wireless networks. The interaction between MAC and physical layers occurs on the same interface card or even on the same chipset. The advanced physical layer techniques have improved the physical layer to enable it to support more optimized cross-layer design and thus, increase the network performance. For example, the transmission rate of a wireless link can be significantly enhanced by means of multiple coding and modulation schemes with advanced antenna techniques (e.g., directional antennas and smart antennas), Multiple-Input and Multiple-Output (MIMO) [Oestges and Clerckx, 2007] and Orthogonal Frequency Division Multiplexing (OFDM) [Nee and Prasad, 2000] which is based on multiple antenas for radio signals transmission, reception and advanced signal-processing techniques.

4.1.4 Interaction between the Network, MAC and Physical Layers

A routing algorithm of a multi-hop wireless network selects a path for any packet from its source to destination. Usually, it only needs to consider connectivity among nodes to select its paths. However, other routing metrics are required to enhance the performance. For instance, the routing algorithm should take into account the interference level and traffic load information to determine the best path. However, these kinds of information are not available in the network layer in a traditional protocol architecture and as a result, these types of layered design approaches still only lead to a sub-optimal performance. Hence, the interaction between network, MAC and physical layers is a useful mechanism in providing information about the link quality for the routing process [Nguyen et al., 2008, Islam et al., 2010]. For this reason, this is the kind of interaction that is highlighted in this thesis.

4.1.5 Interaction between Application and Network Layers

The quality of multimedia applications has been mainly assessed through the Quality of Experience (QoE) which represents the subjective perceptions of end users when they are using network services. Hence, QoE is closely related to the application layer. For this reason, the cross-layer interaction with the application layer has also been investigated in WMN. In this context, both the network and application layers can change their policies in order to improve the service quality. For example, this can involve a routing algorithm that is based on the dynamic choice of routing metrics to calculate the best routes for a specific multimedia application [Gomes et al., 2011] as well as adaptive algorithms that can make decisions to drop some video data packets and VoIP application [Rodrigues et al., 2011]. Indeed, it is worth noting that the cross-layer interaction can occur in two hierarchies in the second example. In other words, in the network, the MAC and PHY layers interact with each other to provide the cross-layer routing metric which is used by the adaptive algorithm.

4.2 Taxonomy for Cross-layer Routing Metrics

Cross-layer routing metrics are a combination of several components that depict different characteristics of the links in the WMN. In previous surveys of routing metrics for WMN [Campista et al., 2008, Liu and Liao, 2008, Guerin et al., 2007, Waharte et al., 2008a, Baumann et al., 2007] there has been a lack of any description of the measurements used in routing metrics and a failure to investigate important issues such as information gathering methods and stability mechanisms. Thus, we are setting out a new taxonomy that provides an in-depth understanding of the main features of cross-layer routing metrics.

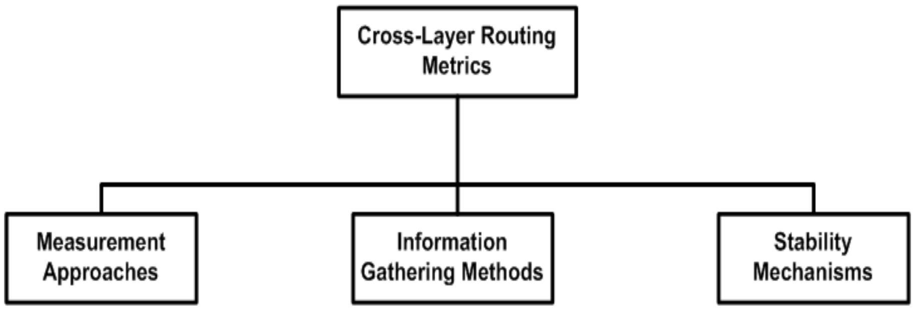

Figure 4.2: Elements of Cross-layer Routing Metrics

Figure 4.2 shows the suggested taxonomy, which comprises the following categories: information-gathering methods (Sub-section 4.2.1), measurements (Sub-section 4.2.2) and stability mechanisms (Sub-section 4.2.3).

4.2.1 Information Gathering Methods

The information gathering methods are the mechanisms to obtain the measurements and must be analysed to understand how the cross-layer routing metrics are implemented in practice. Accuracy and overhead are the main factors that determine the choice of the method employed to obtain the measurements. The methods used to acquire information are node-related, passive monitoring, piggy-back probing and active probing, and are listed as follows [Baumann et al., 2007, Chen et al., 2009]:

- **Node-related.** The measurements are acquired from the nodes, and include fixed, configurable or variable values, such as the number of interfaces of a node, the communication costs and the length of the input and output queues.

- **Passive Monitoring.** This method, which does not cause overhead in the network, is the most employed to collect the cross-layer measurements. Measurements, such as those of the interference and traffic load, are obtained through the traffic that is arriving at and leaving a node. However, in some situations, the passive monitoring can not be employed to capture some parameters. For instance, when the wireless card drivers do not provide adequate capabilities to measure transmission rate, passive measurement is not feasible. Furthermore, the passive monitoring can gather inconsistent measurements when there is a small amount of data, even if the control routing packets are used; the reason for this is that control packets are small in size and thus, fail to depict the conditions of the channel. There are also some measurements that are difficult to gather from the lower layers when using passive monitoring. These measurements will be discussed in sub-section 4.2.2.

- **Active Probing.** With this method, specific packets are generated and included in the traffic to monitor the link characteristics. The first drawback of active probing is that overhead is introduced in the network. Moreover, the active probing can also gather inaccurate cross-layer information, due to the intermittent nature of the wireless medium, as occasional losses of probe packets can occur and lead to an overestimation of the link quality. However, active probing can be used as a short cut to overcome the inability of some drivers to share useful information such as the transmission rate, since this is not provided by some wireless network card drivers [Draves et al., 2004b]. This limitation can be overcome if a driver which shares the required information is chosen. However, since the WMN usually employ the COTS equipments, some wireless network cards may not be able to share the cross-layer information needed in the routing metrics. Hence, the active probing is an alternative and generic approach to obtain information indirectly from the lower layers and thus, it was considered to be a method to gather information for the cross-layer routing metrics. Nonetheless, there is an investigation [Zhang and Davis, 2008] that suggests correlation between active probing and passive monitoring methods to measure link quality precisely, being interchangeably according to the network load. For example, active probing and passive monitoring can be employed when the traffic load is low and high, respectively.

4.2.2 Measurements

The measurements include the key factors involved in the design of the cross-layer routing metrics. They consist of four main categories as shown in Figure 4.3: (a) basic, (b) interference, (c) load and (d) hybrid measurements.

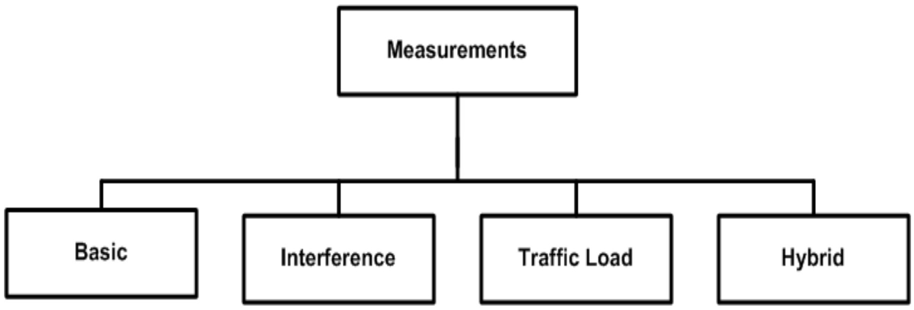

Figure 4.3: Elements of the Sub-taxonomy Measurements

4.2.2.1 Basic measurements

These measurements are called basic because they describe the factors that directly influence the traditional performance parameters (e.g., throughput, delay and packet loss ratio). In addition, the existing cross-layer routing metrics have employed at least one of these measurements.

- **Path Length.** Path length can be defined as the number of links that a path has between the source and the destination nodes. The concept of the path length measurement is quite simple, since it is based on the topological information of the network. On the one hand, the path length achieves useful results when it is used in mobile wireless networks, since it reacts quickly to the topological changes. On the other hand, this measurement does not result in an optimal performance in static wireless networks (e.g., the infrastructure architecture of the WMN). Since shorter paths (in terms of the number of hops) usually correspond to the paths between the nodes that have a higher transmission range, a lower transmission rate will be achieved [Zhao et al., 2005]. Although the path length does not take into account measurements such as interference and traffic load, it should also be treated as a complementary routing metric because of the need to avoid longer paths which also lead to the degradation of the application performance.

- **Transmission Rate.** This measurement shows the amount of data which can be sent over a link within a given time. The routing metrics should take into account the transmission rates of the wireless links, because a node may use different transmission rates to different neighbours. However, drivers of some wireless network cards fail to provide this information. In addition, the transmission rate does not represent the actual link capacity, since it is affected by the interference and traffic load. Hence, the transmission rate should be combined with other measurements to display the quality of the link in a more satisfactory way.

- **Packet Loss Rate.** The loss rate can be calculated either as the complementary value of the delivery rate or as the number of packet retransmissions needed to deliver a packet. This measurement might indicate interference or the poor quality of the channel owing to the fact that the number of retransmissions depends on the interference that is caused both by competing traffic, and the interference stemming from the same data flow on the physical layer. However, the packet loss rate is not the most precise means of measuring the level of interference that is picked up. Furthermore, as it is unaware of the extent of the traffic load, this measurement is a less sensitive way of picking up congested areas, and thus results in an inferior performance when used for high traffic loads.

- **Delay.** Delay is the total amount of time spent by a packet travelling between the source and the destination. Delay can be subdivided into four different components: queuing, processing, transmission and propagation. Routing metrics usually employ active probing to estimate the delay which means that the transmission rate may be implicitly captured. Although delay can pick up the link quality better than the transmission and loss rate, this metric overestimates the link quality, as it is not a very precise way of measuring interference and traffic load. In other words, due to the intermittent nature of the wireless medium, occasional losses of probe packets can occur and lead to an overestimation of the link quality.

4.2.2.2 Interference Measurements

The MAC and physical layers of the wireless networks are far more complex than in wired networks. For example, a wireless link does not have a dedicated bandwidth; the reason for this is that several nodes share the wireless medium, and hence, the neighbouring node transmissions may compete for the same bandwidth, and interfere with the transmissions on the other links. Furthermore, the complexity of these layers tends to increase with multi-channel multi-radio capacity, particularly in the case of WMN. For example, assigning non-overlapping channels for each radio interface to minimize interference is a complicated issue due to the restricted number of channels in the current 802.11 IEEE standard. This means that to select paths that satisfy the requirements of the network applications, the routing process must be aware of the link quality so that it can capture the interference between the neighbouring nodes as well as taking advantage of the multi-channel multi-radio capability. The sub-taxonomy of interference measurements consists of two categories [Yang et al., 2005a] in multi-channel multi-radio WMN and is illustrated in Figure 4.4.

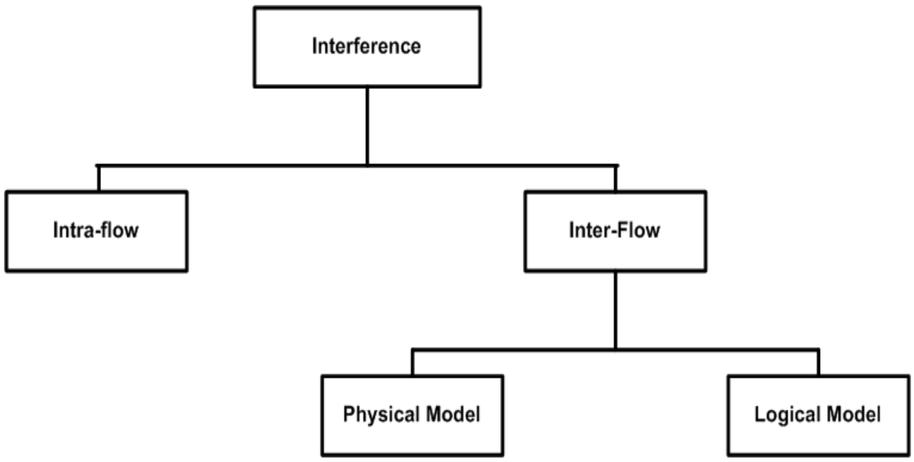

Figure 4.4: Elements of the Interference Subtaxonomy

The inter-flow and intra-flow interference types are shown in Figures 4.5(a) and 4.5(b), respectively. Suppose r is a candidate route ($A \rightarrow B \rightarrow C$) for the connection request. Inter-flow interference is caused by the interference between a wireless link on r ($B \rightarrow C$) and a wireless link which is not on r ($E \rightarrow F$), assuming that B and E use the same channel. It is worth noting that inter-flow interference can also be caused by external networks (e.g., bluetooth) that works in the same frequency.

Suppose p is another candidate route ($A \rightarrow D \rightarrow C$). Intra-flow interference is the contention caused by the interference between two wireless links on p, ($A \rightarrow D$) and ($D \rightarrow C$). The routing metrics should take into consideration both the inter-flow and intra-flow interference in WMN with MCMR capability.

A. Intra-flow Interference Measurements

The intra-flow interference is avoided or reduced by giving more weight to paths with lower channel diversity. In the light of this, two measurements have been used in the routing metrics, namely max_X and the **CSC**, as follows:

- max_X**.** This is the maximum sum of the routing metric for links on the same channel. The routing metric needs global information to obtain the intra-flow interference when it uses this measurement. However, it is of little use, since a

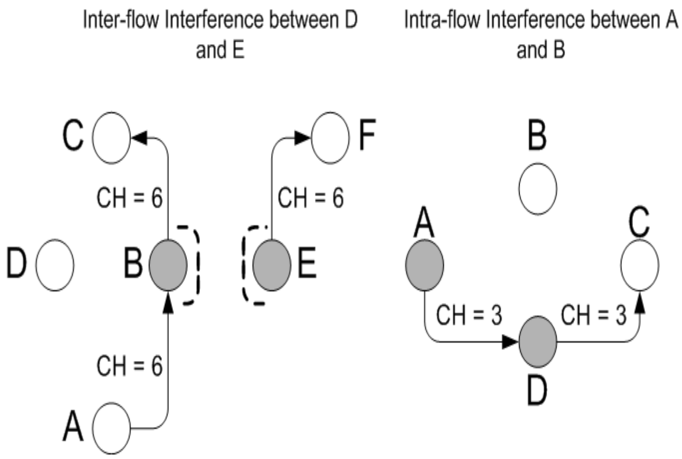

Figure 4.5: (a) Inter-flow and (b) Intra-flow Interfence

node usually does not interfere with other nodes that are more than two hops away even if they share the same channel [Yang et al., 2005a].

- **Channel Switching Cost (CSC).** This reduces the intra-flow interference by drawing on local information, and gives higher weights to paths with consecutive links that use the same channel. Nevertheless, this measurement requires a complex mechanism (i.e., virtual network) to be implemented in the routing metric and this mechanism is also impracticable. A detailed description of this mechanism will be given in the next section.

B. Inter-flow Interference measurements

There are three interference models that have been studied in the literature to pick up the inter-flow interference: the protocol [Gupta and Kumar, 2000], logical [Chen et al., 2009] and physical [Jain et al., 2005] interference models. All these models are influenced by the concept of transmission and interference ranges. On the one hand, the transmission range is the maximum range where a radio frequency signal can be correctly received. On the other hand, the interference range defines the area where a sending node can disturb the transmission from a third node. Figure 4.6 shows an example of interference and transmission ranges of node D, where the solid lines denote the valid transmission range and the dotted lines indicate the interference range. The interference range is difficult to foresee, since it can change quite often [Beuster et al., 2008].

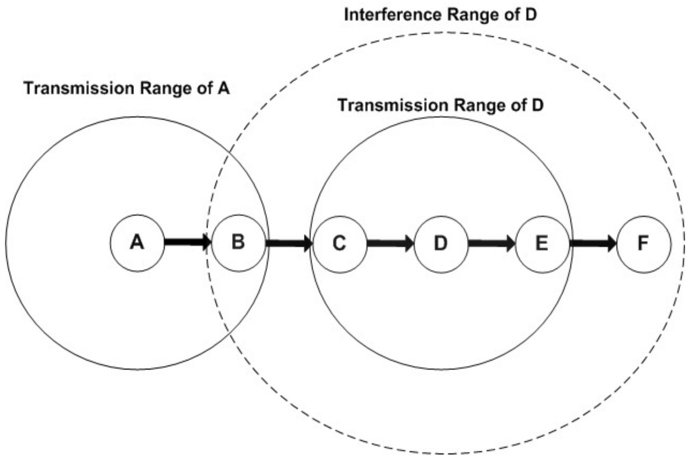

Figure 4.6: Interference and Transmission Ranges

The protocol interference model determines that a transmission from a node X to a node Y is successful if (i) there exists a link between them in the network topology (Figure 4.7) which is used for the transmission; and (ii) any node Z such that $d_{ZY} \leq R$ or $d_{ZX} \leq R$ is neither transmitting nor receiving in the channel used by X and Y. d_{ZY} represents the distance between nodes Z and Y, and R represents the interference range. This model is very strict, since it is designed to guarantee that the links do not interfere with each other through the particular channels assigned for each one and thus, this model becomes an optimum case where the interference is completely avoided. In the light of this, channel assignment algorithms have been adopted in multi-channel multi-radio WMN when assigning the available channels to the radio interfaces of the mesh routers, to minimize interference [Crichigno et al., 2008]. However, the restricted number of available channels in IEEE physical specification [IEEE, 1999] does not allow one channel to be assigned to each wireless link in the WMN. In other words, this specification allows the simultaneous operation of three non-overlapping channels in the 2.4 GHz band and 12 non-overlapping channels in the 5 GHz band. This means that the channels are assigned between the links in a repetitive way, which causes interference. For these reasons, the protocol model is not employed in the routing metrics to show the interference.

The logical interference model also takes into account the interference in the MAC layer. The interference in the MAC layer is also known as channel contention interference [Genetzakis and Siris, 2008], because it stems from the medium access protocol (e.g., Carrier Sense Multiple Access with Collision Avoidance - CSMA/CA) which requires the station to wait until the channel is free before starting the transmission, once that the shared channel may be occupied by transmissions from other nodes that are using the same channel within the interference range. Hence, this

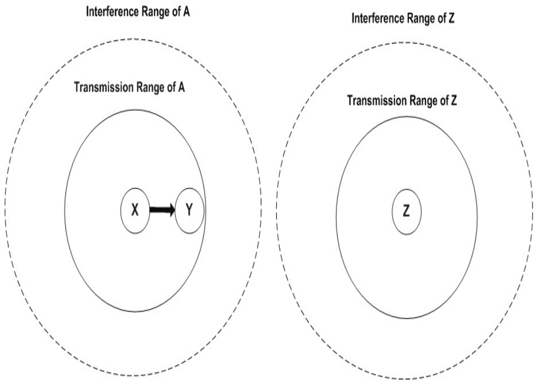

Figure 4.7: Example of a Protocol Interference Model

model incorporates the deferred access time to the wireless medium. The logical model is less restrictive than the protocol model.

The physical interference model captures the interference experienced by the wireless links in the physical layer. The physical interference is caused by superposition of waves which changes the original signal and causes bit alterations. As a result of this, the packets may be dropped. In this model, a communication between nodes X and Y (Figure 4.7) is successful if the signal strength at the receiver Y is above a certain threshold and this depends on the desired transmission characteristics, such as the channel and data rate. The physical model is less restrictive compared to the protocol and logical models, since it only depends on the signal strength values, such as the Signal-to-noise Ratio (SNR) and Signal to Interference plus Noise Ratio (SINR). The physical model has the advantage of depicting measurements by using on-line data traffic. For these reasons, the cross-layer routing metrics have generally used the physical interference model to measure the rate of inter-flow interference. Nonetheless, signal strength is difficult to obtain in an accurate way, since it has a high variation.

In summary, the protocol and the logical interference models form the interference which occurs before transmission. On the other hand, the physical interference model displays the actual transmission of the packet when the interfering signals may cause failed transmissions. Hence, a cross-layer routing metric must take full account of the measurements from both models to depict the interference accurately. The most relevant inter-flow interference measurements are set out in this sub-section.

Some routing metrics pick up the inter-flow interference by employing the logical model. In this particular case, the main measuring device is the Number of Interfering Neighbours (NIN). The smaller the NIN the better, since there is less probability of channel contention because of the relatively small number of interfering neighbours seeking to access the wireless medium. There is a risk that NIN can overestimate the interference levels, because, in showing the extent of interference, it tends to adopt a worst-case scenario approach, where all the neighboring nodes are transmitting packets. However, some of the interfering neighbours may not transmit traffic at a specific time and consequently, the wireless medium may not have a precise interference level.

To overcome the limitations of the NIN measure to depict the inter-flow interference, cross-layer routing metrics have employed measurements based on the physical model. There are several measurements that are used to measure the inter-flow interference based on the physical layer [Vlavianos et al., 2008, Olszewski, 2007], namely, Received Signal Strength Indication (RSSI), SINR, Bit Error Rate (BER) and Frame Error Rate (FER), as follows:

- **RSSI.** RSSI shows the signal strength observed on the receiver's antenna during packet reception. RSSI values vary from R_{max} (i.e., the maximum value of RSSI) and depend on the chipset of the wireless card. Thus, the specifications of each card provide a different formula to convert the RSSI values to power. Furthermore, the RSSI calculation is based on a packet (e.g., packet p) that was received correctly and thus, if the packet of the next computation (e.g., packet q) fails to operate because of interference, the RSSI will not be recorded. Consequently, the RSSI value computed from the packet p is retained and the interference that causes the loss of packet q is not included in the calculation. This means that the RSSI does not depict the interference in the link in an accurate way.

- **SINR.** This captures the power of the received signal that exceeds the sum of noise plus interference at the receiver. Recently, SINR has been regarded as the most appropriate metric to depict the link quality [Manikantan Shila and Anjali, 2008, Borges et al., 2009]. However, the commercial wireless cards do not usually record this measurement. If the commercial wireless card does not provide the SINR value, the SINR could be estimated on the basis of RSSI [Reis et al., 2006, Ares et al., 2007]. Nonetheless, this value of SINR has acquired all the failings of RSSI and, as a result, this value will not be precise.

- **BER.** BER is the ratio between the number of bits with errors and the total number of bits that have been received over a specific time period. In other words, it is a fine-grained metric to measure interference. Hence, it is not simple to measure the BER precisely in real systems, since in making the BER measurements a pseudo-random data sequence transmission has to be taken into consideration. As a result, this approach is of little value when the network conditions are changing quickly over a period of time. Furthermore, BER computation introduces significant overhead, since it requires the processing of a large amount of previously known data.

- **FER.** FER measurements are computed by averaging frame error indicator bits output with the aid of a cyclic redundancy check decoder [Olszewski, 2007]. As in the case of BER, FER also requires the processing of an amount of previously-known data. In addition, it requires repeated computations over extended periods of time in order to provide a more reliable value and thus, FER also takes a long time to capture the interference. Nonetheless, in measuring interference, FER is a coarser-grained measurement than BER, since it takes into account the frame rather than the bits and as a result is simpler to implement than BER. The routing metrics that use BER or FER to measure the interference, should draw on other information so that the interference can be measured in a more accurate way.

4.2.2.3 Measurements of Traffic Load

On the one hand, as is well known, the traffic tends mainly to flow either towards or away from the gateways. In this context, a routing metric should be aware of the traffic load to avoid heavily congested paths around the gateways. On the other hand, if multiple mesh routers select the path with the lowest load to route their traffic to the gateways, the load of this path will increase significantly and thus reduce the overall performance of the network. Hence, a routing metric should depict the load in a way that allows the routing algorithm to calculate paths that provide load balancing between the gateways and thus, improve the capacity of the network by avoiding congested areas [Nguyen et al., 2008, Karrer and Pescape, 2007]. The main measurements used to depict the traffic load in the cross-layer routing metrics consist of the number of flows and queue size and are described below:

- **Number of Flows**. The traffic load can be calculated through the number of flows that are currently being transmitted in the node. It is worth noting that there is a risk that this measurement might overestimate or underestimate the traffic load if it only measurements the number of flows and fails to take into account the data rate of each flow. Furthermore, the number of flows and their data rates may be hard to quantify.

- **Queue Size**. This measurement is employed to depict the traffic load in routing metrics. The queue size is a more precise traffic load measurement than the number of flows, since it is separate from the data rate of the flows.

4.2.2.4 Hybrid Measurements

Hybrid measurements usually depict interference and traffic load together. The hybrid measurements are described as follows:

- **Sum of the Delay of Interfering Links**. This measurement depicts the influence of the amount of traffic generated by the interfering links based on delay. However, as is well known, delay can overestimate the link quality and thus, it is not a precise enough measurement to depict the traffic load and interference.

- **Sum of the Queue length of Interfering Nodes**. This measurement picks up the influence of the interfering traffic more accurately than the previous one, because the queue length shows the interference and traffic load more precisely than delay. As mentioned earlier in sub-section 4.2.2.1, delay can tend to overestimate the degree of the interference and traffic load.

- **Channel Busy Time (CBT).** CBT consists of times spent in distinct states, such as *transmit* (i.e., the time that the node spends in transmitting to other nodes), *receive* (i.e., time when the node is receiving packets from other nodes), *occupied* (i.e., when the node senses the medium is busy because there is transmission from other nodes) and *backoff* (i.e. when the node has some data to send but finds the medium is busy when it tries to transmit it. As a result, the IEEE 802.11 protocol forces the node to wait for a random period of time before trying the transmission again). There is also an *idle* state that represents the time when the node senses that the medium is idle and the node has no data to transmit. This state can be used as a complementary function to calculate the CBT. In addition, the CBT can be regarded as a more precise means of measuring the logical interference than the other measurements discussed earlier, because it is able to pick up the precise time of the channel contention, i.e. the time that a node spends transmitting on the channel. CBT is more accurate than the number of flows in the link, since it does not depend on the data rates of flows. Furthermore, CBT is also considered as the most precise means of measuring the utilization of channels in wireless networks [Athanasiou et al., 2009, Wu and cker Chiueh, 2007]. Nonetheless, the CBT is implemented usually takes into account both the transmission time of the data and the control packets. In computing a more accurate traffic load, the calculation should only take into account the transmission time of the data packets.

4.2.3 Stability Mechanisms

Once a link is recognized as being of a better quality, it attracts a lot of traffic and, as a result, the link can become congested. In other words, as traffic starts to route around this link, its metric value increases and this effect can be propagated to neighbouring links as well. This issue, the so-called *self-interference* [Baumann et al., 2007] causes routing oscillations that degrade the application performance. Moreover, this kind of behaviour tends to increase with the measurements of the wireless networks, that are subject to high variation, such as interference and traffic load measurements.

The level of oscillation may not only depend on the principal characteristics of a metric but also on the implementation details (e.g., information gathering methods). Although the metrics should represent the current state of the network, mechanisms are required that can lessen the metric weights that are based on measurements with high variation values and thus reduce the route oscillation. In this context, some stability mechanisms have been suggested to smooth out the value of a routing metric and thus avoid unnecessary route oscillations. The mechanisms can be employed in the routing metric values as well as in their measurements (i.e., one distinct mechanism for each measurement). The sub-taxonomy of stability mechanisms is shown in Figure 4.8. This sub-taxonomy consists of statistical functions and updated propagation threshold elements.

The statistical functions aim to reduce the excessive sensitivity of the routing metrics to small changes. The main statistical functions are described as follows:

- **Fixed History Window (FHW).** An average figure is calculated from a fixed number of previous measurements or from the measurements in a time interval. This mechanism smooths out the values, but it might not depict the actual network conditions in an appropriate way.

- **Dynamic History Window (DHW).** An average figure is calculated from

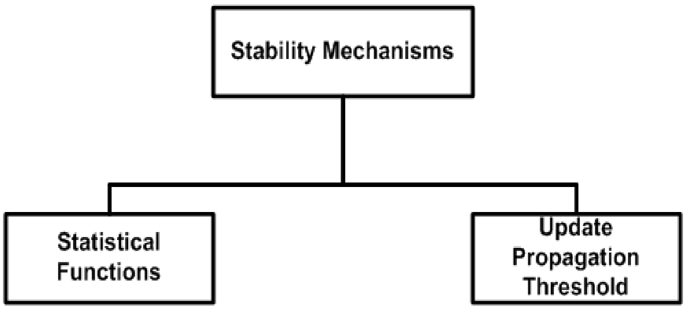

Figure 4.8: Elements of the Stability Mechanisms

a variable number of previous measurements or from the measurements in a time interval. The size of the window might depend on particular factors, such as the higher the network traffic, the smaller the history window. In view of this, this mechanism reveals the state of the current network at a faster rate than fixed history window.

- **Exponential Weighting Moving Average (EWMA).** EWMA gives more weight to recent measurements while not entirely discarding older ones. Thus, the weighting for each older measurement decreases exponentially. In addition, this mechanism depicts the state of the network more accurately than previous mechanisms. For example, delay can be smoothed out as follows:

$$d_{smooth} = \alpha \times d_{current} + (1 - \alpha) \times d_{new} \tag{4.1}$$

where $\alpha \in [0,\ 1]$ is the weighting factor, d_{new} is the new delay value, $d_{current}$ the current delay value and d_{smooth} is the delay calculated value.

The Update Propagation Threshold (UPT) element has a single stability mechanism in which the routing protocol ignores insignificant changes in the end-to-end routing metric weight. This solution defines a threshold so as to identify a "significant" change, for instance, Ramachandran et al. [Ramachandran et al., 2007] set a threshold of 10% difference (better/worse) between the current and previous routing metric value to bring about this change. As a result, this approach improves route stability by reducing any insignificant switching of routes, and is thus able to decrease the overhead of the routing protocols. However, defining a precise threshold is a complex issue. The statistical functions and updated propagation threshold are complementary mechanisms that can be used together to improve the stability of overall routing.

4.3 Related Work on Cross-layer Routing Metrics

A large number of routing metrics for WMN have been recommended in the literature. This section describes the most recent and relevant routing metrics for WMN. The routing metrics will be grouped according to the measurement sub-taxonomy set out in the previous section. It is worth noting that each routing metric is mapped in a group which is matched by the main measure of the metric. Within each group, the routing metrics will be examined in accordance with the timeline of the research path in this area. The routing metrics for WMN have followed four main trends, which are: basic, interference-aware, load-aware and a combination of interference-aware and load-aware routing metrics, as outlined in Sub-sections 4.3.1, 4.3.2, 4.3.3 and 4.3.4, respectively.

4.3.1 Basic Routing Metrics

In a step forward from the hop count routing metric, new routing metrics were designed that take into consideration the packet losses and delay. The following sub-sections examine the most relevant metrics in this group.

4.3.1.1 Expected Transmission Count - ETX

ETX [Couto et al., 2003] can be defined as the expected number of MAC layer transmissions that is needed to successfully deliver a packet through a wireless link, including retransmissions. The weight of a route can be defined as the total sum of the ETX of all the links along the route. In other words, this metric comprises both packet loss ratio and route length. The ETX of a link is calculated on the basis of the forward and reverse delivery ratios of the link. The forward delivery ratio, d_f, is the measured probability that a data packet will successfully reach the receiver; the reverse delivery ratio, d_r, is the probability that the ACK packet has been successfully received. These delivery ratios can be measured in a way that is described in Equation 4.2. The expected probability that a transmission has been successfully received and acknowledged is $(d_f \cdot d_r)$.

$$ETX_i = \frac{1}{d_f \cdot d_r} \tag{4.2}$$

The delivery ratios d_f and d_r are measured with the aid of broadcast probe packets. Each node broadcasts link probes of a fixed size (i.e. 134 bytes), at an average period τ (i.e., every second). Every node knows the number of probe packets it has received during the last w seconds and consequently, it can calculate the delivery ratio from the sender at any time as follows:

$$r(t) = \frac{count\,(t - w, t)}{w/\tau} \tag{4.3}$$

where $count\,(t - w, t)$ is the number of probes received during the window w, and w/τ is the number of probes that should have been received.

Suppose there is a link between the V and U nodes. Equation 4.3 allows node V to measure d_r, and node U to measure d_f. Because U knows it should receive a probe from V every τ seconds. Hence, ETX is calculated based on the expected number of probes and the actual number of received probes.

ETX is based on delivery ratios, which directly affect throughput and accounts for the effects of link loss ratios and asymmetry in the loss ratio in both directions of each link. It prefers paths with higher throughput and a lower number of hops. However, it is a routing metric for a single-channel multihop wireless network, insofar as it does not discriminate between same channel paths and channel-diverse paths. Hence, it does not depict the extent of intra-flow interference. This metric does not explicitly capture the inter-flow interference experienced by the links, which indeed significantly has an impact on the link capacity and the data rate at which the packets are transmitted over each link. It can only detect inter-flow interference indirectly, since the high level of interference will probably have a higher rate of packet loss. Moreover, it fails to take into account differences in transmission rates.

As the transmission rate of the probe packets is typically low, it does not reflect how busy a link is. In other words, it does not allow the traffic load to be entirely caught on the transmission because the sender of a probe packet can defer its transmission if it senses that the channel is busy. Hence, this metrics fails to provide load balancing. Moreover, the active probing method used in this metric introduces inaccuracies on the estimation of the loss rate measurement. These inaccuracies are caused by the different sizes of probe and ACK packets when compared with data packets which causes underestimation and overestimation of loss rate, respectively. Furthermore, due to its lack of knowledge of interference and different transmission rates [Draves et al., 2004a], ETX can result in paths with poor quality (i.e., high level of interference and lower transmission rate) that spend more time to transmit data and consequently, neighbouring nodes are forced to back off from their own transmissions resulting in high contention levels (i.e., logical interference).

Although the experimental results show that ETX performs better than hop count metric under static network conditions, it may perform poorly under highly variable channel conditions and in burst-loss situations. To overcome this limitation, the use of Modified Expected Number of Transmissions (mETX) and the Effective Number of Transmissions (ENT) was recommended in [Koksal et al., 2006]. These routing metrics estimate the losses by means of the bit error probability rather than the packet error probability. The main difference between mETX and ENT is that the latter has a configurable parameter that can be attributed to a tolerable loss rate. These metrics employ similar probing mechanism to ETX (i.e., link-layer probe packets) to estimate the channel parameters. In other words, the data record is at the bit level for each probe packet, rather than at the packet level. The mETX and ENT compute the bit error probability using the position of the corrupted bit in the probe packet and the dependence of these bit errors throughout successive transmissions, since the packet probes are composed by a previously known sequence of bits. However, these metrics have a drawback insofar as they can be impracticable because the packets received with errors in the MAC layer are discarded, without signaling the fact to the higher layers in most of the implementations. Furthermore, the verification of the bit errors is a very complex task.

4.3.1.2 Expected Transmission Time - ETT

The ETT metric [Draves et al., 2004b] predicts the total amount of time it would take to send a data packet along a route, while taking into account the transmission rate of each link and its delivery probability at that transmission rate. In other words, the ETT routing metric extends ETX by taking account of the differences in link transmission rates. The weight of a path is the sum of the ETT of all the links along this path. The ETT of link i is calculated as follows:

$$ETT_i = ETX_i \cdot \frac{S}{B_i} \tag{4.4}$$

where B_i is the transmission rate of link i and S is the packet size. The $\frac{S}{B_i}$ component estimates the expected time to transmit a packet over link i.

The relation $\frac{S}{B_i}$ represents the expected time required to send a packet successfully. The parameters of this relation can be detected in two ways. In the first way, the packet size and transmission rate are obtained from the drivers of the wireless network cards and thus, there is no overhead to capture these parameters. However, there are some drivers that do not provide the bandwidth information. In these cases, the transmission rate has to be estimated. For this reason, probe pair packets are used as a second way of measuring the time expected to transmit a packet. For instance, each node sends probe pair packets to each of its neighbours every 2 s [Draves et al., 2004b]. The first probe packet is small (e.g., about 100 bytes), while the second probe packet is larger than the first packet (e.g., about 1000 bytes). The neighbour measurements the time difference between the reception of the first and the second packets and communicates the value back to the sender. The sender uses a minimum of 10 consecutive samples and then estimates the bandwidth by dividing the size of the second probe packet by the minimum sample. The authours figure out that this estimation is not very accurate for delay, since it does not take into account several factors that affect packet delivery time. Hence, this probing approach has the only purpose to get the links with significantly different bandwidths.

ETT addresses the issue of varying data rates as well as the packet loss rate of different wireless links, because ETX is a part of it. For this reason, there remain a number of drawbacks from ETX in ETT such as not being able to fully capturing the traffic load, intra-flow and inter-flow interference, as well as, not being designed for multi-radio wireless networks. Furthermore, the probing packets used to estimate the transmission rate increase the overhead which means that instability may arise when the medium is very busy. There are other delay-based routing metrics, such as per-hop Round Trip Time (RTT) [Adya et al., 2004], Minimum Delay (MD) [Cordeiro et al., 2007] and Improve Expected Transmission Time (iETT) [Biaz and Qi, 2008], that have the same shortcomings identified in ETT.

The active probing method is also used to estimate implicitly the transmission rate in MD and RTT. The probing method broadcasts a probe packet every 500 ms in RTT, and thus, the overhead is higher than in ETT. Each neighbour responds immediately after receiving a probe packet. The acknowledgment contains a time-stamp so that the delay can be computed. The round-trip technique can underestimate the transmission rate and can overestimate the delay in the wireless links. Namely, this technique estimates about half of the link capacity when compared with the one-way technique, since the two directions of the probing flow share the link capacity; this behaviour is due to the wireless contention resulting from the packet collisions (e.g., between the second probe packet of the packet pair and the acknowledgement of the first probe packet). To overcome this limitation, the MD metric that uses AdHoc Probe [Chen et al., 2005] was proposed. This probing mechanism uses fixed size packet-pairs (e.g., 1000 bytes) to be sent at a constant rate in order to probe link transmission delay in one-way direction. Each node sends probe pair packets to each of its neighbours every 2 s. Hence, this probing approach gathers the delay and transmission rate (implicitly) more precisely than the approaches in [Draves et al., 2004b, Adya et al., 2004]. Nevertheless, the AdHoc Probe may still overestimate the collected values, because of the occasional losses of probe packets in the wireless links.

4.3.2 Interference-aware Routing Metrics

Interference has a significant impact on throughput and delay in WMN [Zhang et al., 2007, Nachtigall et al., 2008] and consequently, this phenomenon influences the application performance. The interference-aware routing metrics have been suggested as a second trend to deal with this. The following sub-sections describe the most relevant interference- aware routing metrics.

4.3.2.1 Weighted Cumulative ETT - WCETT

The WCETT [Draves et al., 2004b] routing metric extends ETT to reduce the number of nodes on the route of a flow that uses the same channel for the whole route. For a route p, WCETT is defined as follows:

$$WCETT(p) = (1-\beta) \sum_{link\ i \in p}^{n} ETT_i + \beta \max_{1 \le j \le k} X_j \tag{4.5}$$

where β is a tunable parameter subject to $0 \le \beta \le 1$. k is the total number of channels in the system, n is the number of links in the path p and X_j is defined as follows:

$$X_j = \sum_{Link\ i\ is\ on\ channel\ j} ETT_i \quad 1 \le j \le k \tag{4.6}$$

The $maxX_j$ component comprises the maximum sum of ETT in the links in the same channel (e.g., channel j). As a result, the total path throughput will be dominated by the bottleneck channel that has the largest X_j, given that the number of links with the same channel tend to increase the $maxX_j$ value. Consequently, WCETT gives more weight to paths with higher $maxX_j$ values and thus, it prefers paths that have greater channel diversity. Furthermore, the original implementation of WCETT uses broadcast probe packets to calculate the ETX and the probe packet pairs required to measure the amount of time $\frac{S}{B_i}$).

The main advantage of WCETT compared to ETT is the fact that it depicts paths with less intra-flow interference. However, WCETT has serious drawbacks. First, it does not take into account the effects of the inter-flow interference and traffic loads and this means that, WCETT may lead to paths through congested areas. Moreover, WCETT lacks an isotonicity property. The isotonic property of a routing metric means that a metric has to ensure that the order of weights of two paths is preserved if they are linked to a common third path. This property is a fundamental requirement for the calculation of minimum weight paths and to achieve loop-free routing. Figure 4.9 illustrates the concept of isonociticity. Examples of these limitations are described in [Yang et al., 2005b].

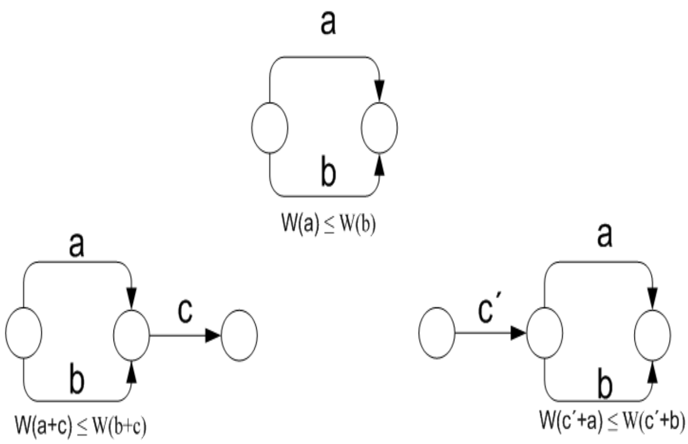

Figure 4.9: Example of isotonicity [Yang et al., 2005b]

Assuming that for any path a, its weight is defined by a weight function $W(a)$ and the concatenation of two paths a and b is denoted by $a \oplus b$.

Definition 1: A weight function $W(\Delta)$ is isotonic if $W(a) = W(b)$ implies both $W(a \oplus c) = W(b \oplus c)$ and $W(c' \oplus a) = W(c' \oplus b)$, for all a, b, c, c' [Yang et al., 2005b]. The isotonicity property of the routing process has the following advantages:

- **Calculation of Minimum Weight Paths.** Both source routing and hop-by- hop routing rely on algorithms, such as Bellman-Ford or Dijkstra's algorithms, to compute the routes. The isotonocity property must be valid to ensure that Bellman Ford and Dijkstra's algorithms find their minimum weight paths.

- **Loop-free Routing.** In the source routing protocols, the source nodes have complete control over the paths of the flows. However, routing loops may occur in hop-by-hop routing, if the metric is non-isotonic. In the specific case of link-state routing by means of the Dijkstra's algorithm, loop-free forwarding requires isotonicity [Sobrinho, 2002].

WCETT, through the $maxX_j$, assumes that if two consecutive links on a path are on the same channel, these links always interfere with each other no matter how long is the distance between them. However, the intra-flow interference in these links will depend on the interference range. This means that, if these links are not in interference range each other, they do not create any interference. Therefore, this assumption is true for short paths, but being outside of each others interference range, this assumption is somehow pessimistic for longer paths.

The Multi-Channel Routing (MCR) metric [Kyasanur and Vaidya, 2006] is an extension of WCETT that takes into account the delay caused by changing channels in multi-channel, multi-radio WMN. However, MCR still has the disadvantages of WCETT.

4.3.2.2 Metric of Interference and Channel-switching - MIC

The MIC [Yang et al., 2005b] metric is an interference-aware routing metric that, like WCETT, also extends ETT to estimate the inter-flow and intra-flow interference. When MIC is used, the cost of a route p is defined as follows:

$$MIC(p) = \alpha \cdot \sum_{link\ i \in p}^{n} IRU_i + \sum_{node\ j \in p}^{m} CSC_j; \tag{4.7}$$

$$\alpha = \frac{1}{(N \cdot \min(ETT))} \tag{4.8}$$

where N is the total number of nodes in the network and the $\min(ETT)$ is the smallest ETT in the network, n and m are the number of links and nodes in the path p, respectively. $\min(ETT)$ can be estimated on the basis of the lowest transmission rate of the wireless cards.

MIC has two components, the Interference-aware Resource Usage (IRU) that depicts the inter-flow interference and the CSC (Channel Switching Cost) that depicts the intra-flow interference. These components are defined in Equation 4.9 and Equation 4.10, respectively:

$$IRU_i = ETT_i \cdot N_i \tag{4.9}$$

where N_i denotes the set of nodes that can interfere in the link i (i.e., the number of interfering neighbours), and ETT_i has been defined in Equation 4.4. The static nature of the infrastructure WMN makes it possible to determine whether two nodes are in each other's interference range at the time when the network is established. In the simulation of MIC, the number of interfering nodes is obtained from the node placement configuration file.

The IRU_i component is the aggregated channel time spent by the transmissions of neighbouring nodes in the link i. This metric captures the rate of inter-flow interference through the multiplication of ETT by all the interfering nodes. The CSC component of MIC allows a reduction of the intra-flow interference since it gives higher weights to paths with consecutive links that use the same channel.

$$CSC_j = \left\{ \begin{array}{ll} w1, & \text{if } CH\left(prev\left(j\right)\right) \neq CH\left(j\right) \\ w2, & \text{if } CH\left(prev\left(j\right)\right) = CH\left(j\right) \end{array} \right\} \tag{4.10}$$

where $0 \leq w1 < w2$, $CH(j)$ represents the channel assigned for node i's transmission and $prev(i)$ represents the previous hop of node i along the route p.

MIC is not an isotonic routing metric. To overcome this limitation, MIC employs a strategy that introduces a virtual network, which is an image of the real network. By adopting this approach, MIC is decomposed into isotonic link weight assignments on virtual links between the virtual nodes [Yang et al., 2005b]. Figure 4.10 shows the non- isotonic behaviour of MIC, where the additional weight that links $(B, C, 1)$ (link $(B \rightarrow C)$ using channel 1) brings it to a path that not only depends on link $(B, C, 1)'s$ own status, but is also related to the channel assignment of the link that precedes link $(B, C, 1)$. Due to the fact that a common channel is used by links $(A, B, 1)$ and $(B, C, 1)$, adding link $(B, C, 1)$ to path $(A, B, 1)$ incurs a higher cost than adding link $(B, C, 1)$ to path $(A, B, 2)$. Hence, even though $MIC((A, B, 1)) < MIC((A, B, 2))$, we have $MIC((A, B, 1) \oplus (B, C, 1)) > MIC((A, B, 2) \oplus (B, C, 1))$, where $\oplus$ indicates a link concatenation.

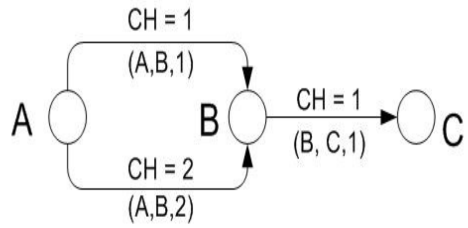

Figure 4.10: MIC without a Virtual Network [Yang et al., 2005b]

By introducing several virtual nodes to represent these possible channel assignments for the preceding link, MIC can be translated into isotonic weight assignments to the links between these virtual nodes. This means that for every channel c that a node $A's$ radios are configured to, two virtual nodes $A_i(c)$ and $A_e(c)$ are introduced. $A_i(c)$ represents the fact that node $prev(A)$ transmits to node A on channel c. $A_e(c)$ indicates that node A transmits to its next hop on channel c.

Figure 4.11 shows an example of the virtual nodes for nodes A, B and C. Links from the ingress virtual nodes to the egress virtual nodes at node A are added and the weights of these links are assigned to establish different CSC costs. Moreover, two additional virtual nodes are introduced, A+ and C- that are the start and end points, respectively.

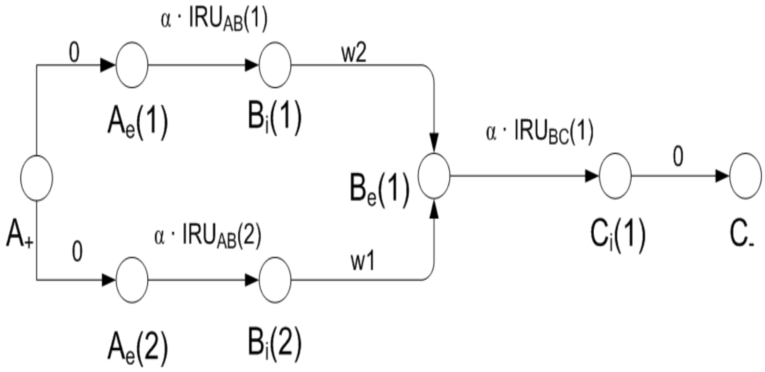

Figure 4.11: Virtual Network of MIC [Yang et al., 2005b]

Link $(A_i(c), A_e(c))$ means that node A does not change channels while forwarding packets and hence weight $w2$ is assigned to this link. Similarly, weight $w1$ is assigned to link $(Ai(c), Ae(c1))$, where $c \neq c1$, to represent the lower cost of changing channels while forwarding packets. Links between the virtual nodes belonging to different real nodes are used to capture the IRU weight. By building the virtual network from a real network, MIC is essentially decomposed from the real network into weight assignments to the links between the virtual nodes. This is because the MIC weight of a real path in a real network can be reconstructed by aggregating all of the weights of the virtual links on the corresponding virtual path. The IRU part of MIC is reflected in the weight of the links between the virtual nodes in different real nodes. The CSC costs are established by routing through different virtual links inside the real nodes. Table 4.1 illustrates the mapping of the real network into the virtual network.

Table 4.1: Real network mapping to the virtual network

Real Path	**Virtual Path**	**MIC Weight**
$(A, B, 1) \oplus (B, C, 1)$	$A_e(1) \rightarrow B_i(1) \rightarrow B_e(1) \rightarrow C_i(1)$	$IRU_{AB}(1) + IRU_{BC}(1) + w2$
$(A, B, 2) \oplus (B, C, 1)$	$A_e(2) \rightarrow B_i(2) \rightarrow B_e(1) \rightarrow C_i(1)$	$IRU_{AB}(2) + IRU_{BC}(1) + w1$

When combined with the virtual network, MIC becomes an isotonic routing metric that pick ups the intra-flow and inter-flow interference in a better way than ETX, ETT and WCETT. Although CSC depicts the intra-flow interference by means of local information, this component is non-scalable and impracticable because the run-time complexity increases significantly with the number of interfaces. However, MIC estimates interference by measuring the number of nodes that can interfere with the transmission. It does not treat interference in a dynamic way. In other words, MIC assumes that all the links located in the interference range of a link lead to the same level of interference. Moreover, it only recognizes the level of interference on a link from the position of the interfering nodes, even though interfering neighbours are not involved in any transmission, whether it occurs simultaneously with that link or not. This is a limitation, since the degree of interference can change over time due to the amount of traffic generated by the interfering nodes. MIC also requires up-to-date information regarding the ETT of each link and this introduces significant overhead and may degrade the overall network performance, particularly in the case of high traffic loads. For example, MIC estimates the inter-flow interference based on the total number of nodes and on the smallest ETT in the network. Furthermore, MIC does not take into account traffic load measurements.

4.3.2.3 Interference Aware Routing - iAWARE

The iAWARE routing metric was recommended to compute paths with a reduction of inter-flow and intra-flow interference by means of signal strength values that continuously reproduce neighbouring interference variations onto routing metrics. The iAWARE metric measurements the degree of interference caused by each interfering node on a link. Moreover, it captures the effects of variation on the loss-ratio link and the differences in the transmission rate through the ETT sub-component, as well as intra- flow interference [Subramanian et al., 2006]. The weighted cumulative path metric iAWARE of a path p is defined as follows:

$$iAWARE(p) = (1-\alpha)\sum_{i=1}^{n} iAWARE_i + \alpha \max_{1\leq j\leq k} X_j \tag{4.11}$$

where $iAWARE_i$ captures the inter-flow interference, X_j depicts the intra-flow interference, α is introduced to represent the trade-off between the inter-flow and intra-flow interference, n is the number of links in the path p and, k is the total number of orthogonal channels available in the network.

The iAWARE metric of a link i is defined as follows:

$$iAWARE_i = \frac{ETT_i}{IR_i} \tag{4.12}$$

The iAWARE uses the HELLO packets sent by the routing protocol to compute ETX. The transmission rate and packet size is needed to compute ETT which is predefined. In addition, the iAWARE uses the transmission rate provided by the driver of the wireless network card. Before it can provide the transmission rate, the network card has to be set in an operating mode called the Radio Frequency Monitor (RFMon) mode. The Interference Ratio (IR) is the component of iAWARE that estimates the interference level in the network through the SNR and SINR, as set out in Equation 4.13.

$$IR_i = \frac{SINR_i}{SNR_i} \tag{4.13}$$

SNR_i and $SINR_i$ are defined in Equations 4.14 and 4.15, respectively.

$$SNR_i = \frac{P_i}{Noise} \tag{4.14}$$

$$SINR_i = \frac{P_i}{Noise + \sum_{w \in N_i - v} \tau_w \cdot P_w} \tag{4.15}$$

where P_i is the signal strength of the link i, N_i denotes the set of nodes that interfere with the link i and the IR_i value pertains to $0 \leq IR_i \leq 1$. τ_w gives the amount of time that node w occupies the channel. It is worth noting that when there is no interference (no interfering neighbours or no traffic generated by interfering neighbours), the $SINR$ of link i is equal to the SNR and thus, IR_i becomes equal to 1.

iAWARE employs $maxX_j$ to take advantage of the diversity of the channel and find paths with less intra-flow interference (i.e., maximum sum of iAWARE over the hops on the same channel), as in the WCETT metric. The main difference is that iAWARE takes full account of the maximum sum of iAWARE over the links on the same channel, whereas WCETT accounts for the maximum sum of ETT. In iAWARE, X_j is defined as follows:

$$X_j = \sum_{conflicting\ links\ i\ on\ channel\ j} iAWARE_i \tag{4.16}$$

where $1 \leq j \leq k$.

iAWARE was the first metric to employ a measurement of inter-flow interference based on the physical interference model, and this is its main advantage. Despite this, the iAWARE routing metric is non-isotonic. The virtual network approach cannot be used by iAWARE to achieve isotonicity, because of its second component which deals with intra-flow interference. The iAWARE and WCETT non-isotonicity is caused by their dependence on the intra-flow interference component that captures the channel assignment of all the links in a path. In other words, the weight increment of adding a link l to a path p depends on how many times each channel has appeared in path p. As the length of p increases, the combination of channel assignments can become infinite and this means that, iAWARE and WCETT cannot be decomposed into virtual networks. Furthermore, another drawback of this metric is that it gives more weight to ETT compared with the interference on the link, i.e. when a link has a higher IR than ETT, the iAWARE metric will have a lower value. This will result in paths with lower ETT but higher interference.

iAWARE uses global information to represent the state of the network. For instance, iAWARE allows a reduction of intra-flow interference by taking into account the maximum number of times that the same channel appears along the whole path. This is because a node does not usually interfere with other nodes that are more than two hops away, even if they share the same channel. The use of global information has several drawbacks associated with the difficulty of gathering information, in particular concerning interference and traffic load characteristics. For example, it is difficult to use max_X in routing protocols that do not rely on flooding and as a result, the nodes do not have information about the channel used for every link in the end-to-end path. Moreover, iAWARE does not take into account traffic load measurements and thus, does not always provide paths with less congestion.

4.3.2.4 Interferer Neighbors Count - INX

The INX [Langar et al., 2009] extends the ETX metric to take into account the interference experienced by the wireless links. Therefore, the INX value of a link i is defined as the product of the ETX of the link i between nodes V and U and the number of all the interferer links resulting from a transmission on that link i. INX can be expressed as follows:

$$INX_i = ETX_i \cdot \sum_{j \in N_i} r_j, \tag{4.17}$$

where N_i is the set of links that can interfere with the transmission on link i and r_j is the available bandwidth of link j. N_i is defined as follows:

$$\left\{ \begin{array}{c} D\left(V,M\right) \leq R_h\left(V\right) \text{ or } D\left(V,N\right) \leq R_h\left(V\right) \text{ or } D\left(M,U\right) \leq R_h\left(M\right): \\ \text{during the transmission of the data packet from } V \text{ to } U, \\ D\left(U,M\right) \leq R_h\left(U\right) \text{ or } D\left(U,N\right) \leq R_h\left(U\right) \text{ or } D\left(M,V\right) \leq R_h\left(M\right): \\ \text{during the transmission of the ACK frame from } U \text{ to } V \end{array} \right\} \quad (4.18)$$

where $R_h\left(V\right)$ is the carrier range (i.e., interference range) of node V and $D\left(M,N\right)$ denotes the Euclidian distance between nodes M and N. If the link (M,N) verifies the condition of Equation 4.18, it is referred to as an interferer link to link i.

INX is an isotonic routing metric that estimates interference by measuring the sum of the transmission rate of links that can disturb the transmission. The asymmetric link is taken into account when defining the set of interfering neighbours. Thus, INX takes into account the interference in a better way than MIC. However, INX is still measured in a way that is based on static information and thus does not include physical interference. Furthermore, INX cannot avoid congested paths, since it is not aware of the traffic load.

4.3.3 Load-aware Routing Metrics

Despite the fact that the transmission rate and interference significantly affect the traffic performance of wireless networks, the traffic load should also be taken into account to improve the path selection decision. Both the basic and interference-aware routing metrics have been enhanced by including load-aware components, as discussed in this section.

4.3.3.1 Weighted Cumulative ETT-Load Balancing - WCETT-LB

The WCETT-LB [Ma and Denko, 2007] routing metric extends WCETT so that it can become integrated with a load balancing component. For a route p, WCETT-LB is defined as follows:

$$WCETT\_LB(p) = WCETT(p) + L(p) \quad (4.19)$$

The load balancing component $L(p)$, has two sub-components, namely, the level of congestion and the level of traffic concentration at each node in a specific path. The congestion level at each node is evaluated by considering the relation between the average queue length and the transmission rate at each node.

$$L(p) = \sum_{node\ i \in p} \left(\frac{QL_i}{b_i} + min\,(ETT) \cdot N_i \right) \tag{4.20}$$

where QL_i is the average queue length and b_i is the transmission rate at a node in a particular path.

The traffic concentration was evaluated in each node by using the N_i parameter, which is the set of children nodes using node i as their next hop. N_i is normalized by the $min\,(ETT)$. This means that if a large number of children nodes choose node i as their next-hop to transmit packets, the traffic at node i will increase. In the light of this, WCETT-LB takes into account the traffic load and logical interference that are not captured by WCETT. However, WCETT is not isotonic and does not detect interference in a dynamic way.

4.3.3.2 Load Aware ETT - LAETT

Load Aware ETT [Aiache et al., 2008] extends ETT so that it can estimate the traffic load of the link. The LAETT of the link between nodes v and u is defined as follows:

$$LAETT_{vu} = ETX_{vu} \cdot \frac{S}{\left(\frac{RC_v + RC_u}{2 \cdot \gamma_{uv}}\right)} \tag{4.21}$$

where γ_{uv} is the link quality factor that is defined according to the distance between the nodes that are defined in [Aiache et al., 2008]. RC_v and RC_u represent the remaining capacity of the nodes.

LAETT uses the HELLO packets sent by the routing protocol to compute ETX. Each node broadcasts periodic HELLO packets with a Time-To-Live (TTL) of one of its neighbouring nodes. Each node recalls the message it received during the previous w seconds. Therefore, the delivery ratios d_f and d_r of ETX can be measured by means of the periodic HELLO packets. The Remaining Capacity (RC) of node v is introduced in order to depict the load of the links. To achieve this, the transmission rates of each flow that traverses a specific node are taken into account. The RC of node i is defined in Equation 4.22.

$$RC_i = B_i - \sum_{flow\ k \in N_f} (f_{ik} \cdot \gamma_{ik}) \tag{4.22}$$

where B_i is the total transmission rate of node i, N_f is the number of flows in the node i, f_{ik} is the transmission rate of each flow and γ_{ik} is the link quality factor of node i.

LAETT improves ETT by adding a traffic load measure. However, this metric assumes that each flow uses the same data rate and thus, the number of flows should be taken into account to estimate the available bandwidth. The available bandwidth is obtained from the network card interface through the sending/receiving transmission rates, while B_i is predefined. However, the bandwidth is difficult to depict in wireless networks in an accurate way, since it is a shared resource and can be degraded as a result of interference. Furthermore, in multi-hop wireless networks, there are flows from different applications that require specific data rates. Hence, LAETT does not include real aspects of the distinct applications of WMN. LAETT still retains some of the drawbacks of ETT and ETX, since it does not depict intra-flow and inter-flow interference and thus is unable to take advantage of the MCMR capability.

4.3.4 Hybrid Routing Metrics

In WMN, most interference is caused by traffic generated in the mesh nodes. As a result of this, interference and traffic load are interrelated. Although interference usually affects the links more than 1 hop way, the influence of traffic load in a link should be regarded at a more local level. In view of this, it is necessary to take both into consideration in order to take advantage of the wireless resources and accurately depict the quality of the link. In this way, hybrid routing metrics would be able to combine interference and traffic load measurements as the main measurements as well being used as basic measurements. Hence, the hybrid routing metrics have emerged as the most recent trend with regard to cross-layer routing metrics.

4.3.4.1 Resource Aware Routing for mEsh - RARE

Despite the improvements made by switching from ETX to iAWARE, some of the metrics previously discussed employ AdHoc probe (i.e., active monitoring) that employs fixed size packet-pairs (e.g., 1000 bytes) to estimate the delay [Chen et al., 2009]. This mechanism may cause an excessive overhead and thus might not scale well in large or high density networks. In addition, there is a need for the active monitoring techniques to access the medium, which may be difficult if the links are congested. This was the main motivation to propose RARE [Kowalik et al., 2007]. This metric only uses the passive monitoring technique to measure the link characteristics, that is, the available bandwidth, signal strength and average contention. RARE is defined as:

$$RARE_i = \alpha \cdot \frac{C - BW_a}{BW_a} + \beta \cdot \frac{RSSI_{max} - RSSI}{RSSI} + \gamma \cdot N_c \tag{4.23}$$

where C is the link capacity, BW_a is the available bandwidth, Received Signal Strength Indicator ($RSSI$) is the signal strength value, $RSSI_{max}$ is the maximum signal strength value, and N_c is the average contention calculated as the average number of deferrals. In addition, α, β and γ are the weights associated with the bandwidth, RSSI and contention components, respectively.

BW_a [Davis and Raimondi, 2005] is used as a traffic load measurement, because it is based on the duration of the busy and idle intervals, which are normalized and combined with the transmission rate (TX_{rate}), as shown in Equation 4.24. Instead of relying on the probe packets, RARE uses passive monitoring mechanisms for this purpose.

$$BW_a = \left(\frac{T_{idle}}{T_{busy} + T_{idle}} \right) \cdot TX_{rate} \tag{4.24}$$

$RSSI$ is also measured in a passive way. The average contention N_c was measured in the wireless cards that are put into the RFMON mode to determine when multiple stations are contending for access. RARE smooths out the measured values of each component by means of an EWMA filter and thus, reduces the routing oscillations.

RARE is the first isotonic routing metric in which all the parameters are measured through a passive approach, and thus does not introduce measurements overhead. In addition, RARE takes into account both physical ($RSSI$) and logical interference (BW_a and N_c measurements). However, RARE does not depict paths with channel diversity and hence, does not result in paths with less intra-flow interference. Moreover, $RSSI$ is not an accurate means of measuring interference, especially at high transmission rates. This means that $RSSI$ cannot depict the fluctuations of interference [Vlavianos et al., 2008] as was described in Section 4.2. Although RARE uses passive measurements, it does not provide accurate information about the quality of the link, when there is little data traffic. As a result, RARE achieves performance results that are very similar to those of ETT.

4.3.4.2 Contention-Aware Transmission Time - CATT

CATT [Genetzakis and Siris, 2008] extends ETT to capture the interference and link congestion levels. By means of CATT, it is possible to obtain a path that minimizes the total packet transmission time and provides load balancing between the links. CATT identifies the congested links by showing the influence that the interfering links, in 1 and 2 hop neighbours, can have on the time needed to transmit a packet over link l. This metric uses the link costs that are averaged over an interval of time, and broadcasts this average, rather than the immediate costs, to reduce route instability. CATT can be expressed as follows:

$$CATT_i = ETX_i \cdot \sum_{j \in N_i} \left(\left(\sum_{k \in N_j} \frac{S_k}{B_k} \right) \cdot \tau_j \cdot \frac{S_j}{B_j} \right) \tag{4.25}$$

where N_i is the set of links that can interfere with the transmission on link i and N_j is the set of links that can interfere with the transmission on link j. S_j and S_k are the packet size of the links in 1 and 2 hop neighbours, respectively. B_j and B_k are the transmission rates of the links in 1 and 2 hop neighbours, respectively. τ_j is the packet transmission attempt rate on link j.

The CATT metric uses the willingness field in HELLO and TOPOLOGY CONTROL messages in the OLSR routing protocol to exchange the transmission rates between the nodes. Each node initially obtains the transmission rate of its interfaces; this information is available through MadWifi's Wireless Extensions API, a set of packages that allows access to information about the wireless network interfaces in the system kernel.

CATT is an isotonic routing metric that depicts the intra-flow and inter-flow interference as well as the traffic load in an uniform way by making a sum of the delays of the interfering neighbours links that are 1 and 2 hops away. Hence, CATT does not require one component for each type of interference which reduces the complexity of this routing metric. Nonetheless, there are two serious drawbacks with CATT. First, it assumes a worst-case approach for estimating interference, in that it assumes that all the interfering links constantly interfere with the transmission over the link and this can result in an overestimated link quality. Secondly, delay does not capture the traffic load over wireless links in an accurate manner as was explained in the previous section. Following similar approach, the Exclusive Expected Transmission Time (EETT) routing metric [Jiang et al., 2007] applies the sum of the delays in the interfering links that are only 1 hop away, which is also very similar to MIC.

4.3.4.3 Interference-Load Aware - ILA

The ILA [Manikantan Shila and Anjali, 2008] metric is a load and interference-aware routing metric. This metric has two components, the Metric of Traffic Interference (MTI) and the Channel Switching Cost (CSC). These components depict the effects of inter-flow and intra-flow interference, respectively. ILA can be represented as follows:

$$ILA(p) = \frac{1}{\alpha} \cdot \sum_{link\, i \in p}^{n} MTI_i + \sum_{node\, j \in p}^{m} CSC_j \tag{4.26}$$

where n is the number of links and m is the number of nodes of the path p, α is defined in Equation 4.28.

MTI_i assumes that the interference levels depend on the traffic load of the interfering nodes and not only on the number of interfering nodes. Equation 4.27 shows the MTI sub-component.

$$MTI_i(C) = \left\{ \begin{array}{cc} ETT_{ij}(C) \cdot AIL_{ij}(C), & N_l \neq 0 \\ ETT_{ij}(C), & N_l = 0 \end{array} \right\} \tag{4.27}$$

where Average Interfering Load (AIL_{ij}) is the average load of the neighbours that may interfere with the transmission between nodes i and j over channel C. ETT is also used in order to identify the difference in transmission rates and packet loss ratio.

The α parameter is used to weight the influence of the inter-flow and intra-flow interference in the metric, as follows:

$$\alpha = \left\{ \begin{array}{cc} \min(ETT) \cdot \min(AIL), & N_l \neq 0 \\ \min(ETT), & N_l = 0 \end{array} \right\} \tag{4.28}$$

where $\min(ETT)$ and $\min(AIL)$ are the smallest ETT and AIL in the network, respectively.

The AIL describes the neighbouring activity of the interfering nodes so that it can avoid the congested links, as defined in Equation 4.29,

$$AIL_{ij}(C) = \frac{\sum_{N_l} IL_{ij}(C)}{N_l(C)} \tag{4.29}$$

where $N_l(C)$ is the set of interfering neighbours of nodes i and j and Interfering Load ($IL_{ij}(C)$) is the load of the interfering neighbour. IL is measured by the average queue length, which is depicted in the number of packets.

ILA also uses the HELLO packets sent by the routing protocol to calculate ETX. In addition, the ETT of a link is computed by means of the ETX, link bandwidth and the size of the packet that is fixed.

This metric addresses the limitations of existing metrics referred to earlier, such as ETX, ETT, WCETT and MIC, by focusing on measurements regarding traffic load, loss packet rate, transmission rate, intra-flow and inter-flow interference. Although ILA does not employ devices from the physical model to measure interference, it picks up the inter-flow interference by only taking account of the amount of traffic generated by interfering neighbours. The intra-flow component (i.e., CSC) becomes increasingly complex as a result of the need of the virtual network to become isotonic. Furthermore, ILA also employs ETX and ETT and thus, may overestimate the link quality.

4.3.4.4 Contention Window Based - CWB

The CWB [Nguyen et al., 2008] takes into account traffic load in two components: congestion window level (CW), and the channel utilization, referred as β_i. CWB is defined as follows:

$$CWB_i = \beta_i \cdot CW_i, \tag{4.30}$$

The channel utilization component relies on the Channel Busy Time (CBT) that represents the fraction of channel time in which the channel is sensed busy (i.e. traffic is sent/received through the channel). Hence, CBT combines the *transmit*, *receive* and *occupied* states. However, the authors do not show how CBT is exactly computed. A mathematical function is used to standardize the way that the channel is utilized on the basis of the threshold values, as expressed in Equation 4.31.

$$\beta_i = \left\{ \begin{array}{cc} 1, & \text{if } u \leq T_1 \\ \min\left(\alpha \cdot (u - T_1) + \exp\left(\frac{u-T_1}{T_2-u}\right), \beta_{max}\right), & \text{if } T_1 < u < T_2 \\ \beta_{max}, & \text{if } u \geq T_2 \end{array} \right\} \tag{4.31}$$

where u is the percentage of channel utilization, T_1 and T_2 are the minimum and maximum threshold of the channel utilization, and β_{max} is the maximum value that the channel utilization can reach.

CBT has been employed in other cross-layer routing metrics, such as Interference-Aware Routing (IAR) [Waharte et al., 2008b] and Expected Forwarding Time (EFT) [Islam et al., 2010]. IAR draws on CBT to measure traffic load and logical interference, while EFT only uses this measurement to depict the logical interference, while relying on queueing delay to measure the traffic load. Both IAR and EFT use the delay measurement which means that some issues from delay remain unresolved (e.g., link quality overestimation). Expected Link Performance (ELP) [Ashraf et al., 2008] extends ETX by including the transmission rate, logical interference through CBT and asymmetry link when depicting the loss rate and interference.

The contention window level on link i is measured through the Frame Error Rate (FER) as defined in Equation 4.32. FER iis obtained from the average of the values over a period of time.

$$CW_i = \frac{1 - FER}{1 - FER^{r+1}} \frac{1 - (2 \cdot FER)^{r+1}}{1 - (2 \cdot FER)} CW_0 \tag{4.32}$$

where CW_0 is the minimum Contention Window and the r maximum back-off stage.

CWB is an isotonic routing metric that combines one measurement of the physical model (FER) with one measurement that reflects the traffic load and interference which are based on the logical model (CBT). The FER is a coarser-grain metric used to depict the link quality when it takes a long time to capture the interference and as a result may not provide a precise value for the interference. Hence, this routing metric is not reliable when the network conditions are quickly changing over a period of time. In addition, there is no clear specification of CBT that can allow one to identify how far the logical interference has been taken into account. Furthermore, CWB does not deal with intra-flow interference. The Airtime routing metric [Hiertz et al., 2007] also employs FER to depict interference. This metric defines the amount of channel resources that are consumed by transmitting the frame over a link. Airtime does not include traffic load and logical interference measurements, and thus, can result in congested paths. This means that the performance of Airtime is very similar to that of ETT.

4.3.5 Discussion

This sub-section summarizes the main aspects of sub-sections 4.2 and 4.3. Table 4.2 shows the components and characteristics supported by each cross-layer routing metric.

Most routing metrics combine measurements or metrics provided by other routing metrics. ETX and ETT are frequently reutilized in most cross-layer routing metrics. As a result, the routing metrics have advantages and drawbacks from the ETX and ETT. The main measurement used in these metrics are transmission rate, delay and loss ratio. Depending on the information made available by the wireless card driver, either active or passive monitoring mechanisms can be used to measure the ETT and ETX components, for instance, to measure the transmission rate.

Table 4.2: Main Components and Characteristics of Cross-Layer Routing Metrics

Routing Metrics	Interference Type	Interference Model	Traffic Load	Isotonicity	Stability Mechanism	Asymmetry Link
ETX	None	None	No	Yes	FHW	Yes
mETX	None	None	No	Yes	FHW	Yes
ENT	None	None	No	Yes	FHW	Yes
ETT	None	None	No	Yes	None	Yes
RTT	None	None	No	Yes	None	No
MD	None	None	No	Yes	None	No
iETT	None	None	No	Yes	None	Yes
WCETT	Intra	None	No	No	None	Yes
MCR	Intra	None	No	No	None	Yes
MIC	Intra and Inter	Logical	No	Yes	None	Yes
iAWARE	Intra and Inter	Physical	No	No	None	Yes
INX	Inter	Logical	No	Yes	None	Yes
ELP	Inter	Logical	No	Yes	FHW and UPT	Yes
Airtime	Inter	Physical	No	Yes	FHW	No
WCETT-LB	Intra	None	Yes	No	None	Yes
LAETT	None	None	Yes	Yes	None	Yes
RARE	Inter	Physical	Yes	Yes	EWMA	No
CATT	Intra and Inter	Logical	Yes	Yes	FHW	Yes
EETT	Inter	Logical	Yes	Yes	None	Yes
CWB	Inter	Physical and Logical	Yes	Yes	FHW	No
IAR	Inter	Logical	Yes	Yes	FHW	No
EFT	Inter	Logical	Yes	Yes	FHW	No
ILA	Intra and Inter	Logical	Yes	Yes	FHW	Yes

Interference-aware routing metrics were suggested as a second approach. Basically, these metrics use measurements derived from the physical and logical model to depict both intra-flow and inter-flow interference. First, there was WCETT, which is only concerned with intra-flow interference. Next, MIC was developed to include both intra-flow and inter-flow interference. However, MIC takes into account the interference in a static way, by depicting the inter-flow interference through the number of nodes that interfere with the transmission of a specific link. INX also views inter-flow interference in a static way like MIC, although it takes into account the sum of the transmission rate of neighbours. iAWARE was the first routing metric (followed by RARE) that took into consideration the physical interference model (i.e., signal strength).

As can also be observed in Table 4.2, most of the routing metrics use two measurements to pick up the intra-flow interference. The first of these is the $maxX$ intra-flow interference component which causes the non-isotonicity property. The non-isotonicity of WCETT and iAWARE is caused by this measurement. The second is the *CSC* component which is a complex and non-scalable means (i.e., the virtual network) of providing isotonicity to the MIC and ILA routing metrics. It is important to note that CATT depicts inter-flow and intra-flow interference in a single component. However, the routing metrics that combine all the weights in a single component, such as ETX, ETT, LAETT, RARE, CWB, CATT and INX, are isotonic. The reason why they are isotonic metrics is that the aggregate path weight is the sum of the weights for all the links in the path and the link weights are non-negative.

Following this, an investigation has been carried out into the load-aware routing metrics, such as LAETT and WCETT-LB that only take into account the traffic load and transmission rates, and use the available bandwidth (i.e., number of flows) and average queue length, respectively. However, they do not depict interference precisely. The load-aware routing metrics help to provide load balancing between the paths, which can smooth out the interference levels over the whole network. In addition, the use of less congested paths results in better traffic performance.

It is worth noting that the load-aware metrics depict the effect of the traffic load in the link quality from a local perspective, while the effect of interference is extended to nodes and links that are at a distance of one or two hops from the generated traffic load. This means that interference and traffic load measurements should be combined in the same routing metric to pick up the link quality; these metrics are described as load-aware and interference-aware routing metrics or hybrid routing metrics, such as ILA, CWB and CATT. Although these metrics enable the link quality to be measured with greater precision, there still remain some shortcomings in the routing metrics reutilized, for instance ETX and ETT. Nonetheless, there are some aspects of this research area that need further investigation such as, assymetric links and logical interference. This particularly applies to CBT, which could be investigated further to allow the contention and traffic load to be depicted more precisely. For example, an attempt could be made to find a mathematical equation based on the time states of the channel such as, *transmit*, *receive*, *occupied*, *idle* and *backoff time*. It is important to point out that none of the analysed cross-layer routing metrics combines the most precise measurements for the physical (i.e. SINR) and logical interference models (i.e. CBT) at the same approach. For instance, CWB uses measurements for these two models, but the FER which depicts the physical interference is not the most accurate measurement for this purpose.

The monitoring methods, being either passive and active probing, employed to obtain measurements are an important characteristic of cross-layer routing metrics.

Table 4.3: Information Gathering Methods of the Routing Metrics

Routing Metrics	Node-related	Monitoring
ETX	No	Active
mETX	No	Active
ENT	No	Active
ETT	Yes	Active
RTT	Yes	Active
MD	Yes	Active
iETT	Yes	Passive
WCETT	Yes	Active
MCR	Yes	Passive
MIC	Yes	Passive
iAWARE	Yes	Passive
INX	Yes	Passive
ELP	Yes	Active
Airtime	No	Passive
WCETT-LB	Yes	Passive
LAETT	Yes	Passive
RARE	No	Passive
CATT	Yes	Passive
EETT	Yes	Passive
CWB	No	Passive
IAR	Yes	Active
EFT	Yes	Passive
ILA	Yes	Passive

On the one hand, passive monitoring is the most employed method in the design of the cross-layer routing metrics since it relies on cross-layer information exchange. In addition, this method does not introduce overhead when obtaining the measurements. Nevertheless, there are some measurements (e.g., transmission rate) that cannot be obtained in a passive way owing to the limitations of the network interface driver and the problem of inaccuracy when there is little data in the network. On the other hand, active probing allows the estimation of measurements that are not available to some of the network drivers, acting as a short cut to overcome this limitation of passive monitoring. However, active probing increases the overhead in the network, can have some inaccuracies and causes routing oscillation in medium and high loads. Despite of these aspects, the active probing used in [Cordeiro et al., 2007] has overcome most of the issues in previous active probing techniques. The transmission rate, delay and packet loss rate are the cross-layer information generally collected by the active probing. Table 4.3 shows the methods that each cross-layer routing metric supports in the original implementation.

Most routing metrics rely on a stability mechanism to smooth out the wide variation of their values. An average figure based on the fixed history window is commonly used in the routing metrics. Despite this, the routing metrics usually do not carry out some performance evaluations or theoretical studies of the parameters of the stability mechanisms adopted. Furthermore, there is a lack of scientific research into the impact of different stability mechanisms in the routing metrics of WMN. Finally, Figure 4.12 groups the cross-layer routing metrics analysed in this section in accordance with the sub-taxonomy of the measurements.

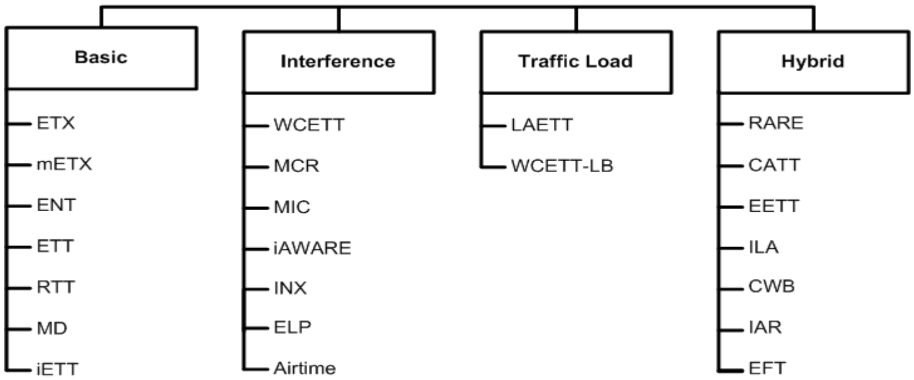

Figure 4.12: Mapping the Routing Metrics in the Sub-taxonomy of the Measurements

Despite the set of analysed cross-layer routing metrics tends to be distributed equally between the categories of the sub-taxonomy (excepting for the traffic load category), it is important to notice that the most of recent metrics belongs to the hybrid category. Hence, the cross-layer routing metrics has combined more different kinds of measurements at a same approach than using only one type.

4.4 Metric for INterference and channel Diversity (MIND)

In this section, the MIND proposal [Borges et al., 2009] is described as one of components of the ACRoMa architecture which is proposed in this thesis. Although several cross-layer routing metrics have been proposed for WMN, they have some limitations which were addressed in Section 4.3. For example, none of the analysed cross-layer routing metrics employs precise measurements for the physical and logical interference models at the same metric which is the main motivation for MIND. In this view, MIND combines measurements that take into account interference (e.g. physical and logical) and traffic load through accurate and passive measurements. This metric includes two components: the first component concerns the inter-flow interference and load awareness ($INTER\_LOAD$), while the second component depicts the intra-flow interference (CSC) which is very similar to [Yang et al., 2005b]. $MIND$ can be defined as follows:

$$MIND(p) = \sum_{link\ i \in p}^{n} INTER\_LOAD_i + \sum_{node\ j \in p}^{m} CSC_j \tag{4.33}$$

where n is the number of links and m is the number of nodes of path p.

The $INTER\_LOAD$ component depicts information about interference and traffic load simultaneously. Interference Ratio (IR) is also extended from [Subramanian et al., 2006] to capture the interference between the links. MIND regards CBT as a smooth function of multiple weighting through IR. For this reason, MIND strikes a combination between interference and load, in which interference has a higher weight than traffic load. τ is a configurable parameter that determines the higher weight of the interference in the $MIND$ component. The $INTER\_LOAD$ component is defined in Equation 4.34,

$$INTER\_LOAD_i = ((1 - IR_i) \cdot \tau) \cdot CBT_i \tag{4.34}$$

where $0 \leq IR \leq 1$.

The IR component of MIND is similar to the IR measurement outlined in [Subramanian et al., 2006]. However, the $SINR$ adopted in MIND does not take into consideration that a node only occupies the channel (i.e., CBT). In fact, MIND uses the CBT as sub-component. The $SINR$ can be defined as follows:

$$SINR_i = \frac{P_i}{Noise + \sum_{w \in N_i} P_w} \tag{4.35}$$

where P_i is a signal strength of the link i, N_i denotes the set of nodes from which interference in the link i and the IR_i value pertains to $0 \leq IR_i \leq 1$.

CBT is also employed to estimate the traffic load and logical interference. In MIND, this measurement is estimated by applying Equation 4.36. The CBT calculation is based on the time that it takes for the packets in the wireless medium to complete a successful transmission. In other words, it uses an estimation based on the idle and total periods. The $TotalTime$ includes the time between the first attempt to send the packet and the reception of its ACK. In other words, CBT is a percentage of time that includes the times from the *transmit*, *receive* and *occupied* states during the attempt to transmit a packet.

The value of the $IdleTime$ (Equation 4.36) includes the backoff times and the time periods in which the nodes regard the radio medium as being available for access. There are inter-frame spaces in which the channel is idle before and after each busy period, such as Distribute Coordination Function (DCF), InterFrame Space (DIFS), Short InterFrame Space (SIFS) and Extended InterFrame Space (EIFS). These can be described as follows: any node has to be aware of the status of the wireless medium before the transmission in the DFC protocol; if the medium is continuously idle for the DIFS duration, it is only allowed to transmit a frame, while if the channel is found busy during the DIFS interval, the node defers its transmission; SIFS is the short time period between the data frame and its acknowledgment. If a previously received frame contains an error, it has to defer the EIFS period, rather than the DIFS period, before transmitting a frame. Hence, $IdleTime$ combines the interframe space and backoff time period. Hence, CBT is defined as follows:

$$CBT_i = \frac{TotalTime - IdleTime}{TotalTime} \tag{4.36}$$

MIND also uses smoothing out functions to avoid routing instability. For instance, the IR and CBT components are smoothed out through their respective averages of a set of packets. In this way, CBT can be computed as the average of a specific number of packets transmitted, including both data and control packets. With this approach, there will always be packets to calculate CBT. However, as mentioned earlier, passive measurements can not be precise when traffic is reduced.

The MIND metric provides an approach to integrate physical and logical interference as well as to capture both the intra-flow and inter-flow interference components. The main advantage of MIND is that it employs precise measurements of physical and logical interference as well as traffic load, in a passive monitoring approach. In addition, in a similar way to CWB and RARE, MIND does not employ ETX or ETT as measurements and thus, avoids the drawbacks of these routing metrics such as, routing oscillation and poor performance in high loads.

4.5 Performance Evaluation

A thorough simulation study is undertaken to validate the MIND metric and to carry out this goal, MIND is compared with the most important routing metrics for WMN. This section is structured as follows: the impact of cross-layer routing metrics on the triple-play service is discussed in Sub-section 4.5.1. Sub-sections 4.5.2 and 4.5.3 analyse the influence of outdoor and indoor environments on the cross-layer routing metrics, respectively. The effects of cross-layer routing metrics on routing stability and the QoE performance parameters are shown in Sub-section 4.5.4.

4.5.1 Effects of Cross-layer Routing Metrics on the Triple Play Service

The simulation study outlined in this sub-section aims at shedding light on the capabilities of the different cross-layer routing metrics in WMN with a high degree of interference. The simulated scenario establishes configuration parameters to achieve this aim (e.g. traffic pattern, topology size and placement). A comparison was drawn between the most suitable and recent hybrid routing metrics for the FTP traffic performance, VoIP and video streaming, as well as how they operate in combination to configure triple-play services. The number of companies and academic projects which focus on this traffic configuration has increased [IBM, 2006][Boccolini et al., 2011][Azcorra et al., 2009]. Furthermore, as far as we know, no cross-layer routing metric has been evaluated for this kind of traffic. It is worth noting that at this stage clustering has not yet been included in the simulation. Hence, this simulation evaluates the cross-layer routing metrics in a network similar to that of the intra-cluster structure. This sub-section is structured as follows: the scenario configuration is outlined in sub-section 4.5.1.1. Sub-section 4.5.1.2 discusses the results of the evaluation.

4.5.1.1 Simulation Configuration

The NS-2 simulation tool version 2.31 [NS-2, 2012] was used to evaluate MIND and compare it with the MIC, iAWARE, INX and CATT metrics, since they are the main interference-aware routing metrics that have been analysed. In addition, this study aims to simulate a realistic WMN, and thus it provides a fair evaluation of the selected cross-layer routing metrics. The source code of developed cross-layer routing metrics in NS-2 is made available to download in [Borges and Pereira, 2009]. For this reason, ETX and ETT are designed in the same way for all the routing metrics that employ them. As a result, ETX is computed through the HELLO packets sent by the routing protocol [Manikantan Shila and Anjali, 2008, Subramanian et al., 2006, Langar et al., 2009, Aïache et al., 2008], and ETT [Draves et al., 2004b, Cordeiro et al., 2007] is implemented by means of probe packets. These routing metrics were implemented in an extended version of the OLSR routing protocol [Cordeiro et al., 2007] by means of the NS-2 simulator version 2.31 [NS-2, 2012]. This routing protocol is commonly used in WMN, because it allows link state information to be disseminated efficiently [Genetzakis and Siris, 2008, Campista et al., 2008, Nguyen et al., 2008].

The main features of the scenario used for the evaluation of the routing metrics are shown in Table 4.4. These configurations are usually employed in outdoor city-wide deployments [Genetzakis and Siris, 2008]. The performance is evaluated in terms of application level throughput, since this is the parameter used to describe the global performance of the applications. Each data point in the graphical results is computed as the average of 10 different runs where the confidence intervals of the performance parameters (which have a confidence level of 95%), are also shown. Furthermore, each run has a specific scenario with different random node placement.

Table 4.4: Scenario Setup

Parameter	Value
Simulation Time	120s
Flow Lifetime	115s
Network Size	50
Topology Size	1000m x 1000m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

All of the nodes have the same physical configuration; there are two channels and two network interfaces. Each channel is combined with one particular network interface, and no channel assignment algorithm has been employed. The nodes usually have multiple radios in WMN, where each radio can define different link capacities, depending on environmental conditions. The transmission rate of each radio is based on the Adaptive Auto Rate Fallback (AARF) algorithm [Lacage et al., 2004] so that it can simulate a realistic environment. Moreover, the transmission rate changes over a period of time in accordance with the degree of packet loss in the wireless channel and as a result, this scenario can have links with a heterogeneous capacity. In this algorithm, the transmission rates vary according to the following values 1, 2, 5.5 and 11 Mpbs. This algorithm is available in a patch for NS-2 [Fiore, 2009].

The standard NS-2 channel model only takes into account the received signal power to determine the correct reception of a frame. The effect of interference and different thermal noises, as well as the impact of the transmission rates employed, are ignored. This means that the transmission range of a wireless station is the same for any data transmission rate, which is unrealistic. It should be pointed out that all the evaluated routing metrics are interference-aware and thus require a simulator that includes a channel propagation that simulates the effects of interference in a realistic way. For this reason, Marco Fiore's patch [Fiore, 2009] also includes both of the recommended improvements [Xiuchao, 2004] by taking into account the effect of interference and different thermal noises to compute the cumulative Signal to Interference-plus-Noise Ratio (SINR) cumulative thus accounting for the different Bit Error Rate (BER) SINR curves for the various transmission rates used. For these reasons, this patch is employed in all tests of simulation presented in this thesis.

The traffic proportion of each application in the mixed traffic at Table 4.5 was based on [Quintero et al., 2004][Kim et al., 2008], that is, the percentage of flows for VoIP, FTP and video are 60%, 30% and 10% of the total load, respectively. Thus, a set of four combinations of mixed traffic were prepared, as follows: combination A (1, 2 and 4 flows of video, FTP and VoIP, respectively), combination B (1, 3 and 6 flows of video, FTP and VoIP, respectively), combination C (2, 4 and 8 flows of video, FTP and VoIP, respectively) and combination D (2, 5 and 10 flows of video, FTP and VoIP, respectively).

Table 4.5: Traffic mix

Applications/Combinations	Video	FTP	VoIP
A	1	2	4
B	1	3	6
C	2	4	8
D	2	5	10

The video streaming simulation uses Variable Bit Rate (VBR) flows with an average rate of 264 Kb/s (standard deviation of 3 Kb/s). The Evalvid platform [EvalVid, 2012] was configured to support Moving Picture Experts Group (MPEG-4) with I and P frames, and to control the quality of the real video traces in NS-2. The ns2voip [Bacioccola et al., 2007] module was used to simulate VoIP in NS-2. The VoIP traffic was modelled on the basis of [Chuah and Katz, 2002], which uses the G729 codec, $Weibull\,(\delta, \beta)$ function distribution for the talkspurt ($\delta = 1.42s, \beta = 0.82s$) and silence ($\delta = 0.89s, \beta = 1.08s$) periods. FTP employs the default settings of the NS-2 (Tahoe TCP). The scenario uses a typical Client WMN traffic pattern feature that is very similar to MANETs, where several flows originated from the source nodes to different destination nodes, and the source and destination nodes were chosen at random. The reason why this traffic pattern was chosen is because it spreads the traffic load throughout the network and hence, the interference is also caused through the whole network. Every flow starts and finishes at the same time. The lifetime of each flow is 115s and there is a warmup period of 5s for a total simulation time of 120s.

4.5.1.2 Simulation Results

Figures 4.13, 4.14, 4.15 and 4.16 show the throughput of MIND, CATT, MIC, iAWARE and INX for all the load configurations (i.e. number of flows). The total throughput is calculated by summing up the flows in each load configuration. All the applications and load conditions have the highest goodput when MIND is used, due to the fact that MIND makes a combination between interference and traffic load measurements through accurate and passive measurements. The precise measurements of interference and traffic load improve the throughput significantly. For example, the performance difference among routing metrics is higher when the FTP single and mixed traffic are evaluated, since these types of traffic require a greater amount of bandwidth than VoIP and video. CATT provides better throughput than INX, MIC and iAWARE in high loads, since it recognises the influence of the interference and traffic load of the 1 and 2 hop neighbours, whereas MIC and INX only recognise the influence of the 1 hop neighbours.

Although iAWARE views interference in a more dynamic way than MIC through the signal strengths used in this metric, both iAWARE and MIC give more weight to the component that measures delay than to the interference-aware component, and thus, iAWARE and MIC show a very similar throughput in all the evaluated applications. It is worth noting that iAWARE and MIC use the same component to measure delay (i.e. ETT). As expected, INX results in worse throughput than MIC and iAWARE with high loads for video and VoIP, and moreover, when the FTP and mixed traffic for all traffic load conditions, because these cases generate a large amount of data. The reason is that the packet loss measurement used in INX is not accurate enough to show the link quality in high loads as well as environments with high level of interference, and thus, the packet loss fails to distinguish accurately between the links which have high interference and congestion. This is an evidence that the choice of cross-layer routing metric can affect the traffic performance for some appications and traffic load.

CATT, MIC and INX routing metrics assume that all the neighboring nodes are

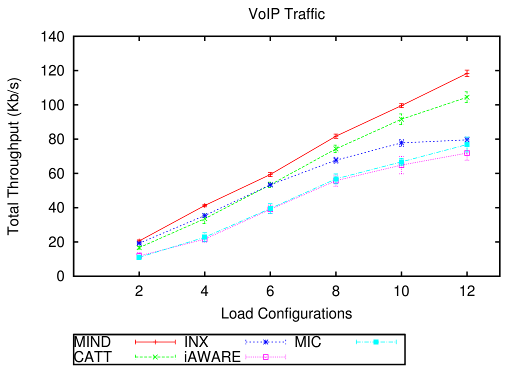

Figure 4.13: VoIP Traffic

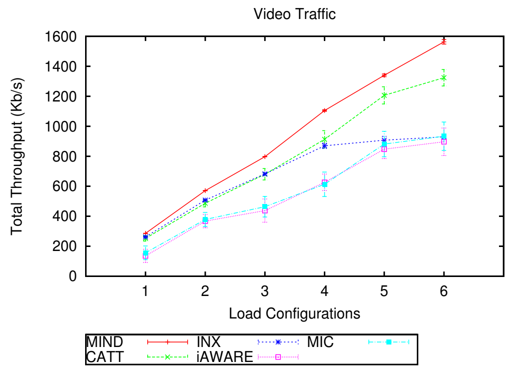

Figure 4.14: Video Streaming Traffic

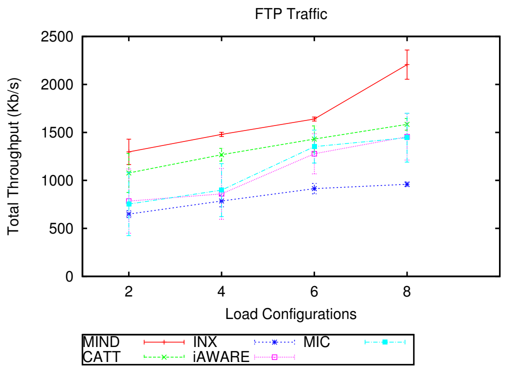

Figure 4.15: FTP Traffic

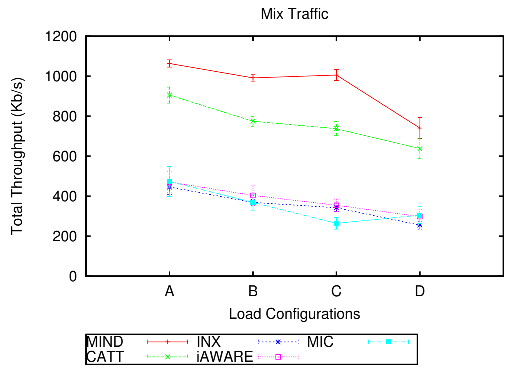

Figure 4.16: Mixed Traffic

transmitting packets. Thus, these metrics can not depict links with distinct interference levels in areas with a large number of nodes. Moreover, CATT, iAWARE, MIC and INX employ probing to pick up the link quality. As this technique also tends to overestimate the link quality, it can result in an unstable behavior with applications that require high bandwidth [Genetzakis and Siris, 2008].

With regard to the mixed traffic, in particular, the throughput does not increase with higher load, as is the case when the routing metrics are evaluated with a single application. The reason for this is that there is a huge amount of flows and data and thus, the traffic experiences high losses due to the high levels of interference. Furthermore, it should be noted that both applications (as well as the mixed application traffic) achieved the best throughput when MIND is used, particularly in mixed traffic. It can thus be inferred that MIND is a useful means of supporting triple play services in a scalable way, since it takes into account measurements that show a better link quality than the other routing metrics.

4.5.2 Effects of Outdoor Environment on the Cross-layer Routing Metrics

The main objective of this sub-section and the next sub-section is to analyse representative results of two different groups of simulation experiments when video streaming is used. The first is carried out to determine the performance of the routing metrics in an outdoor environment. The second group aims to shed light on the performance of cross-layer routing metrics in an indoor environment. It is advisable to evaluate the sensitivity of the cross-layer routing metrics to the interference found in these two environments, because of their specific characteristics that influence the levels of interference. Each of these environments has a distinct propagation model, network area size and number of nodes as well as varying transmission ranges. In view of this, both groups are evaluated in the light of certain values of the scenario configuration that are combined to model the conditions of each specific environment. This sub-section is organized as follows: scenario configuration and evaluation parameters are described on sub-sections 4.5.2.1 and 4.5.2.2, respectively. Sub-section 4.5.2.3 discusses the results of the evaluation.

4.5.2.1 Simulation Configuration

The cross-layer routing metrics used in the previous section were also compared for outdoor and indoor environments. Furthermore, to ensure that accurate results were obtained, 10 scenarios were examined with different random node placements, and thus, each data point in the graphical results is computed as being the average of 10 distinct simulations as well all the graphs show the confidence intervals of the performance parameters which have a confidence level of 95%. Table 4.6 shows the simulation parameters for the video streaming following the settings proposed in [Park and Han, 2008].

The scenario deploys a typical traffic pattern for infrastructure WMN, where several flows were originated from the source nodes (i.e. mesh routers) towards a destination node (i.e., gateway), and the source nodes were chosen at random. The gateway is located in the central position [Bejerano et al., 2007]. All the flows start and finish at the same time. An initial period of 30 seconds (i.e. a warm-up period) is undertaken for the simulation before the flows start. When the lifetime of the flow is over, an end period of 30s is adopted to ensure that all the data has been transmitted.

The simulation focuses on video streaming with Variable Bit Rate (VBR) at an

Table 4.6: Simulation Parameters

Parameter	Value
Flow Variable Bit Rate	Randomly chosen between 253 Kb/s and 259 Kb/s
Simulation Time	300s
Flow Lifetime	240s
Frames per Second (FPS)	25
GOP Size	25
Network Interface Cards	2
Data Channel Rate	11Mb/s
PHY Specification	802.11 b/g
Antenna	Omnidirectional
Runs	10

average rate of 256 Kb/s (standard deviation of 3 Kb/s), generated by the Evalvid platform [Lie and Klaue, 2008]. In the interests of simplicity and owing to the greater complexity caused by the B (Bi-directional coded) frames, the Evalvid platform assumes MPEG-4 with I (Intra coded) and P (Predictive coded) frames [Schwarz et al., 2007] to control the quality of the delivery of real video traces in this simulation study. The I frames can be reconstructed without any reference to other frames. The P frames are predicted in advance of the last I frame or P frame and thus, it is impossible to reconstruct them without using the data from another frame (I or P). Hence, the I frame is of greater significance in the user perception quality than the P frame and therefore I frame is larger in size than the P frame.

The simulation parameters of the outdoor environment are defined according [Capone and Martignon, 2007, Xiuchao, 2004], as listed in Table 4.7. The outdoor scenario consists of 1 gateway (located in the central position [Bejerano et al., 2007]) and 49 static mesh routers with Multi-Channel Multi-Radio (MCMR) capability and is typical of outdoor city-wide deployments.

Table 4.7: Simulation Parameters of Outdoor Environment

Parameter	Value
Number of Nodes	50
Size of Network Area	1500mx1500m
Number of flows	6
Transmission Range	250, 300, 350 and 400m
Interference Range	550, 600, 650 and 700m
Propagation Model	TwoRayGround

It should be noted that each transmission range value corresponds to a single value of the interference range. For instance, for a transmission range of 250 and

300 meters, the interference range is 550 and 600 meters, respectively.

4.5.2.2 Evaluation Parameters

This section describes the parameters used to evaluate the routing process and traffic performance. Traffic performance can be assessed from two perspectives: QoS (i.e network level) and QoE (i.e. user level) parameters. The QoS parameters employed to evaluate the network performance are the packet delay and throughput, as follows:

- **Packet Delay.** This parameter states the time that a packet takes to travel across the network from a source node to a destination node [G.114, 2003].

- **Jitter**. This parameter measures the statistical variance of data packet interarrival time. The jitter is critical in applications where decoders have to receive packets at a constant rate such as, video and VoIP. To overcome this problem, a buffer is usually introduced on the client side. Nevertheless, this solution brings another problem that is the definition of the buffer size. It is worth noting that large buffers lead to further delays, while small buffers may have a low adaptation capacity and, consequently, high losses [Schulzrinne et al., 2003].

- **Throughput.** This parameter represents the amount of data successfully moved from one place to another in a given time period. The throughput of a network may be lower than the input rate due to loss and delay experienced in the network [Rappaport, 2001].

The QoE parameter used to evaluate the quality of the video applications is described as follows:

- **Structural SIMilarity Index (SSIM)** [Wang et al., 2004a]. This parameter is based on frame-to-frame measurement of three Human Visual System components: luminance similarity, contrast similarity and structural similarity [Rouse and Hemami, 2008]. The SSIM index outputs a decimal value between 0 and 1, where 0 means no correlation with the original image, and 1 means the same image. This parameter has three components that are relatively independent. For example, the change of luminance and contrast will not affect the structure of the image. Hence, SSIM gives details about the level of the video quality and takes human perception into account.

The routing process is evaluated through the total routing overhead associated with each evaluated cross-layer routing metric, since an excessive routing overhead can be caused by frequently changing of the values in the routing metrics. The routing overhead parameter used is defined as follows:

- **Total Routing Overhead**. This parameter measures the number of received control routing packets by all the nodes.

4.5.2.3 Simulation Results

The results obtained for all evaluated parameters follow a common pattern, where the video performance decreases in proportion with the increase of transmission and interference ranges, since the interference in this environment also increases, thus leading to a deterioration of traffic performance. However, the routing metrics have distinct impacts on the video performance. Figures 4.17, 4.18 and 4.19 show the lowest throughput, highest delay and highest jitter, when MIC and iAWARE are used. Similar to the previous sub-section, the traffic performance has a slight difference when the MIC and iAWARE are employed. They give more weight to the component that measures delay than to the interference-aware component and furthermore, they use the same component to depict delay (i.e. ETT).

Despite the fact that CATT is also based on the expected transmission time, this metric results in higher throughput and lower delay than MIC and iAWARE, because CATT is able to pick up the influence of the interference and traffic load weight of the 1 and 2 links away from a link. Surprisingly, INX achieves a higher throughput than MIC, iAWARE and CATT, by avoiding the use of ETT. This can be explained by the difference of simulated scenarios. For example, the simulated outdoor scenario of this sub-section has lower degree of interference than the scenario used in the triple play service evaluation, since the previous scenario employed a smaller topology size which increases the inteference levels. Also, as an evidence of this, when INX is used, the throughput decreases in proportion to the increase of the transmission ranges (Figure 4.17); for instance, INX and CATT display a similar throughput in higher transmission ranges (i.e. 350m and 400m). The reason for this is that INX is based on a packet loss rate measurement that is not sensitive to an increase in interference or channel variations [Beuster et al., 2008]. Furthermore, CATT results in a lower delay than INX in higher transmission ranges (Figure 4.18), because CATT is based on a delay measurement (i.e. ETT). MIND achieves the best performance in all the network performance parameters because it takes into account precise and passive measurements when depicting interference. Moreover,

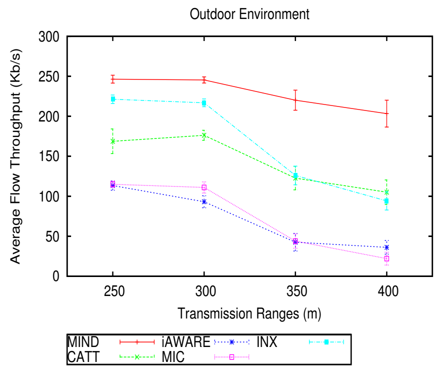

Figure 4.17: Average Flow Throughput

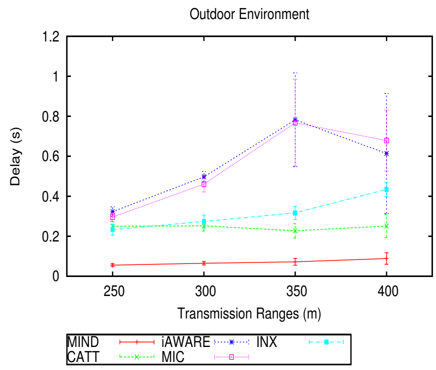

Figure 4.18: Average Flow Delay

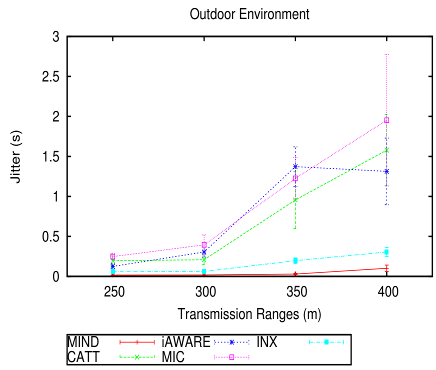

Figure 4.19: Average Flow Jitter

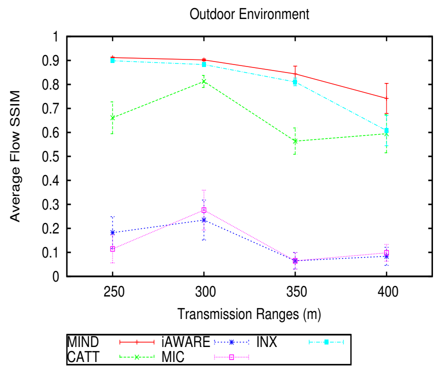

Figure 4.20: Average Flow SSIM

MIND does not employ neither ETX nor ETT metrics.

Figure 4.20 shows that the user perception is influenced by the IP parameters. However, in some cases, jitter and delay result in more impact than throughput. For example, although INX and CATT show a very similar throughput (Figure 4.17), within the transmission range of 350m, CATT has a higher jitter than INX (Figure 4.19) and therefore, CATT achieves lower SSIM than INX in this range. Moreover, in the transmission range of 400m, INX and CATT result in similar SSIM values, even though CATT achieves a higher jitter than INX, because INX has a higher delay and slightly lower throughput than CATT in the transmission range of 400m (Figure 4.18). Hence, as a result of the delay and jitter values, these network performance parameters may have a more significant impact on the user perception than on throughput.

Figure 4.21 illustrates the way that the increase of the routing overhead matches that of the transmission range, the reason being that there is a rise in the number of neighbours as well as in the received control packets of the routing protocol. The lowest routing overhead, when MIND is employed, can be explained by the fact

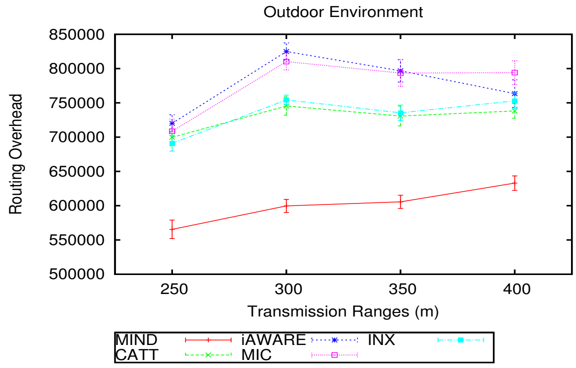

Figure 4.21: Routing Overhead in Outdoor Environment

that MIND only uses passive measurements, while the other metrics employ active probing methods. It should be stressed that the rise of update routing messages (i.e. control packets) can also increase the convergence time of the routing algorithm, so that the computed paths may not reflect the real state of the network. Thus, the results show that the information gathering method affects significantly the routing overhead. Furthermore, the routing overhead also shows to be different from the network performance parameters in some cases. For example, in the transmission ranges of 250m and 300m, INX and CATT result in similar routing overhead, although INX achieves a higher throughput than CATT. In this view, the high level of routing overhead does not always imply a worse performance while, at the same time, a low level of routing overhead does not always mean a better performance. Hence, the traffic performance is mainly affected by the measurements that are combined in the cross-layer routing metrics rather than the routing overhead.

4.5.3 Effects of Indoor Environment on the Cross-layer Routing Metrics

The purpose of this sub-section is to assess the video traffic performance in an indoor environment. For this reason, the performance parameters which were used to evaluate the cross-layer routing metrics in the outdoor environment are also employed in this sub-section. This sub-section is organized as follows: scenario configuration is shown in sub-section 4.5.3.1 and sub-section 4.5.3.2 discusses the results of the evaluation.

4.5.3.1 Simulation Configuration

The simulation parameters of the indoor environment are defined in Table 4.8, the remainder details of the scenario configuration shown in this table are the same as employed in outdoor environment. The topology size, interference and transmission ranges of the indoor environment are proportionally defined in accordance with the topology size and ranges of the outdoor scenario. The indoor scenario consists of 1 gateway (located in the corner) and 14 static mesh routers [Draves et al., 2004b][Yousefi et al., 2006] with MCMR capability and is typical of a university building lay-out, where there are open corridors, classrooms and computer labs. Hence, the path loss exponent and the shadowing deviation standard were defined for this specific indoor scenario in agreement with [Chuah and Katz, 2002][Bacioccola et al., 2007].

Table 4.8: Simulation Parameters of Indoor Environment

Parameter	Value
Number of Nodes	15
Size of Network Area	150mx150m
Number of flows	4
Transmission Range	25, 30, 35 and 40m
Interference Range	50, 60, 70 and 80m
Propagation Model	Shadowing
Path Loss Exponent	2
Shadowing Standard Deviation	4

4.5.3.2 Simulation Results

In the indoor environment, the results obtained in Figures 4.22, 4.23 and 4.24 follow a different pattern from those of the outdoor environment. In this case, the video performance increases slightly so that it conforms with the rise of the transmission range. There are several reasons that explain this such as, the lower number of nodes, the lower number of links, and the fact that the links have the same transmission rate and are at a much shorter distance than in the outdoor environment. As a result, this specific indoor environment provides links which have a higher capacity than in the outdoor scenario and this explains why the difference between MIND and INX decreases in proportion with the increase of the transmission range. However, MIND still achieves a better performance than INX due to the fact that it measures interference with more accuracy than INX, since MIND takes into account logical and physical interference.

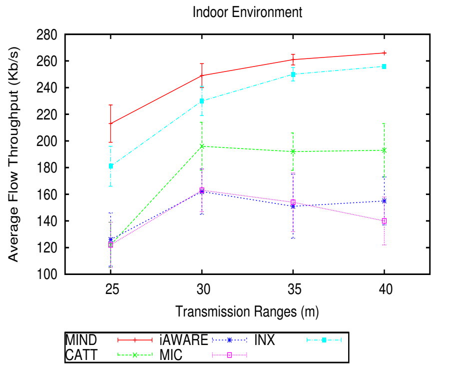

Figure 4.22: Average Flow Throughput

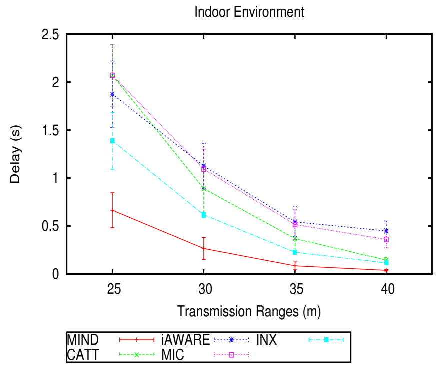

Figure 4.23: Average Flow Delay

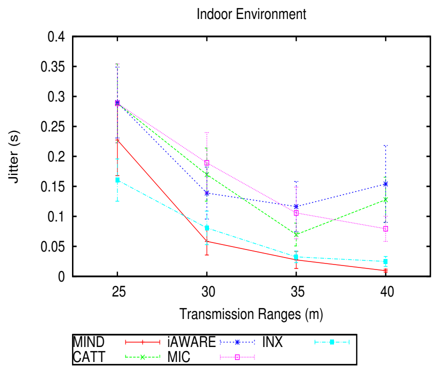

Figure 4.24: Average Flow Jitter

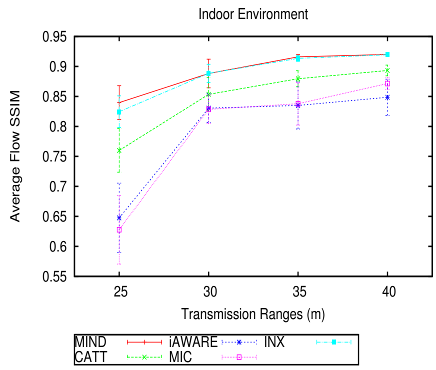

Figure 4.25: Average Flow SSIM

Figure 4.25 illustrates the SSIM parameters. Video streaming achieves a good level of quality in all of the transmission ranges when MIND and INX are employed, because the jitter achieved is very similar when both are used. Whilst the CATT routing metric results in very similar network performance parameters to MIC and iAWARE in the transmission range of 25m, CATT has the highest SSIM value in this range. The reason for this is that CATT results in lower I and P frame losses than MIC and iAWARE, and CATT and INX result in similar I and P frame losses (Figure 4.26). Despite that INX and CATT show similar frame losses in the transmission range of 25m, CATT does not achieve a SSIM similar to INX, because it has a higher delay and jitter than INX in the transmission range of 25m. Hence, in some cases, the impact of the routing metrics on the network and user level parameters cannot be directly correlated. It should be stressed that the loss of specific frames, I or P frames, can have different impact on the user perception (e.g. motion frames). Furthermore, the losses experienced by I frames are greater than the losses of P frames because I frames are larger and consequently, there are more chances of losses in the transmission.

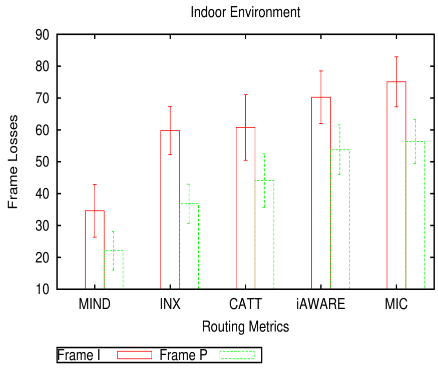

Figure 4.26: Transmission range of 25m

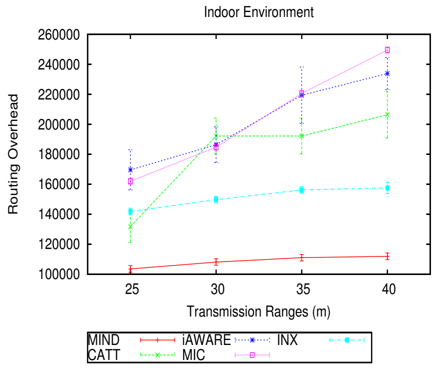

Figure 4.27: Routing Overhead

Figure 4.27 shows that the routing overhead also increases in proportion with the rise of the transmission range. Unlike the outdoor scenario, INX results in less overhead than CATT in almost all transmission ranges. Once more, the level of routing overhead resulting from by the routing metrics does not always imply a worse or better performance. For instance, in the transmission range of 25m, CATT results in lower routing overhead than INX, but INX achieves higher throughput, lowest delay and lowest jitter. Despite of the large difference of routing overhead between MIND and INX, the network performance parameters are very similar when MIND and INX are employed.

4.5.4 Effects of Cross-layer Routing Metrics on Routing Stability and QoE parameters

The simulation study in this sub-section shows a comparison of the most relevant and recent cross-layer routing metrics in terms of routing stability and QoE parameters, when VoIP traffic is used. However, iAWARE is not included in this evaluation, since the iAWARE has resulted in very similar traffic performance MIC. Furthermore, ILA and CWB are taken into account, since they use measurements which were not yet evaluated (e.g, queue length and Frame Error Rate). This sub-section is organized as follows: scenario configuration and evaluation parameters are described on sub-sections 4.5.4.1 and 4.5.4.2, respectively. Sub-section 4.5.4.3 discusses the results of the evaluation.

4.5.4.1 Simulation Configuration

The characteristics of the scenario used for the evaluation of the routing metrics are shown in Table 4.9. These configurations are defined according to [He et al., 2009, Capone and Martignon, 2007, Xiuchao, 2004], where configurations of the transmission and interference ranges, propagation model, and topology of the size are defined to outdoor scenarios.

Table 4.9: Scenario Setup

Parameter	Value
Simulation Time	100s
Flow Lifetime	60s
Network Size	50
Topology Size	1500m x 1500m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

The scenario consists of 1 gateway and 49 static mesh routers with multi-channel multi-radio capability and is typical of outdoor city-wide deployments. There are two channels and two network interfaces. The scenario uses a typical WMN traffic pattern characteristic, where several flows were originated from the source nodes (i.e. mesh routers) to a destination node (i.e., gateway), and the source nodes were chosen at random. The gateway is located in the central position [Bejerano et al., 2007]. All the flows start and finish at the same time. An initial period of 10 seconds of the simulation (i.e. a warm-up period) is undertaken before the flows start. When the lifetime of the flow is over, an end period of 30s is adopted to ensure that all the data has been transmitted. The total number of flows is based on [Dai and Han, 2003], where the number of voice calls is defined in order to maintain a good VoIP quality.

4.5.4.2 Evaluation Parameters

The QoS parameters used in this section is throughput and delay. In addition, the QoE parameter used to evaluate the quality of the VoIP applications is described as follows:

- **Mean Opinion Score (MOS)** is a subjective parameter that is used to evaluate the quality of the multimedia content. In VoIP evaluation, the scale varies according to the following values 1, 2, 3, 4 and 5 (that are equivalent to Bad, Poor, Fair, Good and Excellent, respectively) [p.800.1, 2006]. The ns2voip [Bacioccola et al., 2007] module was used, since it enables to emulate a realistic traffic model for VoIP application and it also calculates this parameter.

The routing process is evaluated by measuring the path characteristics that can assess its level of stability. The stability evaluation parameters are obtained from an average of all the flows for all the load configurations employed. Two stability parameters were used to evaluate the routing oscillation, namely the prevalence and routing flap [Ramachandran et al., 2007]. The routing table is stored in a text file every 5s to measure the prevalence and the routing flap. The interval time was defined in accordance with the update interval of the topology controls. The evaluation parameters for stability are described as follows:

- **Prevalence**. The prevalence is based on the dominant route concept. A dominant route is the route that is most often observed in the set of routes. The prevalence pre_d of the dominant route is calculated in Equation 4.37.

$$pre_d = \frac{k_p}{n_p} \tag{4.37}$$

  where n_p is the total number of times that any route was available in the set of routes computed and k_p is the number of times that the dominant route was used in the set of routes. The set of routes contains all the calculated routes that are found during the simulation time.

- **Routing Flap**. This stability parameter is the number of route changes for a given source-destination pair.

As an illustration of these parameters, suppose a simulation time of 30s, where the set of routes $S = \{A, B, A, A, B, C\}$ is calculated during the whole simulation. In this example, the routing flap is equal to 4. Route A is the dominant route, since it appears more often than B (twice) and C (once) routes. In this context, the prevalence is calculated as follows:

$$pre_d = \frac{3}{6} = 0,5(50\%) \tag{4.38}$$

4.5.4.3 Simulation Results

Figures 4.28 and 4.29 show the lowest throughput and highest delay when MIC is used. This routing metric does not measure interference in a dynamic way and does not take into account the traffic load. In addition, MIC employs the expected transmission time metric (i.e., ETT) as a sub-component that overestimates the link quality. Although ILA and CATT are also based on the expected transmission time, these metrics result in higher throughput and lower delay than MIC, since they are able to pick up interference and traffic load. CATT takes into account the influence of the interference and traffic load weight of the 1 and 2 links away from a link, whereas ILA is only concerned with this influence when it is 1 link away. As a result, the traffic achieves its better performance when CATT is used.

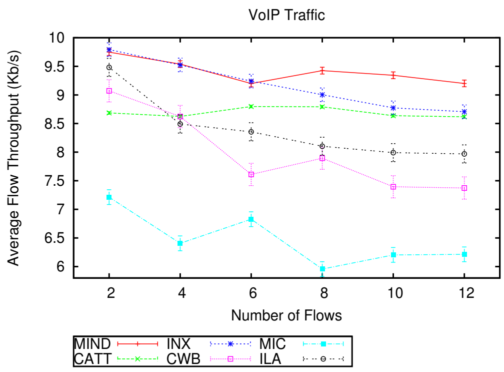

Figure 4.28: Average Flow Throughput

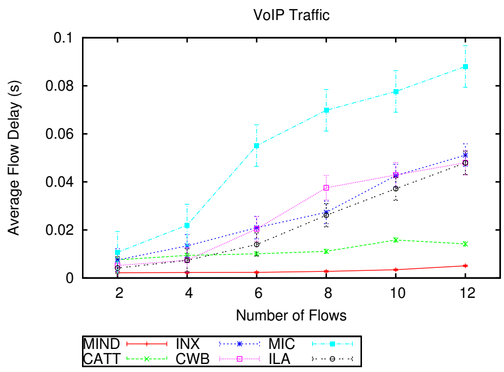

Figure 4.29: Average Flow Delay

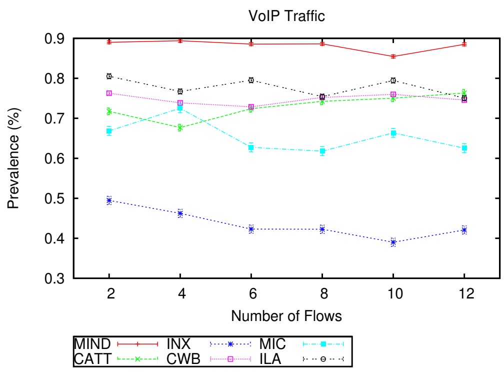

Figure 4.30: Average Flow Prevalence

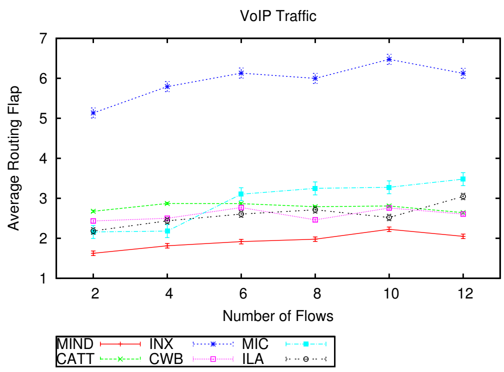

Figure 4.31: Average Flow Routing Flap

Although INX does not take into account traffic load and interference in a dynamic way through the physical model, INX achieves higher throughput than MIC, ILA and CATT, by refraining from using the ETT. However, when INX is used, the average flow throughput decreases in proportion to the increase of load configurations (i.e., number of flows), because it is based on a packet loss ratio measurement that is not sensitive to the increase in traffic load. However, even though INX achieves a high throughput, it does not result in a low delay, since it is not based on a delay measurement. For example, INX results in a delay that is very similar to ILA and CATT, but ILA and CATT achieve a lower throughput than INX.

MIND and CWB are metrics that employ both a logical and physical model to depict interference, while using the same measure to depict both logical interference and traffic load (i.e., channel busy time). CWB achieves a worse traffic performance than MIND, since it uses measurements of interference that does not depict interference precisely when the network conditions are rapidly changing during the time period (i.e. frame error rate). As a result, it was also expected that CWB would have a lower throughput than ILA and CATT in medium and high loads, where the network conditions can vary quickly. MIND and CWB behave in an unstable way when throughput is analysed, i.e. in MIND and CWB result in peaks in 2 and 8 flows where the throughput decreases from 2 up to 6 flows, then increases at 8 flows and again starts to decrease at 10 and 12 flows. This is because the channel busy time measure used in CWB and MIND regards the backoff time as an idle period of the channel and consequently, the routing algorithm can select links with higher logical interference.

Figures 4.30 and 4.31 provide evidence that all the studied routing metrics are consistent when they are analysed for different stability routing parameters. It is also worth noting that the stability level has an effect on the network performance parameters. However, this impact is less noticeable in the case of some routing metrics, such as INX and CWB. INX has the worst stability of all the evaluated loads, because it is mainly based on the measurement of packet losses that is very sensitive in highly variable conditions and does not employ any stability mechanism. Nevertheless, INX achieves a higher throughput than ILA, CATT, CWB and MIC routing metrics in all the load configurations. In addition, CWB achieves stability levels that are very similar to ILA and CATT in all the load configurations, since the frame error rate takes a long time to pick up changes in the degree of interference. Despite this, ILA and CATT result in a higher throughput than CWB in almost all load configurations. In view of this, it provides better routing stability and a worse traditional network performance, owing to its inability to pick up the interference level precisely. The high level of routing instability provided by the routing metrics does not always imply a worse performance while, at the same time, a low level of routing instability does not always mean a better performance. Hence, the traffic performance is mainly influenced by the measurements that are combined in the routing metrics rather than the routing instability aspect.

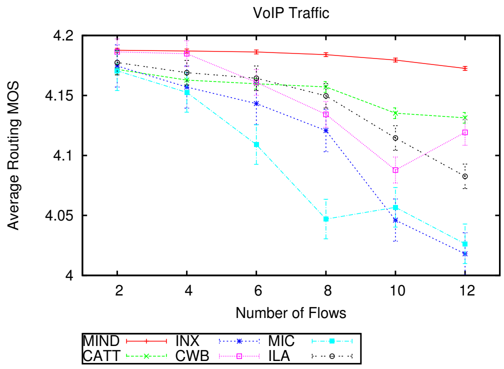

Figure 4.32: MOS of Talkspurt periods

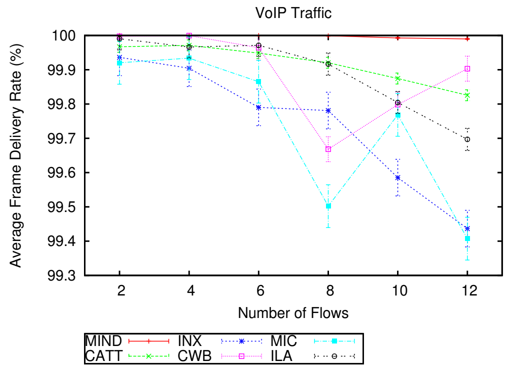

Figure 4.33: Average Flow Frame Delivery

Figure 4.32 shows that the routing metrics have a different impact on the user level from at the network level evoluation parameter. For instance, the INX routing metric results in a higher throughput than MIC, ILA, CWB and CATT in all the load configurations. However, when INX and CWB are used, the MOS values are the lowest for high loads (10 and 12 flows). The delay and frame losses are the main parameters that explain the VoIP quality achieved by INX and CWB. Moreover, these results are also due to the simple static buffer with a fixed length. Thus, the received frames can be either dropped or accommodated in the buffer, depending on whether the playout time has expired or not. Furthermore, it should be noted that the loss of a single packet can cause the loss of an entire VoIP frame. Although the remaining packets of this frame are taken into account for the throughput calculation, they do not influence the assessment of VoIP quality. In fact, the MOS parameter is affected more by the frame losses and delay than by the throughput measurement (see Figure 4.33). For example, the load configuration of 10 flows for MIC, INX and CWB are evidence that delay and frame losses can also have an impact on the MOS parameter. In this case, CWB also has a higher MOS than INX even though INX results in a very similar delay to that of CWB; this is because INX has more frame losses than CWB. In addition, the frame losses are very similar when CWB and MIC are used, even if CWB achieves higher MOS. The reason for this is that the delay in 10 flows is very high when MIC is used.

4.6 Summary

In this chapter, there has been an in-depth investigation of cross-layer routing metrics. First of all, the cross-layer concept and the kind of interactions it involves, were described to emphasize its importance for the WMN. Following this, a new taxonomy of the specific measurements taken in the cross-layer routing metrics was set out as the main contribution of this chapter, which seeks to provide an in-depth understanding of the main features of this important subject. After this, a survey of several cross-layer routing metrics was conducted which drew attention to some key issues. It was found that the most serious limitation of the current cross-layer routing metrics was the lack of coordination between the most precise measurements that were used to pick up interference and traffic load. To overcome this drawback, the MIND cross-layer routing metric was proposed as a means of combining the most accurate measurements of traffic load, physical and logical interference (i.e. inter-flow interference), while using passive mechanisms to obtain the cross-layer measurements. MIND also took full account of the intra-flow interference. Extensive simulation results showed that MIND can bring about considerable performance improvements for WMN in diverse conditions, such as different types of scenarios, traffic patterns and applications. The results demonstrated that MIND improves scalability with regard to triple play services, and outdoor and indoor environments as well as when traffic patterns are deployed for the client and WMN infrastructure. Furthermore, other factors were observed, such as the results of these simulation tests which are as follows: most of the assessed cross-layer routing metrics resulted in a different performance which varied in accordance with the environmental conditions (i.e., whether it was indoors or outdoors). Although cross-layer routing metrics affect the performance of both the network and the user levels, there are some cases where they have less impact on the latter, because the user perception parameters are less influenced by the behaviour of the network. The results also demonstrate that the high level of routing stability attained with some cross-layer routing metrics do not imply there is a good traffic performance and vice-versa. Hence, these aspects are not closely interdependent.

Chapter 5
Clustering Approach

Finding low-cost solutions to build up Wireless Mesh Networks (WMN) can help provide coverage for larger areas. However, as was shown in Chapter 2, the existing routing protocols for WMN have scalability limitations, since they cause a large routing overhead and intolerable delay in the presence of a large number of nodes [Yu and Chong, 2005, Woo and Singh, 2001]. Consequently, the network performance degrades significantly when the size of the WMN increases. As a result of this, the routing protocols may also not be able to find a reliable routing path, causing the loss of transport-level connections. In addition, the MAC protocols may experience a significant reduction in throughput. In the light of this, clustering approaches have been employed to mitigate the overhead and delay of the routing process in WMN. This chapter addresses clustering in WMN and is structured as follows: there is an overview of clustering in Section 5.1 where the main advantages of clustering are described. Section 5.2 analyses a taxonomy of clustering for MANET and proposes a taxonomy for the classification of clustering in WMN. The most relevant clustering approaches for WMN are outlined in Section 5.3. A clustering approach, called Clustering Approach for Routing MAnagment (CARMA), which is designed and proposed for the ACRoMa architecture is investigated in Section 5.4 and it is shown that CARMA consists of the mesh traffic migration method, the Collaborative CLustering Scheme (CoCLuS) and Routing Algorithm for Inter-cluster Load Balancing (RAILoB). Section 5.5 describes the simulation study that validates CARMA. Section 5.6 summarizes the findings of this chapter.

5.1 Overview of Clustering

Clustering provides a virtual and hierarchical structure for WMN in a partitioned way by dividing the WMN into different virtual groups. The nodes are allocated geographically alongside the same cluster in accordance with specific rules. In this structure, the nodes may have different functions, such as clusterhead, gateway, or member. A clusterhead serves as a local leader for its cluster, and performs intra-cluster transmission arrangements and data forwarding. A gateway node is a non-clusterhead node with inter-cluster links, which means that it can access neighboring clusters and forward information between the clusters. A member is a non-clusterhead node without any inter-cluster links, but only intra-cluster links. In addition, there are also approaches that assume that the clusterhead and gateway are the same entity in the clustering structure and others that the gateway or clusterhead cannot exist [Yu and Chong, 2005].

A clustering diagram can be seen in Figure 5.1.

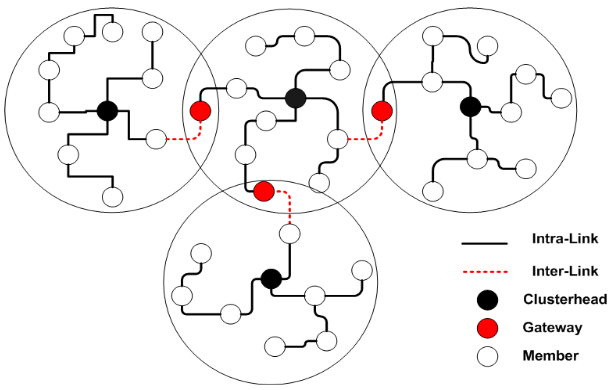

Figure 5.1: Example of a Cluster Structure - Adapted from [Yu and Chong, 2005]

Clustering can provide the following benefits to WMN:

- **An ability to increase the system capacity**. It allows a spatial reuse of resources. In other words, with the non-overlapping multicluster structure, two or more clusters can use the same frequency thus allowing simultaneous data transmissions.

- **An ability to distribute and exchange routing information**. The clusterheads and gateways can become a virtual backbone for inter-cluster routing, and thus the generation and propagation of routing information can be more restricted than in the reactive and proactive routing protocols.

- **Smaller and more stable structure**. When a node (that is strictly a member node) changes to another cluster, only the nodes that are in the corresponding clusters need to update the information. This means that local changes do not have to be spread and updated by the entire network, and there is a considerable reduction of the amount of information as such, that is processed and stored by each node.

Some key features of self-organizable systems [Prehofer and Bettstetter, 2005, Tang and Tianfield, 2006] can be seen in this scheme such as the following:

- **Self-configuration**. The clustering structure is formed automatically on the basis of certain criteria. Moreover, new elements can be automatically configured and integrated into the network.

- **Self-optimization**. Parameters of a clustering element can be frequently adjusted according to the status of some neighbouring elements. Thus, all the elements cooperate to achieve a common objective.

- **Self-healing**. Owing to the redundancy of entities, elements which failed such as clusterhead and gateways, can be easily replaced.

5.2 Taxonomy for Clustering Approaches

Clustering has been employed in MANET [Yu and Chong, 2005], and are classified according to several criteria (e.g., clusterhead or not, 1-hop or Multi-Hop) as shown in Figure 5.2.

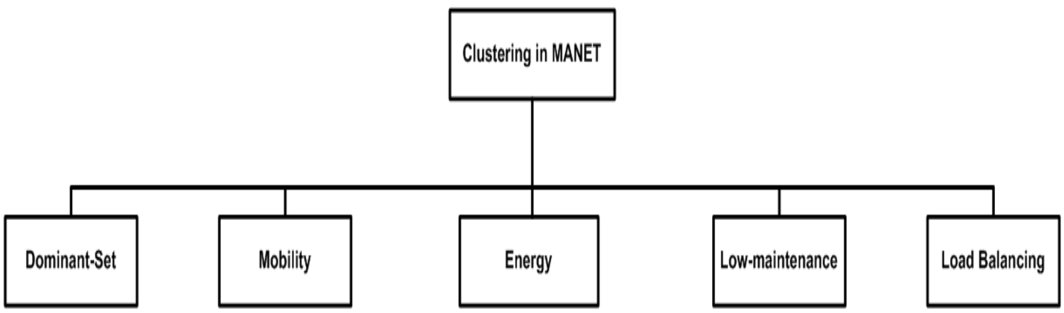

Figure 5.2: Taxonomy for Clustering in MANET - Adapted from [Yu and Chong, 2005]

Yu and Chong (2005) classify the clustering approaches in a way that can meet their objectives and thus they can be grouped into five categories, as follows:

- **Dominating-Set-based**. This seeks to find a set of nodes, called Dominating Set (DS), for MANET to reduce the routing information.

- **Low-maintenance**. This provides a cluster infrastructure for upper layer applications with reduced cluster-related maintenance costs (e.g., fewer reclustering situations and a reduced number of explicit control messages). Thus, the cluster can be maintained in a more economical way.

- **Mobility-aware**. This group the nodes in a clustering scheme on the basis of mobility, since mobility is a characteristic of MANET and the main reason for making changes to the network topology.

- **Energy-aware**. Energy consumption is a crucial factor in MANET and this scheme is designed to manage the energy consumption of the nodes.

- **Load-balancing**. It distributes the workload into clusters more fairly.

Clustering has also been adopted in WMN to handle scalability in the routing process. However, the taxonomy which classifies clustering in MANET is not suitable for systematically organizing the clustering approaches, since it fails to take into account the main features of WMN. An example of this is traffic patterns and fewer restrictions on the consumption of energy. This is one reason why a new taxonomy for clustering in WMN is also proposed in this thesis.

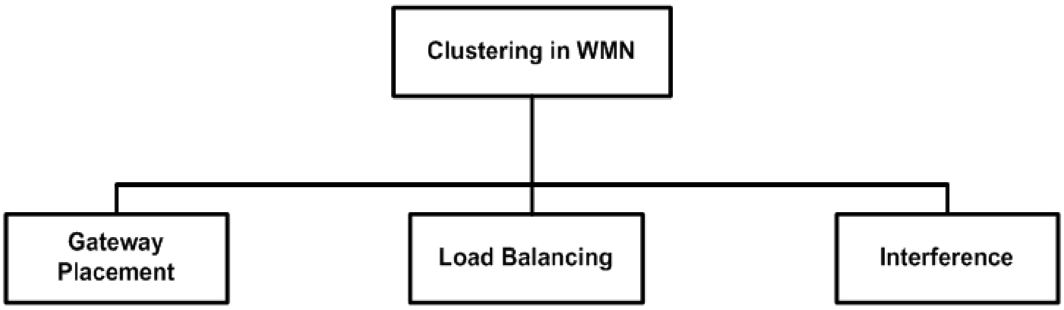

Figure 5.3: Taxonomy for Clustering in WMN

Figure 5.3 illustrates the taxonomy for a clustering in WMN. This proposed taxonomy groups the solution of clustering into three categories, which are as follows:

- **Gateway Placement**. This category is concerned with how to deploy the gateways in order to balance performance and cost.

- **Load-balancing**. This is very similar to the load balancing category in the MANET taxonomy. However, the typical traffic pattern of WMN requires the clustering approaches to be designed differently.

- **Interference**. This category divides the network into clusters so that the interference levels are mitigated.

5.3 Related Work on Clustering Approaches

The most relevant approaches that are based on clustering for WMN, are outlined in this section. The approaches are grouped into three sub-sections, Gateway Placement (sub-section 5.3.1), Load Balancing (sub-section 5.3.2) and Interference (sub-section 5.3.3).

5.3.1 Gateway Placement

The gateways are the data output and input points of a WMN, since they provide the Internet access. This means that the traffic performance would be improved if more gateways were placed in the WMN [Aoun et al., 2006, Bejerano, 2004, He et al., 2008]. However, an increase in the number of gateways can also lead to an increase in costs, since the gateways have special requirements (e.g. Internet access by wired network). Hence, the main challenge of gateway placement is to determine the minimum number of gateways that meet the QoS constraints with regard to the service provider and as a result, keep costs under control.

5.3.1.1 Efficient Integration of Multi-Hop Wireless and Wired Networks with QoS Constraints

Bejerano et al. propose a clustered algorithm for WMN that selects a set of nodes for the gateway functions [Bejerano, 2004], called as Iterative Greedy Dominating Set (IGDS). It also uses a spanning tree rooted at each clusterhead (i.e., gateway) for message delivery. In this way, the gateways are subject to three constraints, cluster radius (maximum depth of the spanning tree), cluster size (number of nodes inside the cluster) and relay load (traffic load aggregated and forwarded by the intermediate nodes). It should be noted that these constraints have an impact on the throughput and delay, which explains why they are described as the QoS constraints. The advantage of this approach is that it allows the gateway placement to take the QoS constraints into account.

This proposal breaks the problem of clustering and ensuring QoS into two sub-problems. The first seeks to find a minimal number of disjoint clusters' containing all the nodes subject to an upper bound on clusters' radius. The second one considers placing a spanning tree in each cluster, in which clusters that violate the relay load or cluster size constraints are further subdivided. Nevertheless, there is a problem in the IGDS [Bejerano, 2004], namely when the cluster radius is large enough to accommodate a large number of nodes in the initial clustering process. Thus, at a later stage, whenever various constraints are imposed, the IGDS subdivides the clusters to satisfy the constraints. As a result, a large number of small clusters is obtained. In addition, this approach is an intra-cluster solution that does not take interference into consideration and since it is validated through a conceptual evaluation, it is not evaluated with any traffic or routing protocol. For this reason, it is difficult to assess the impact of the IGDS on the traffic application.

5.3.1.2 Gateway Placement Optimization in Wireless Mesh Networks with QoS Constraints

The Gateway Placement Optimization (GPO) approach also aims to calculate the optimum number and position of gateways in a WMN [Aoun et al., 2006]. Translating this problem to the clustering scheme, the clusterheads will be linked to the gateways. Thus, this problem is specified by an Integer Linear Problem (ILP) in which a way to achieve the best outcome (such as the maximum profit or the lowest cost) is subject to a list of constraints represented as a linear equation [Gerla and Tsai, 1995].

The GPO algorithm extends the IGDS algorithm to enhance the excessive subdivision of the clusters, which are reduced through the recursive algorithm proposed in [Aoun et al., 2006] by means of which the nodes are attached to the generated clusters that are constrained by the QoS requirements, and can otherwise be considered as new clusterheads. The authors also consider the cluster radius, cluster relay load and cluster size as QoS constraints. The main advantage of the proposed recursive algorithm compared to the IGDS algorithm, is that clusters have the chance to merge with other clusters at earlier iterations where they can form feasible clusters that satisfy all the QoS constraints.

The GPO algorithm employs a spanning tree rooted at the gateway for forwarding intra-cluster traffic. In view of this, GPO is neither concerned about the load balancing between the different clusters (i.e. inter-cluster routing) nor interference. In addition, the evaluation of GPO only consists of a comparison with IGDS that takes into account the number of clusters, since GPO aims to reduce the number of clusters. Hence, it fails to carry out a more realistic evaluation that involves a traffic model and routing process.

5.3.1.3 Optimizing deployment of Internet gateway in Wireless Mesh Networks

He et al. address the gateway placement in a WMN from two novel aspects [He et al., 2008]: (a) modeling the throughput capacity of mesh routers and gateways with Multi-Channel Multi-Radio (MCMR) capability, and (b) proposing two heuristic algorithms to minimize the number of gateways that are subject to various constraints, such as, full coverage, gateway and mesh router throughput capacity, co-channel interference. These aspects can be translated in the following constraints: size, relay load and radius. In other words, this approach extends GPO and IGDS by considering the previously identified characteristics.

Two algorithms were conceived in this approach to reduce the number of gateways and position them, while satisfying the QoS constraints. The first algorithm, called degree-based Greedy Dominating Tree Set Partitioning (GDTSP), emphasizes the connectivity degree of the gateway. This refers to the fact that, the degree-based GDTSP algorithm computes all degrees for every node in the graph and the node which has the highest connectivity degree is selected as the gateway. The second algorithm, called weight-based GDTSP, is based on the connectivity weight of each node in its *R-hop zone* (i.e., nodes that are neighbours, if they are within the R hop range). The node with the largest connectivity weight is selected as the gateway. The connectivity weight of a node is calculated as follows:

$$W(v_i, R) = \sum_{\forall v_j \in N_R(v_i)} \frac{1}{Hop(v_i, v_j)} \tag{5.1}$$

where $N_R(v_i)$ is the set of nodes that are neighbours of v_i in the graph; $Hop(v_i, v_j)$ is the shortest distance in number of hops between node v_i and node v_j in the original graph. $W(v_i, R)$ measures the path length from all the other nodes to a specific node v_i. Thus, if the number of hops increases, its connectivity weight decreases. The weight-based GDTSP algorithm reduces the MR-IGW hops, that is, the number of transmission hops between the mesh router and the gateway.

In both the proposed algorithms, the graph is divided into disjoint clusters (i.e., there is no inter-cluster communication) that satisfy the QoS constraints and make use of three interference models. However, the employed interference models do not depict interference in a dynamic way, since the main factors that determine this physical phenomenon in the WMN are not taken into consideration, such as changing traffic patterns, multi-hop routing and the physical conditions of the environment. In addition, the evaluation of the GDTSP algorithms is similar to the GPO and

IGDS evaluation, i.e. since it does not include a more realistic WMN. Nonetheless, the proposed approaches in [Aoun et al., 2006, Bejerano, 2004, He et al., 2008] do not consider that the gateways can be congested, on account of the traffic that is going through the gateways. This means that all these approaches only consider intra-cluster routing.

5.3.2 Load Balancing among Gateways

Some consideration should be given to load balancing in WMN, in particular among gateways that are the main WMN bottlenecks. This is because the load balancing among gateways improves the network capacity by avoiding congestion. There have been several research studies on this and these are examined in subsections 5.3.2.1 and 5.3.2.2. All these approaches are based on a single load balancing method which is mesh router migration. This method is defined in this sub-section, but will be described in greater detail in subsection 5.4.1.

5.3.2.1 Load-balanced Mesh Router Migration for Wireless Mesh Networks

Xie et al. propose a load-balancing approach among gateways that also provides Internet mobility for mesh clients [Xie et al., 2008], called the Load Balancing Approach (LBA). This approach divides the networks into several domains (i.e., clusters), where the clusterhead also has a gateway role. In addition, there are two algorithms, called the initial domain partitioning and load adjustment algorithms. In the proposed approach, the migration of a mesh router to a domain is defined by means of two key metrics, hop count and bandwidth utilization of a domain. In the initial partitioning algorithm, the hop count metric defines which mesh routers should be assigned to the nearest gateway.

The load balancing among gateways is carried out by the load adjustment algorithm that determines the mesh router migration. This migration is performed when a specific domain is identified by the second metric as having a high load. In the case of an overloaded domain, the migration of the mesh router with the lowest hop count value to a gateway, in a domain with a lower load, has a higher priority. The gateway capacity follows the tie-breaking criteria when some mesh routers have the same hop count. The bandwidth utilization of the routers of a domain is employed as the load metric (i.e., second metric), which is the percentage of used bandwidth compared with the total available bandwidth of a domain. Each domain calculates its bandwidth utilization in real time and periodically exchanges this information with its all neighbouring domains. Thus, two neighbouring domains can estimate the utilization differential between themselves. Unlike the previous clustering approach, LBA is validated trhough a more realistic simulation model of WMN, which involves a traffic model and a routing protocol.

The interference is not considered to be a clustering metric and therefore, the calculated available bandwidth can be overestimated. In addition, this approach prioritizes the hop count as the main metric required to make the mesh router migration. As a result, a node with lower load can be migrated instead of a node with higher load. Another factor is that the hop count metric does not depict the link quality.

5.3.2.2 Load Balancing Routing for Wireless Mesh Networks: An Adaptive Partitioning Approach

The Partition-based Load Balancing (PLB) approach divides the network into several clusters to provide load balancing routing among gateways (i.e., sink nodes) [Choi and Han, 2010]. Unlike the LBA approach [Xie et al., 2008], PLB considers the load metric for the mesh router migration. A tree-style single path routing is assumed, where each gateway is considered as the root. The number of trees (i.e., clusters) is equal to the number of gateways. The links between a gateway and its neighbouring nodes are termed top sub-links and the gateways' neighbours are called top sub-nodes. PLB enables load balancing between clusters (inter-cluster) and within clusters (intra-cluster). The PLB has three algorithms that are defined as follows: Load-Adaptive Clustering phase (LAC), Inner Domain Load Balancing phase (IDLB) and Outer Domain Load Balancing phase (ODLB). The LAC algorithm is executed first, followed by ODLB. IDLB is executed both during LAC and ODLB.

The LAC algorithm is responsible for partitioning the entire network into domains, using hop-count and network load as the clustering metrics, and also performing mesh router migrations. Each mesh router is combined with a weight which is the number of session requests (i.e. flows) from users that are attached to this mesh router. It is assumed that each user will request one session with the same traffic load. The IDLB algorithm balances the load in the intra-cluster nodes that have multiple downlinks, called division points, from the gateway to boundary nodes, in an iterative way. At each round, a division point selects two adjacent sub-links from its downlinks. Following this, it calculates the difference between their cumulative loads, which is called Imbalance Difference (ID). Finally, the division point finds the nodes whose cumulative load is closest to the ID among the heavier sub-tree, and changes the uplink of the selected node to the lighter subtree. In addition, the ODLB algorithm solves the unbalanced load cases through the inter-cluster load balancing that the IDLB algorithm is unable to solve at the intra-cluster level.

The PLB algorithms overestimate the traffic load in the network, since the number of sessions does not represent channel occupancy in an accurate way. In other words, the traffic load of the user requests may be different, due to the fact that different applications require distinct data rates such as video streaming and VoIP. The PLB approach assumes that there is a well-defined MAC protocol for employing MCMR capability between nodes, which means the interference issue is not considered. However, interference is a serious problem in wireless networks, including WMN, and even though the MAC protocol is well defined, it remains present. For example, it might be external interference or the fact that the current MAC protocols have a restricted number of channels. Both the PLB and LBA approaches can again overestimate the link quality when they only use the number of sessions and hop count metrics. Furthermore, PLB does not evaluate the impact of the mesh router migration method on the traffic performance, as it only involves a conceptual evaluation where PLB and LBA are compared through a fairness index parameter.

5.3.3 Interference

Although several proposed technologies for wireless networks have reduced the interference levels, such as directional and smart antennas, interference still persists and continues to degrade the performance of the wireless networks. This sub-section outlines a clustering approach to address this phenomenon.

5.3.3.1 Mobility-Aware Clustering Algorithms with Interference Constraints in Wireless Mesh Networks

There are some approaches that have investigated the problem of gateway placement in WMN, such as [Aoun et al., 2006, Bejerano, 2004, He et al., 2008]. These approaches divide the network into a minimum number of disjoint macro-clusters, where each macro-cluster is assigned to a clusterhead node (i.e., gateway) that connects directly to the wired network. Langar et al. focus on clustering algorithms that divide the macro-cluster into virtual micro-clusters (sub-clusters) with the aim of reducing the Radio Resource Utilization (RRU) cost in the WMN, and then maximizing network throughput [Langar et al., 2009]. This approach is called Virtual

Micro-Clusters (VMC). The RRU cost of a mesh client comprises two components - the resource utilization for the data packets and the resource utilization for the signaling messages that are used for managing user mobility. This approach considers the properties of user mobility as well as the effects of interference between links.

In this proposal, the clustering configuration adopted is defined as follows: a node serves as a clusterhead and it operates as an intermediate node between the gateway and the mesh routers inside the cluster. The clusterhead replaces the gateway inside the cluster and manages the mobility of local users. Thus, the signaling messages are reduced when there is mobility of the mesh clients to a local area (i.e., inside the cluster). As a result, this configuration reduces the RRU cost and improves the network performance. In the light of this, VMC proposes two clustering algorithms to reduce the RRU cost, Optimal Static Clustering (OSC) and the Distributed Clustering Algorithm (DCA).

The OSC algorithm assumes that clusters are static and disjoint, and that the micro-cluster placement can be formulated as an integer linear programming problem that optimizes the RRU cost. In the second algorithm, the clusters may overlap and cluster placement is carried out in a distributed manner. The OSC algorithm has larger time complexity than DCA, due to the time-consuming resolution of the associated ILP problem. In addition, both the proposed clustering algorithms take into account the interference effect among the neighbouring links during the cluster process (i.e., cluster formation). To reach this, an interference-aware routing metric for the clustering process, called Interferer Neighbors Count (INX) was used (described in the previous chapter).

The INX value of a link (v, u) is defined as the product of the Expected Transmission Count (ETX) [Couto et al., 2003] of the link (v, u) by the total number of interferer links resulting from a transmission on that link. The INX routing metric does not consider interference in a dynamic way, which is a limitation, since the interference can change over time due to signal strength variations and the amount of traffic generated by the interfering nodes. Both clustering algorithms employ the INX routing metric to form the clusters. However, the proposed algorithms do not consider inter-cluster routing for load-balancing among macro-clusters, since the load balancing achieved by OSC and DCA is at a micro-cluster level.

5.3.4 Comparison among Related Works in Clustered WMN

Table 5.1 summarizes the related work on clustering approaches in WMN, on the basis of their main characteristics, such as gateway placement, load balancing,

Table 5.1: Comparison among Related Work in Clustered WMN

Related Works	Gateway Placement	Interference	Inter-Cluster Routing	Clustering Scheme	Algorithm Type	Emulated Traffic Model
IGDS	Yes	No	No	QoS Constraints	Centralized	–
GPO	Yes	No	No	QoS Constraints	Centralized	–
GDTSP	Yes	Yes	No	QoS Constraints	Centralized	–
LBA	No	No	Yes	Load Balancing	Centralized	CBR
PLB	No	No	Yes	Load Balancing	Centralized	–
VMC	No	Yes	No	QoS Constraints, Macro- and Micro-clusters	Centralized and Distributed	CBR

and interference. The analysis of this Table shows that the existing solutions only address some of the distinct features and fail to provide the gateway placement and interference awareness while enabling the load balancing to occur among gateways in a single approach. For example, the proposed approaches in [Aoun et al., 2006, Bejerano, 2004, He et al., 2008] present very different characteristics from the approaches in [Kim et al., 2008, Ma and Denko, 2007]. Nevertheless, VMC combines the gateway placement and interference awareness features, although the cross-layer routing metric used in VMC has some limitations and does not depict interference precisely. The interaction between load balancing and interference awareness is an important synergy to provide a more scalable solution for WMN, even though the related work lacks solutions that can be used to explore this interaction. Furthermore, most of the related work employs a conceptual evaluation of their approaches without taking into account a traffic model or routing process. Only VMC and LBA emulate a simple traffic model and routing process.

5.4 Clustering Approach for Routing MAnagement (CARMA)

The overall objective of this section is to set out CARMA which is composed of the mesh traffic migration method, the Collaborative CLustering Scheme (CoCLuS) and the Routing Algorithm for Inter-cluster Load Balancing (RAILoB). The next sub-sections will clarify what this approach entails. It deals with open issues arising from the mesh router migration method [Xie et al., 2008][Choi and Han, 2010]. Sub-section 5.4.1 describes the main drawbacks of the mesh router migration method which is employed in the PLB approach. Sub-section 5.4.2 provides an overview of the mesh traffic migration method, while also describing the main similarities and differences with the mesh router migration method. Sub-section 5.4.3 presents CoCLuS. The routing algorithm for inter-cluster load-balancing will be explained in sub-section 5.4.4.

5.4.1 Problem Description

The network model for the mesh router migration method that is employed for WMN can be outlined as follows: WMN are represented by graph *G(V, E)*, where V is the set of nodes and E is the set of links. V is divided into two subsets I and R, which are a set of *I*nternet gateways (referred to as gateways from this point onwards) and mesh *R*outers, respectively. The mesh clients are not included. There is a link between two nodes if the nodes are within each other´s transmission range. A grid topology was defined to limit the maximum number of neighbour mesh routers. It is assumed that the gateway placement is well defined.

Mesh routers form a tree structure that is used to communicate with the gateway. Every cluster has a single gateway which is located in the central position of the cluster. In this way, the network is partitioned into clusters in which the root is a gateway. Each mesh router is characterized by its weight which depicts the load level and is usually represented by the number of active flows. These flows are normally derived from mesh clients which attach themselves to the mesh router. The *Cumulative Load* (CL) is the sum of the weights of all the nodes in the sub-tree, including the weight of the root. Thus, the CL of a node is the number of uplink traffic incidents on the node. The links between a gateway and its neighbouring nodes are called *Top Sub-Links* (TSL), and the neighbours that are one hop from the gateways are called *Top Sub-Nodes* (TSN). A TSN of an adjacent cluster is called an adjacent TSN. The overload condition occurs when the Cumulative Load of TSN exceeds the defined maximum load threshold.

Mesh router migration is a topological change of the clustering structure, which allows the migration of mesh routers, as well as their attached application traffic, from a heavily-loaded cluster to a lightly-loaded cluster. Mesh router migration only occurs between adjacent clusters. In particular, only the mesh routers which are border nodes (i.e. nodes which have connectivity with nodes in an adjacent cluster) are able to migrate. Moreover, mesh router migration only occurs when there is an overload condition in any TSN of a cluster.

Figure 5.4 shows the network model used in the mesh router migration method. The network is indicated as a matrix *M*, where *x* is the *x-axis* index and *y* is the *y-axis* index. It should be noted that both *x-axis* and *y-axis* start from the upper-left corner position instead of the lower-left corner position, which is commonly used in matrices notation for mathematics. Moreover, *m(x,y)* is used to refer to the elements of matrix *M*. We also assume that the load threshold of TSN is 4, thus *m(3,2)* which is on the right-hand side of *G1*, is overloaded (i.e. CL with value 5). The numbers in the squares correspond to the weight of each node. The numbers which are alongside the Top Sub-Links are the Cumulative Load in the TSN sub-tree. The gray arrow illustrates an example of the application traffic coming from *m(3,4)* and going to *G1* passing by *m(3,2)*.

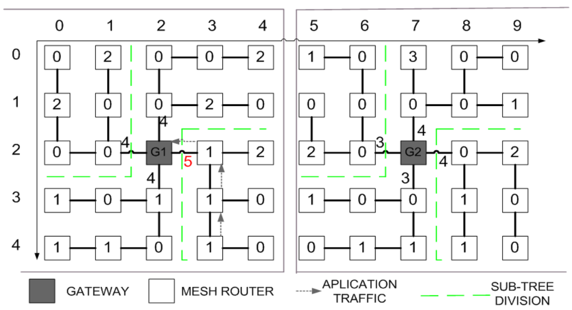

Figure 5.4: Mesh router migration: PLB approach

Following this, the gateway with the role of a clusterhead chooses one candidate mesh router for defection (i.e. migration). The candidate is defected only if the sum of the CL of the adjacent Top Sub-Nodes and the CL of the candidate does not exceed the defined maximum load threshold. In this process, the border nodes with lower CL are more likely to be selected. After this, the gateway sends a defection request message to the candidate node which forwards this message to the adjacent TSN. When this defection message arrives at the adjacent TSN, it is checked to determine whether or not the candidate can be accepted, in accordance

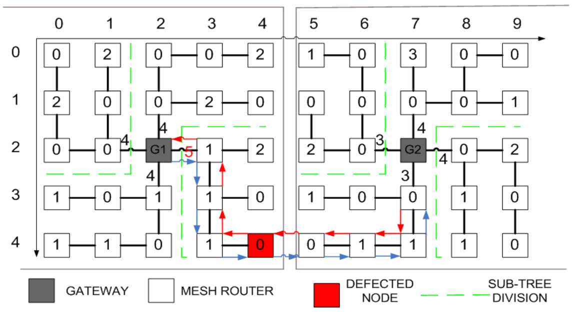

Figure 5.5: Mesh router migration: Step 1

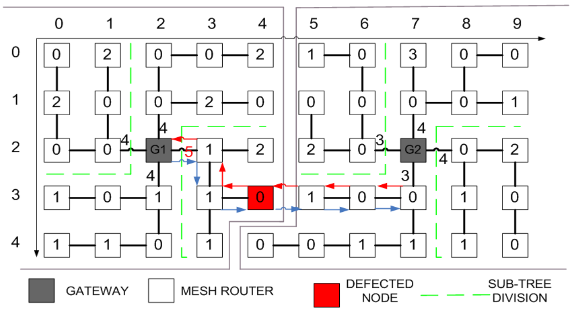

Figure 5.6: Mesh router migration: Step 2

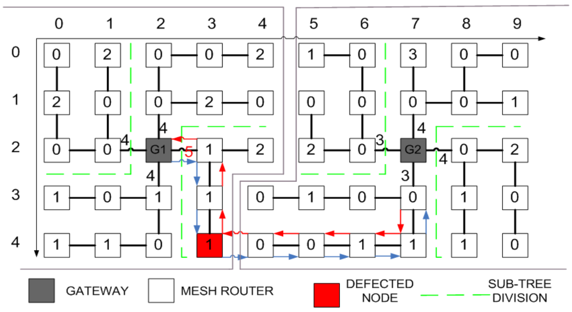

Figure 5.7: Mesh router migration: Step 3

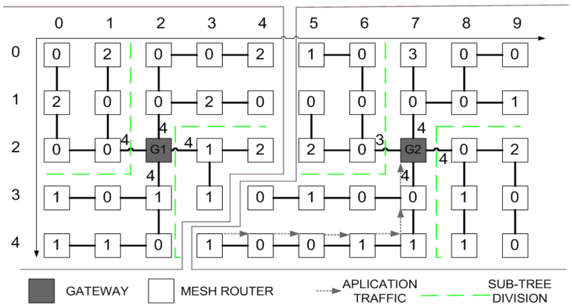

Figure 5.8: Mesh router migration: Step 4

with the conditions described earlier. Then, the adjacent TSN sends back a defection response message to the candidate and to the gateway, notifying them of its defection decision. Figures 5.5, 5.6, 5.7 and 5.8 provide a step-by-step illustration of how the mesh router migration method works.

First, *m(4,4)* is migrated (Figure 5.5), since it is a border node and has one of the smallest CL. Next, *m(4,3)* is also migrated (Figure 5.6) since it is the next border node which has one of the smallest CL. However, they do not help to improve the load balancing of the network, since it actually has no traffic load. It should be noted that *m(3,4)* is a better candidate to make the load balancing more efficient, but is not yet a border node in Figure 5.5. Hence, *m(3,4)* has to wait to become a border node with smallest CL, which occurs when *m(4,4)* and *m(4,3)* migrate to the adjacent cluster (Figure 5.7). Figure 5.8 shows the balanced clusters *G1* and *G2* after the migration of three mesh routers. It is important to point out that the clustering structure was modified by the migration process.

The messages required by this method are illustrated in Figure 5.5, which can also be used to illustrate this. The *G1* gateway sends the defection request message (blue arrow) to *m(4,4)* which then forwards it to *m(7,3)*, the adjacent TSN. When *m(7,3)* receives this message, it sends back a defection response message (red arrow) to the *G1* gateway and *m(4,4)* to confirm the acceptance status of *m(4,4)*. The defection decision could have been made locally at *m(4,4)*, if the nodes had had the information about the CL of the TSN in the adjacent clusters. In this case, *m(4,4)* would not need to forward the defection request message to *m(7,3)* and thus, could reduce the time needed to make the inter-cluster routing decision.

The mesh router migration method may take a long time to achieve load balancing between the gateways, due to the limitations discussed above. In the first place, the mesh routers which are selected to migrate may not be border nodes at the time when the overload condition occurs. Since mesh router migration can only occur when it includes border nodes, this solution must wait until this condition occurs. In addition, the mesh router migration uses an on-demand (i.e. reactive) strategy to enable the mesh routers to migrate and thus requires more time. This means that the mesh router migration always has to send messages to find out the load level of the adjacent TSN (i.e. reactive stragegy), since the mesh routers do not have a priori information about the load of the adjacent TSN.

5.4.2 Mesh Traffic Migration - Proposed Migration Method

This sub-section describes the mesh traffic migration method [Borges et al., 2012c] and points out its main similarities to and differences from the mechanisms associated with mesh router migration. The main reason for this mesh traffic migration is that the mesh router migration is very slow to make the inter-cluster load balancing. In view of this, the mesh traffic migration method only seeks to migrate the traffic load between the clusters, instead of migrating traffic and nodes, as is the case with the mesh router migration method. However, a new clustering scheme and new inter-cluster load balancing routing algorithms are needed, since neither the scheme nor the algorithms from the mesh router migration are able to choose candidate nodes that are not border nodes. In the light of this, CoCLuS and RAILoB are proposed in this thesis. Figure 5.9 shows the dependencies between mesh traffic migration, CoCLuS and RAILoB.

It is worth noting that all the components of CARMA form a significative part of ACRoMa architecture in Figure 5.9. In other words, mesh traffic migration and RAILoB are a process management plan to provide load balancing between the gateways, while CoCLuS is a topology management component to reduce the routing overhead and provide a clustering structure to enable the load balancing. CARMA is used as a term in this chapter to clarify the proposed clustering approach and its key role in the ACRoMa goals. Some of the concepts that are adopted in the mesh router migration method are also used in RAILoB, such as gateway, mesh

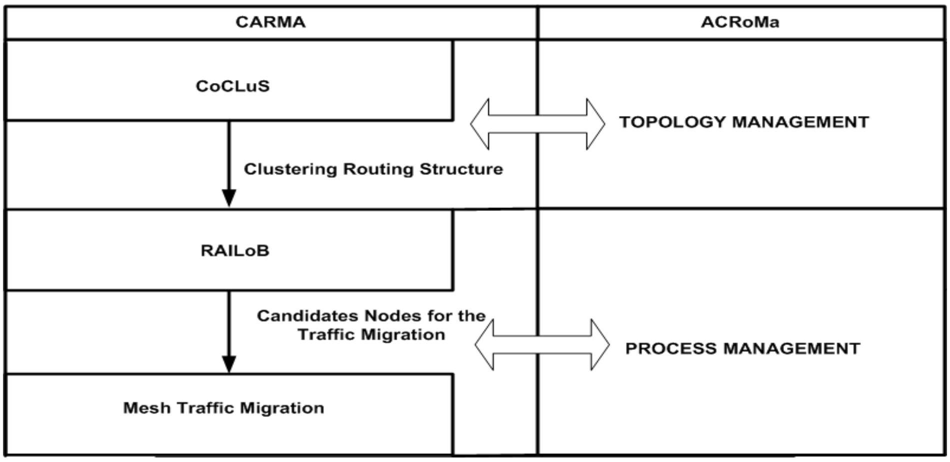

Figure 5.9: Mapping the CARMA approach in the ACRoMa architectural model

router, tree routing, TSN, adjacent TSN, CL, node weight and overload condition. In addition, the mesh traffic migration also restricts the migration of data so that it only occurs between the adjacent clusters and thus limits the number of hops to reach the destination, because, as it is well known, a small number of hops leads to a better traffic performance [He et al., 2008]. Every cluster has a single gateway which is located in the central position of the cluster. It is also assumed that the gateway placement is well defined.

In this context, some new factors should be borne in mind. First, there are some criteria that must be used to select nodes for the mesh traffic migration. As well as this, there is a need to determine what information is required for these candidate nodes to make an inter-cluster routing decision. This includes the kind of information that the nodes should be aware of, to support the selection of lightly-loaded clusters, the best way to calculate the whole path that leads to the lighter adjacent cluster and the kind of new functions the nodes should play in the clustering structure to support the mesh traffic migration. These factors will be discussed in the sub-sections that follow.

5.4.3 Collaborative CLustering Scheme (CoCLuS)

The main purpose of the CoCLuS [Borges et al., 2012c] is to provide a flexible clustering structure that enables an agile inter-cluster load balancing routing to occur through RAILoB. Moreover, as a result of clustering, routing decisions become more precise, due to the smaller scale of the area where cross-layer routing metrics are used. CoCLuS is described in the next paragraphs.

This scheme is an outline of the clustering structure and it defines how the clustering is formed (i.e. the clustering elements as well as their functionalities). Apart from the gateway and mesh router, CoCLuS contains two elements that did not exist in PLB, the relay node and the boundary node. The relay nodes are egress points from the heavily-loaded clusters, whilst the boundary nodes are ingress points to the lightly-loaded clusters. The relay and boundary nodes share the load information (e.g. CL of adjacent TSN) with all mesh routers belonging to the adjacent clusters, since they are within each other's transmission range. As a result, the relay and boundary nodes provide information to support the inter-cluster routing decision.

In fact, the boundary and relay nodes play a similar role in the mesh traffic migration process, but they are described in distinct ways, depending on the cluster in which the mesh routers are located. For instance *m(4,4)* is a relay node for all the mesh routers in the *G1* cluster and is a boundary node for all the mesh routers in the *G2* cluster. In other words, a boundary node does not belong to the cluster, whereas a relay node does. Figure 5.10 shows an example of the network model adopted in this proposed clustering scheme.

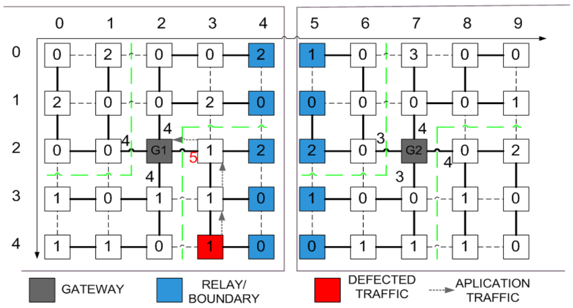

Figure 5.10: CoCLuS - Network Model

CoCLuS uses a new hybrid routing scheme that combines two different routing structures. First, the load balancing routing scheme provides the spanning tree structure (solid line) to communicate with the gateway which is built by the intra-cluster load balancing routing algorithm (IDLB [Choi and Han, 2008]). Later on, the nodes calculate the routes to every neighbour (excepting the gateway) inside the cluster by means of the Dijkstra algorithm and the MIND cross-layer routing metric (i.e. the link state routing scheme). This latter routing scheme (dotted line) is necessary to forward data from the member nodes to the relay nodes. It is worth noting that these routing schemes can be overlapped. Figure 5.11 shows the complete path in the mesh traffic migration method.

The complete path for the RAILoB algorithm consists of two main sub-paths which are as follows: intra-path (the path between the selected node and the relay node based on the link state routing scheme) and inter-path (the path between the

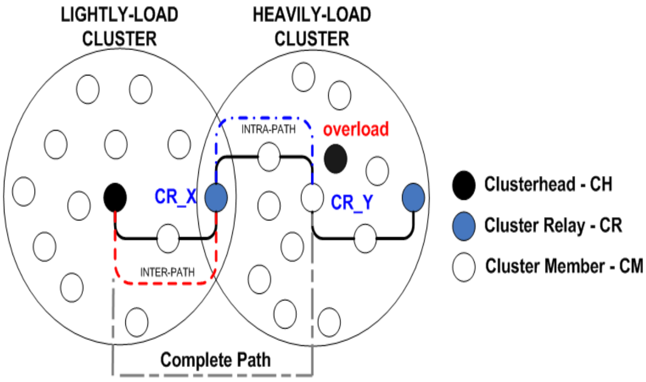

Figure 5.11: Intra-Path and Inter-Path

relay node and the lighter gateway based on the load balancing routing scheme).

5.4.4 Routing Algorithm for Inter-cluster Load Balancing (RAILoB)

RAILoB is the last component to be included in the ACRoMA architecture using the bottom-up approach of integration, and it also interacts with all remainder components of architecture. As result of this, RAILoB represents conceptually ACRoMA. RAILoB [Borges et al., 2012c] selects the candidate nodes for the mesh traffic migration method which allows the traffic of the selected mesh routers to migrate without needing mesh router migration.

When the mesh traffic migration method is employed, defected nodes are not required to be border nodes. This is significant because the main goal of RAILoB is to allow agility in reducing the traffic load in the nodes which are close to the TSN, while keeping control over the number of hops required to reach the destination. There are two criteria that are combined for selecting the candidate nodes. The first criteria shows a preference for the nodes which are farther away from the gateway in the sub-tree of the TSN overload (line 4 in Figure 5.12). This criterion has two advantages. First, it avoids congested links close to the gateway. Second, it means that the nodes that are closer to the adjacent cluster are more likely to be selected. By adopting this flexible method, RAILoB can add agility to the traffic migration and thus, reduce the time needed to carry out the inter-cluster traffic routing.

The second criterion seeks to select candidate nodes which are aware of at least one lighter TSN (i.e., the sum of the traffic load of the candidate node and CL of adjacent TSN does not exceed the load threshold - line 6 in Figure 5.14). It is important to stress that load of the adjacent TSN is shared by the relay/boundary nodes.

Hence, the mesh traffic migration is performed in two phases in which each phase requires different procedures or algorithms. Figure 5.12 shows the first procedure (performed in the gateways) that is used to select the candidate nodes.

```
//Key Notation
maxTSN - The number of TSN in the cluster
maxHopTSN_i - The maximum hop counts of the sub-tree rooted in i'th TSN
SubTreeNum_i^k - The number of nodes that are k hop away from the i'th TSN
node_i^(k,n) - The node at the n'th position among the k hop away nodes in the i'th TSN
weight_i^(k,n) - The weight at the n'th position among the k hop away nodes in the i'th TSN
//Selecting the candidates for traffic migration (this is performed in the gateway)

1:  for i=0 up to maxTSN do
2:        overload = CL[i] – LOAD_THRESHOLD
3:        Set defected_load and j to 0
4:     for k= maxHopTSN_i up to 0 step -1 do
5:          for n=0 up to SubTreeNum_i^k do
6:
7:               overload -= get_defected_load(node_i^(0,0));     //node_i^(0,0) is a TSN node
8:               if (overload <= 0)
9:                    break
10:
11:              if (((node_i^(k,n)->state() == WAITING) ||
12:                 (node_i^(k,n)->state() == REJECTED))
13:                 && ((node_i^(k,n)->time() >= TIMEOUT_THRESHOLD ))) {
14:                          Update the node state to NORMAL
15:                          candidates[j] = node_i^(k,n)
16:                          defected_load += weight_i^(k,n)
17:                          j++
18:
19:              } else if (node_i^(k,n)->state()==NORMAL) {
20:
21:                          candidates[j] = node_i^(k,n)
22:                          defected_load += weight_i^(k,n)
23:                          j++
24:                    }
25:          done
26:     done
27:
28:          for j=0 up to length(candidate) do
29:
30:               Send the DEFECTION_REQUEST message to candidate[j]
31:               Change the state of the candidate[j] to WAITING
32:
33:          done
34: done
```

Figure 5.12: Procedure for Candidate Selection

RAILoB uses different states so that the nodes for traffic migration process can be selected. Only nodes which are in the *NORMAL* state can be selected as candidates (line 19 in Figure 5.12). There are *TIMEOUT* and *BACKOFF_TIME* timers for the *WAITING* and *REJECTED* states, respectively, which means the nodes that are in these states have a chance to participate in the mesh traffic migration process (lines 11, 12 and 13 in Figure 5.12). These timers are based on the time used in the routing protocol to spread the control messages. The node where the traffic load is being migrated to (i.e. *ACCEPTED* state) can be released from the migration process in some situations, as follows: first, if the Cumulative Load of the adjacent

TSN where the traffic has been forwarded to, is greater than the maximum load threshold; and, second, if the Cumulative Load of the TSN which the node belongs to, is smaller than the maximum load threshold. As a result, the node will change its own state to *NORMAL* and notify the gateway. In this way, its load will be forwarded back to the original gateway. This also reduces the time for inter-cluster routing, since the released nodes must follow the slow process of the mesh router migration. In other words, the local decision of mesh traffic migration speeds up the migration of data as well the release of traffic between the appropriate clusters. Figure 5.13 shows the state machine with the different states and transitions.

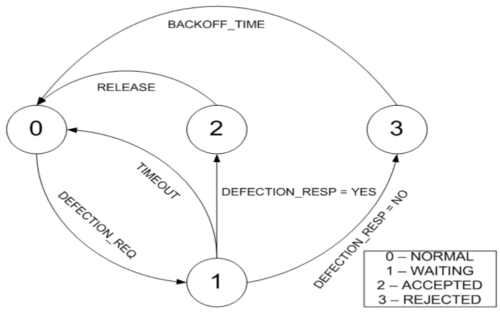

Figure 5.13: Node States

After the candidate nodes have been selected, the C-OLSR routing protocol, which supports clustering [Ros and Ruiz, 2007], sends messages to every candidate node requesting the migration of all the traffic. When a candidate node receives the defection request message, it checks if two conditions have been met before selecting the relay node. First, it establishes if there is a relay node which is able to communicate with at least one boundary node, in a sub-tree of an adjacent TSN which is not overloaded. In this process, it determines whether or not the sum of the weight of the candidate node and CL of the adjacent TSN will surpass that of the load threshold (line 6 in Figure 5.14). Second, it ensures that the number of hops between the candidate node and relay node is not greater than the average path length of all the nodes inside the cluster (i.e. *APL_THRESHOLD*). This second restriction of the relay node selection supports the main objective of mesh traffic migration, i.e. to avoid congestion in the gateway and TSN, while keeping control of the number of hops to reach the destination. It should be pointed out that the candidate node can also be a relay node.

All the nodes are familiar with all the relay nodes inside their own cluster. This is because each node which enters into communication with any boundary node can be defined as a relay node and regularly conveys the load information through control routing messages to boundary nodes. Furthermore, after receiving the load information from a boundary node, the relay node also spreads the Cumulative Load of the adjacent TSN to each neighbour inside the cluster, to support the inter-cluster routing decision. All the messages and information which are generated in this approach are piggy-backed in the default routing messages (i.e. HELLO and TOPOLOGY CONTROL) of the C-OLSR routing protocol.

```
//Key Notation
maxRelay - The number of relay nodes for this candidate
node_i - The relay node at the i'th position
LOAD - The weight of the candidate
PATH_LENGTH() - The number of hops from this node to node_i
APL_THRESHOLD – The average path length of all nodes needed to reach the gateway

//Checking if the candidate can migrate its traffic to some adjacent TSN (this is
performed in the candidate node)

Set CR_ID, GW_ID to -1 //CR_ID ➔ relay node and GW_ID ➔ adjacent gateway
Set response to 0

  for i=0 up to maxRelay do

        if ((node_i->boundary_CL() + LOAD) < LOAD_THRESHOLD) &&
           (node_i->PATH_LENGTH() <= APL_THRESHOLD)) {

           //boundary node is attached to some adjacent TSN

              response = 1;
              CR_ID = node_i->relay_addr();
              GW_ID = node_i->boundary_gw_addr();
              break

        }

  done

if (response==1) {

        CRInterCluster = CR_ID;         //Setting the relay for traffic migration
        GWInterCluster = GW_ID;         //Setting the gateway for traffic migration
}
```

Figure 5.14: Procedure for Relay Node Selection

Although the relay node and its respective boundary node are not in the same cluster, the relay node receives the CL of the adjacent TSN because the boundary nodes disseminate this information to their neighbours inside the cluster, as well to the relay nodes. It is important to point out that each relay node may communicate with more than one boundary node, but the relay node only takes into account the boundary nodes in which the TSN is not overloaded. When the relay node is connected to two or more effective boundary nodes (i.e. belong to lighter TSN), the boundary node is selected randomly.

The relay and boundary nodes play a key role in the mesh traffic migration process, since they share the load information of the adjacent TSN between the adjacent clusters. In this way, the candidate node is able to select the lighter adjacent cluster locally. Hence, the candidate nodes do not need to send a defection request to the adjacent TSN, since the RAILoB employs a proactive migration strategy to start the traffic migration, which further reduces the time required to start the traffic migration.

The mesh traffic migration method requires both a more complex clustering structure, since it uses new clustering elements (e.g. relay/boundary nodes) and establishes the inter-cluster path in two phases (1 - selecting the candidates nodes, 2 - selecting the relay node in a lighter adjacent TSN). Furthermore, mesh traffic migration incurs an additional routing overhead to make the inter-cluster routing decision locally. Figure 5.15 also shows an example of mesh traffic migration when CoCLuS and RAILoB are employed.

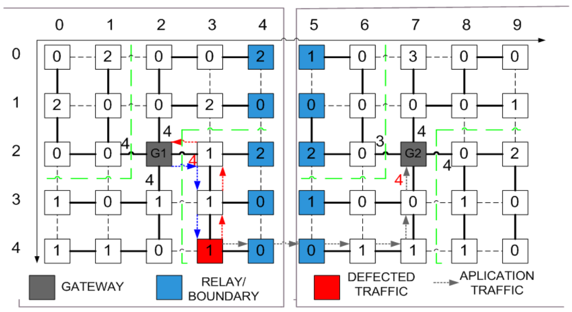

Figure 5.15: Mesh traffic migration - Example

There is an overload condition in *m(3,2)* in Figure 5.15. The *G1* gateway chooses *m(3,4)* for traffic migration (phase 1) and sends it a defection request message (blue dotted arrow). Next, in phase 2, *m(3,4)* checks in its routing database and finds *m(7,3)* (i.e. adjacent TSN that can accept the traffic in *m(3,4)* without overloading it) and then, *m(3,4)* sends back a defection response message (red dotted arrow) and starts to allow the traffic to migrate (dotted gray arrow) using *m(4,4)* and *m(5,4)* as relay and boundary nodes, respectively. Although it is not show the link weights in Figure 5.15, it is worth noting that the intra-path to *m(4,4)* is calculated based on the link-state routing scheme using MIND metric, whereas the inter-path calculation is based on the load balancing routing scheme of the IDLB procedure.

5.4.5 Discussion

The mesh traffic migration and mesh router migration can be defined as proactive and reactive migration strategies, respectively. However, it is important to stress that as soon as the traffic load has migrated to the lighter adjacent clusters, the congestion in the gateways decreases and the overall capacity of the network is improved. Regarding this matter, the mesh traffic migration is a more efficient method than the mesh router migration, since it uses a proactive migration strategy.

Table 5.2 summarizes the main aspects of each load balancing method.

Table 5.2: Inter-cluster Load Balancing Routing Methods

Methods	Migration Strategy	Routing Metrics	Clustering Elements
Mesh Router Migration	Reactive	No	Clusterhead
Mesh Traffic Migration	Proactive	MIND	Clusterhead, Relay and Boundary

Although the mesh traffic migration provides an agile and flexible method to make the inter-cluster routing through the local inter-cluster routing decision, it requires a more complex clustering structure. Moreover, mesh traffic migration generates an additional routing overhead when distributing information about the Cumulative Load of adjacent TSN and new cluster elements (relay and boundary nodes). Nevertheless, the mesh traffic migration may not have a higher total routing overhead than mesh router migration. The reason for this is that the mesh traffic migration does not need to send defection messages up to the adjacent TSN, since the inter-cluster routing decision is made locally by the candidate node. Thus, it is important to verify the impact of these methods on the routing overhead. It is also necessary to evaluate the approaches that have different traffic loads, nodes and gateways, since these factors are significant when evaluating load balancing methods. All these aspects will be included in the performance evaluation in the next section.

5.5 Performance Evaluation

The simulation study outlined in this section aims at throwing light on the ability of ACRoMa to confirm the supposition that it has the potential to achieve a greater degree of traffic performance when a more efficient inter-cluster load balancing routing is used. It is also concerned with drawing a comparison between RAILoB and the most effective inter-cluster load balancing routing (i.e. Partition Load Balancing (PLB)) [Choi and Han, 2010]), since RAILoB combines all the ACRoMa's components, including MIND metric. This section is structured as follows: the impact of intra-cluster and inter-cluster load balancing routing approaches, as well as the traffic load on the triple play service are discussed in Sub-sections 5.5.1 and 5.5.2, respectively. Sub-sections 5.5.3, 5.5.4 and 5.5.5 show the influence of the number of gateways, network (cluster) size and topology type on the inter-cluster routing approaches, respectively.

5.5.1 Effects of Intra-cluster Routing Schemes on the Triple Play Service

The main objective of this sub-section is to compare the triple play service performance when employing two different intra-cluster routing schemes. The first routing scheme, which is called Load Balancing (LB), is based on a spanning tree routing, and intra-cluster load balancing routing algorithm (i.e. Inner Domain Load Balancing (IDLB) [Choi and Han, 2010]) and the number of flows that the cross-layer routing metric used to balance the traffic load between the sub-trees rooted in the gateway. The number of sub-trees is equal to the number of neighbours that each gateway has. It should be pointed out that each sub-tree contains a spanning tree which enables the communication to occur with the gateway. The second routing scheme, which is called Link State (LS), is based on a link state routing scheme where all the nodes have all the routing information of the network and the Dijkstra routing algorithm. In this scheme, MIND was chosen as the cross-layer routing metric, since it employs precise measurements of interference and traffic load and has been validated with different applications and scenarios [Borges et al., 2009, Borges et al., 2010].

5.5.1.1 Simulation Configuration

In this sub-section, the performance evaluation will examine a mixed traffic comprising the VoIP, video and FTP applications which configure triple play services. In this way, we will be able to evaluate the impact of routing schemes on each application of the triple play services. Table 5.3 shows the scenario configuration.

The traffic combination of each application was based on [Quintero et al., 2004][Kim et al., 2008] which is very similar to the traffic model for each application used in sub-section 4.5.1. That is, the percentage of flows for VoIP, FTP and video are 60%, 30% and 10% of the total load, respectively. Thus, a set of four combinations of mixed traffic were formed, as shown in Table 5.4.

The scenario consists of 1 gateway and 24 static mesh routers with multi-channel multi-radio capability and is typical of outdoor city-wide deployments. There are two channels and two network interfaces. On each node, one particular channel is combined with one particular network interface, and no channel assignment algorithm

Table 5.3: Scenario Setup

Parameter	Value
Simulation Time	300s
Flow Lifetime	275s
Network Size	25
Grid Topology Size	1500m x 1500m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

Table 5.4: Traffic Combination for Evaluation of Intra-cluster Algorithms

Applications/Combinations	Video	FTP	VoIP
A	1	2	4
B	1	3	6
C	2	4	8
D	2	5	10

has been employed. Furthermore, grid topology is used to limit the maximum number of neighbours of a mesh router (i.e. four at maximum). Each data point in the graphical results is computed as the average of 10 different simulations and the graphs also show the confidence intervals of the performance parameters which have a confidence level of 95%.

The compared routing schemes were implemented on the OLSR routing protocol by means of the NS-2 simulator version 2.31 [NS-2, 2012]. The scenario uses a typical WMN backbone traffic pattern feature, where several flows originated from the source nodes (i.e. mesh routers) to a destination node (i.e., gateway), and the source nodes were chosen at random. The gateway is located in the central position [Bejerano et al., 2007].

5.5.1.2 Simulation Results

Figure 5.16 shows that the LS scheme results in slightly higher throughput than the Load Balancing scheme in medium loads (i.e. B and C load configurations). However, the traffic performance is better in the highest load (i.e. D configuration) when the Load Balancing scheme is employed. Both schemes are very similar in the lowest load configuration (i.e. 20 flows). Figure 5.17 shows that Load Balancing achieves lower delay in all the load configurations in VoIP, but the difference of performance decreases as the traffic load increases. The Link State and Load Balancing schemes result in a very similar pattern for FTP and video traffic when the delay is analysed. It should be noted that the VoIP traffic in Figure 5.16 has been multiplied by 10 to facilitate the visualization of the VoIP performance in the graph.

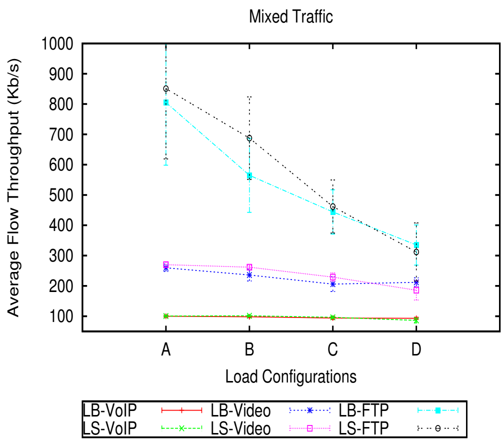

Figure 5.16: Average Flow Throughput

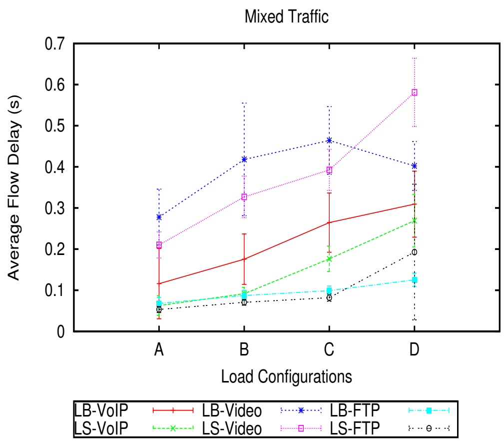

Figure 5.17: Average Flow Delay

These graphs provide evidence that the Link State routing scheme offers a good performance but has some limitations in its ability to provide a scalable solution for traffic performance. For example, although MIND results in the best performance when compared with other cross-layer routing metrics, it does not achieve precise load balancing routing, since it is not only a load-aware routing metric but is also an interference routing metric. On the other hand, the Load Balancing scheme enables it to distribute the traffic more easily and uniformly between the subtrees and hence, the interference is spread out. As a result of this, the Load Balancing routing scheme with spanning tree has a greater impact on the traffic performance (mainly the highest loads) than the Link State routing scheme with a interference-aware and load-aware routing metric. Hence, it can be concluded that the Load Balancing scheme provides a more scalable solution than the Link State scheme.

5.5.2 Effects of Inter-cluster Load Balancing Routing on the Triple Play Service

The simulation study outlined in this sub-section seeks to throw light on the impact of the load balancing methods on the traffic performance of FTP, VoIP and video streaming, as well as in combination (i.e. triple play services). In addition, this study makes a comparison between RAILoB which uses the mesh traffic migration method and the PLB which employs the mesh router migration method. Both approaches adopt the same intra-cluster load balancing routing procedure, i.e. Inner Domain Load Balancing (IDLB) [Choi and Han, 2010]. The scenario configuration and traffic model are outlined in sub-section 5.5.2.1 and the simulation results are discussed in sub-section 5.5.2.2.

5.5.2.1 Simulation Configuration

The scenario configuration includes 50 nodes, three of which are gateways [Ros and Ruiz, 2007]. As a result of this, the cluster size value is 17 nodes on each cluster. The approaches in [Bejerano, 2004, Aoun et al., 2006, He et al., 2008] adopted 20 nodes as the QoS constraint of the maximum cluster size. Hence, the clusters used in the simulation tests follow this rule. Each data point in the graphical results is computed as the average of 10 different simulations and the graphs also show the confidence intervals of the performance parameters which have a confidence level of 95%. The inter-cluster load-balancing approaches were implemented in an extended version of the OLSR routing protocol [Ros and Ruiz, 2007] by means of the NS-2, which supports the clustering. All of the nodes have the same physical configuration; there are two channels and two network interfaces. Table 5.5 displays a scenario configuration used in this sub-section.

Table 5.5: Scenario Setup

Parameter	Value
Simulation Time	300s
Flow Lifetime	275s
Network Size	50
Cluster Size	17
Number of Gateways	3
Grid Topology Size	2000m x 2000m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

5.5.2.2 Simulation Results

The main purpose of adopting these approaches for inter-cluster routing is to provide load balancing between the gateways and thus improve the traffic performance by managing the resource utilization in a better way. Thus, it is important to evaluate these approaches with different traffic loads as well as different applications, since each application causes a different traffic load in the network. Furthermore, it is necessary to check the impact of the distinct inter-cluster approaches on the traffic performance when a single application and a mix of applications is used in the network. To achieve this, the next paragraphs will describe two simulation studies.

Effects of Inter-cluster Routing on a Single Application

First, there is an analysis of the impact of inter-cluster routing approaches on the traffic performance for each application with four kinds of traffic load, as shown in Table 5.6. The traffic combination of each application was based on [Quintero et al., 2004][Kim et al., 2008] which is also very similar to the traffic model used in sub-section 4.5.1. The methods will be evaluated for the short-term (VoIP) and the long-term flows (video and FTP). It should be noted that the VoIP traffic in Figure 5.18 has been multiplied by 10 to facilitate the visualization of the VoIP performance in the graph.

Table 5.6: Traffic Combination for a Single Application

Applications/Combinations	Video	FTP	VoIP
A	8	10	20
B	10	13	27
C	12	16	34
D	15	20	40

Figures 5.18 and 5.19 show that the load balancing inter-cluster routing methods and traffic load have a significant impact on the traffic performance for each application, because there are significant differences in each method.

RAILoB uses the mesh traffic migration method which is more agile and flexible than PLB, while keeping the same cluster structure. By means of the method outlined above, the VoIP, video and FTP traffic are able to reach lighter adjacent clusters more quickly and the overloaded gateways are made lighter more quickly. As a result, the overall network capacity is improved because the traffic load is dis-

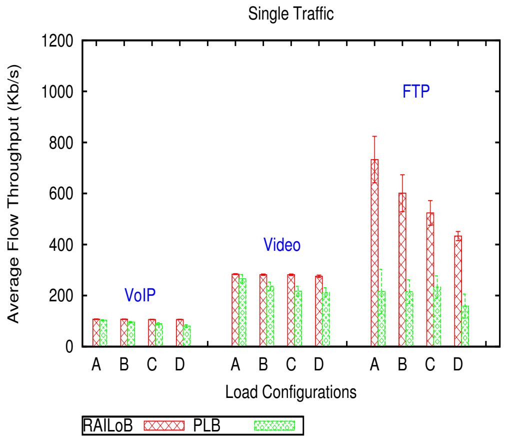

Figure 5.18: Average Flow Throughput

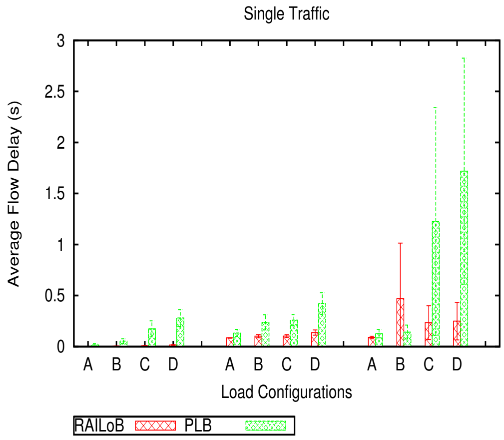

Figure 5.19: Average Flow Delay

tributed in a more uniform way. In addition, it should be pointed out that there is an increased difference in performance between the two approaches in the applications which require more bandwidth. For example, FTP achieves 433 Kb/s and 158 Kb/s in high loads when using RAILoB and PLB respectively, video reaches 275.63 Kb/s and 212.18 Kb/s in the highest load when using RAILoB and PLB, respectively, while VoIP reaches 10.63 Kb/s and 8.01 Kb/s in highest load when using RAILoB and PLB, respectively. TCP uses a transmission rate control policy so that it can be adapted in network congestion situations and thus, load balancing approaches have a greater impact on applications based on this transport protocol.

Effects of Inter-cluster Routing on Triple Play Services

In this case, the performance assessment will examine a mixed traffic comprising the VoIP, video and FTP applications which configure the triple play services. In this way, we will be able to evaluate the impact of load balancing methods on each application of these services. The traffic combination of each application was also based on [Quintero et al., 2004][Kim et al., 2008]. Thus, the percentage of flows for VoIP, FTP and video are 60%, 30% and 10% of the total load, respectively. A set of four combinations of mixed traffic were also formed, as shown in Table 5.7.

It should be stressed that none of these approaches or methods makes an inter-cluster routing decision that prioritizes a particular application. In other words, the total traffic of the selected mesh routers are migrated and hence, the improvement in performance achieved by these methods is restricted by what kind of application flows are attached to the selected mesh routers. As a result, these methods and

Table 5.7: Traffic Combination for Triple Play Services

Applications/Combinations	Video	FTP	VoIP
A	1	4	12
B	2	6	16
C	3	8	20
D	4	12	24

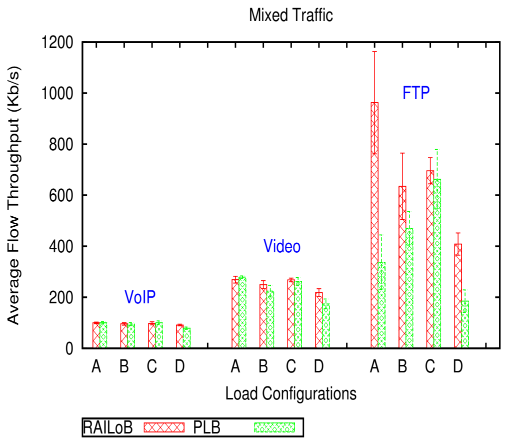

Figure 5.20: Average Flow Throughput

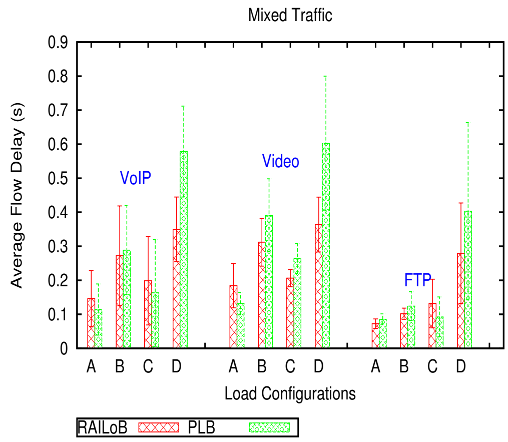

Figure 5.21: Average Flow Delay

traffic load had lesser impact on each application of the triple play services than on the single application configuration. Despite this, Figures 5.20 and 5.21 demonstrate that the load balancing inter-cluster routing methods and traffic load can still have an impact on the traffic performance for each application in a configuration of triple play services.

Figures 5.22 and 5.23 show the user-level assessment when RAILoB and PLB are used. Although RAILoB and PLB achieve a very similar average flow throughput in VoIP (Figure 2.18), PLB results in higher MOS than RAILoB in the lowest load, since PLB has lower delay than RAILoB. On the other hand, even though the RAILoB achieves lower throughput and higher delay than PLB in the lowest video load, RAILoB results in higher SSIM than PLB. This can be explained by the fact that there are more losses of I and P frames when PLB is used. For example, PLB has 32.55% and 27.15% of losses for frames I and P respectively, whereas RAILoB has 11.38 % and 6.6 % of losses for frames I and P, respectively. Nonetheless, RAILoB demonstrated that it was more scalable than PLB for triple play services in a user-level assessment. In addition, RAILoB achieves an acceptable average

Quality of Experience (QoE) for VoIP and video in the highest load (i.e. 3.65 in MOS and 0.93 in SSIM) [Kashyap et al., 2007a, Wang et al., 2004b]. Hence, the maximum acceptable number of VoIP, FTP and video flows is, in fact, defined by the highest traffic load.

It is also worth underlining that the performance gains of RAILoB for all the applications in the mixed traffic decrease when they are compared with its gains in performance (which were evaluated in the previous sub-section). For example, the performance gain of RAILoB for throughput in the highest load, is 32%, 29% and 172% for VoIP, video and FTP respectively in the case of a single application and, 15%, 25% and 120% for the same applications and load in the case of mixed traffic.

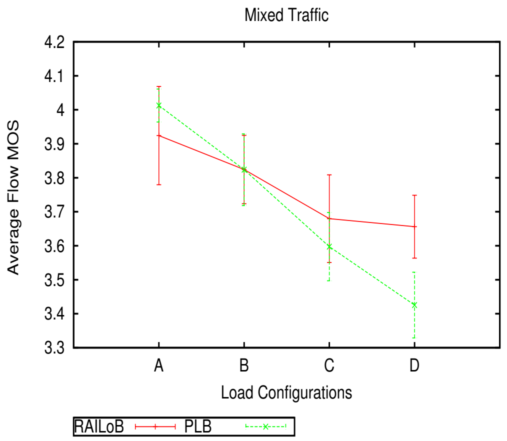

Figure 5.22: Average Flow MOS

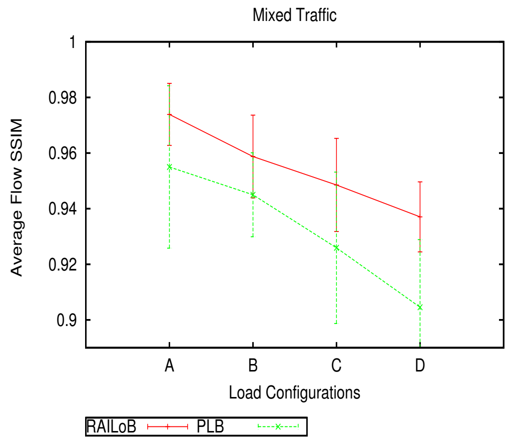

Figure 5.23: Average Flow SSIM

Although the traffic performance of each application is similar to both methods in medium loads (i.e. combinations B and C) particularly in VoIP and video, RAILoB results in a slightly better performance in most cases. However, PLB achieves a better performance in the lowest load (combination A), where RAILoB achieves a higher delay than PLB for VoIP and video. This can be explained by a number of factors. First, there is no application classification for inter-cluster routing decisions when the migration is made in both approaches. In addition, VoIP and video are more sensitive to the change and length of the path, which tends to increase when the mesh traffic migration method is employed because of its flexibility. Furthermore, there are fewer overload situations in low loads and thus a lower number of migrations. The flexibility of the mesh traffic migration causes superfluous migrations of VoIP and video traffic in low loads because it increases the traffic migration without any application classification.

Even though combination C contains more traffic load than combination B, the performance traffic of all the applications is better in combination C than in combination B for both methods (for example, RAILoB results in delay values of 272 ms and 198 ms for combinations B and C, respectively). This supports the supposition that the inter-cluster routing load balancing decision should take into account the kind of applications in the selected mesh routers that are required for the migration process. Figure 5.24 illustrates the number of migrations from two different perspectives and supplies evidence of the superfluous migrations for VoIP and video traffic, the number of Migrated Nodes (MN) and the traffic load (i.e. number of flows) attached to these migrated nodes, which is called Migrated Load (ML).

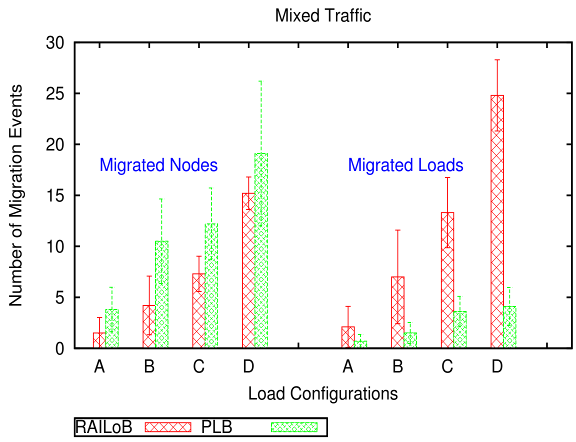

Figure 5.24: Number of Migration Events

Figure 5.24 shows the lowest number of migrations for all the load configurations that were found in each method when PLB was used. As expected, when both methods were employed, the number of migrated nodes and amount of migrated traffic increased when there was a greater traffic load. In addition, the PLB resulted in a smaller amount of migrated traffic than RAILoB for all the load configurations. The reason for this is that the mesh router migration method only detects border nodes which cannot have any or have the lowest traffic attached to them.

For example, despite the fact that PLB results in a higher number of migrated nodes for all the load configurations, the amount of migrated traffic load is higher when RAILoB is used, especially in the case of the highest load. This suggests that the inflexibility of the mesh router migration method prevents a wireless medium from being fully optimized for the highest load. In addition, it should also be stressed that PLB achieves a smaller number of migrated nodes and traffic in the lowest load. Nevertheless, VoIP and video show a better performance in low loads when they are evaluated through the PLB approach. This is an evidence that the mesh traffic migration method causes superfluous migrations which decrease the traffic performance for some applications in the lowest load.

5.5.3 Effects of Multiple Gateways on Inter-cluster Load Balancing Routing

In this sub-section, the performance evaluation will examine a mixed traffic flow comprising the VoIP, video and FTP applications which configure triple play services when varying the number of gateways. In this way, we will be able to evaluate the impact of the number of gateways on the inter-cluster routing methods. This sub-section is described as follows: the scenario configuration and traffic model are outlined in sub-section 5.5.3.1. The simulation results are discussed in sub-section 5.5.3.2.

5.5.3.1 Simulation Configuration

The aim of the inter-cluster routing approaches is to enable load balancing to occur between multiple gateways and this means that they must be evaluated with a variable number of gateways. To meet this requirement, four network scenarios are used to undertake this, each of which has a different number of gateways, flows, and nodes. The definition of the number of flows for each application is shown in Table 5.8.

Table 5.8: Traffic Combination for Multiple Gateways

Applications/Number of Gateways	**Video**	**FTP**	**VoIP**
2	2	8	16
3	4	12	24
4	6	16	32
5	8	20	40

Every cluster has 10 nodes which are attached to each gateway. It should be noted that the amount of traffic load for each gateway is very similar (i.e. 8, 4 and 1 flows of VoIP, FTP and video, respectively). The mixed traffic model uses the same configuration that was employed to assess the routing schemes in the previous sub-section.

5.5.3.2 Simulation Results

Figures 5.25 and 5.26 show that RAILoB also achieves the best performance for most of the evaluated parameters and network scenarios. RAILoB is more agile and flexible than PLB, while keeping the same cluster structure. As explained above, the triple play services are able to reach lighter adjacent clusters more quickly and the overloaded gateways are lightened at a faster rate. As a result, the overall network capacity is improved.

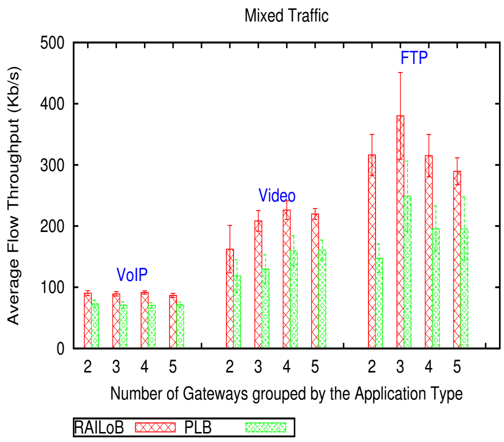

Figure 5.25: Average Flow Throughput

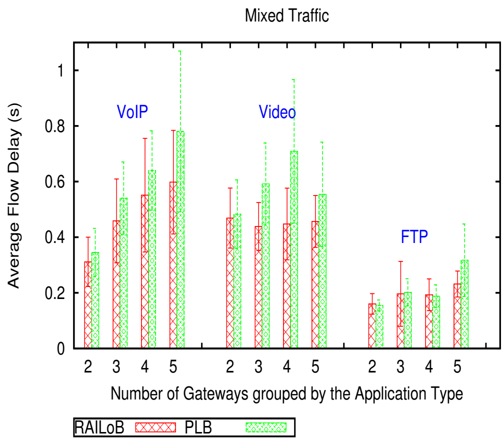

Figure 5.26: Average Flow Delay

The inter-cluster routing approaches have a greater effect on the multimedia applications (i.e. VoIP and video) than FTP, due to the fact that VoIP and video are more sensitive to the change and length of the path which tends to increase as there is a rise in the number of flows and nodes. The results obtained follow a different pattern for both approaches, i.e. the application performance tends to be different when the number of gateways increases for distinct applications of triple play services. On the one hand, the throughput value varies in all the scenario configurations for FTP and video, while the throughput remains the same as the number of gateways increases for both approaches in the VoIP application. It is important to stress that the video throughput increases still more when RAILoB is used. On the other hand, the delay parameter tends to rise as the number of gateways increases for all applications, except for video which has the same delay value when RAILoB is employed. The delay increase can be explained by the higher interference and congestion levels, since the number of nodes and flows increases with the number of gateways, which also increase the interference in the network. Nonetheless, these results provide evidence that RAILoB distributes the traffic load in a more uniform way than PLB and thus also reduces interference. Hence, RAILoB takes more advantage of multiple gateways than PLB.

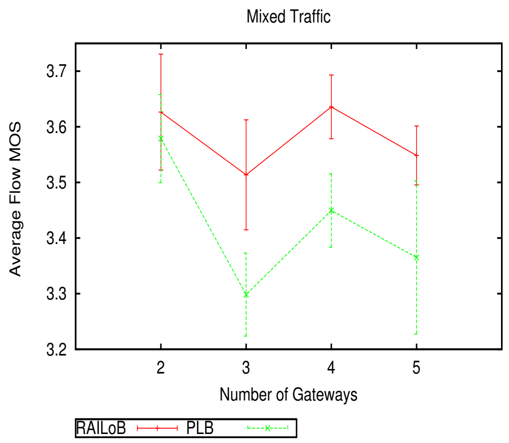

Figure 5.27: Average Flow MOS

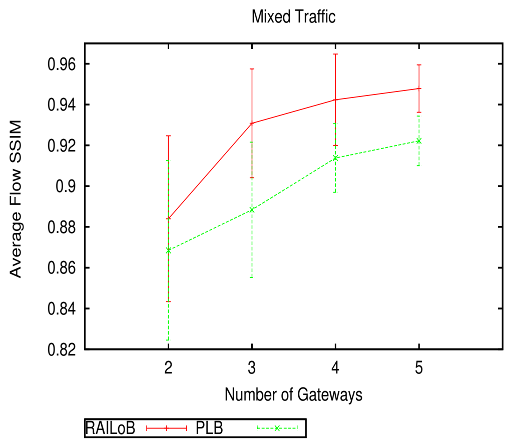

Figure 5.28: Average Flow SSIM

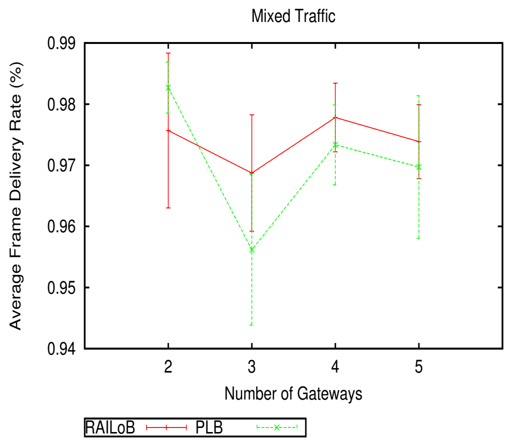

Figure 5.29: Average Frame Delivery Rate for VoIP application

Figures 5.27 and 5.28 provide evidence that user-level assessment is still higher when RAILoB is used. However, the QoE parameters of VoIP and video show a different tendency when the number of gateways varies. Whereas video improves its user perception quality, MOS decreases slightly when there is an increase in the number of gateways for both approaches. This can be explained by the fact that delay of VoIP application increases proportionally with a rise in the number of gateways in the network, while the video delay is very similar for all the scenario configurations. The throughput also improves for video, but the VoiP throughput is very similar for the distinct number of gateways. Furthermore, VoIP shows an unstable behaviour that can be explained by its conformity with the average flow service delivery framework (vide Figure 5.29).

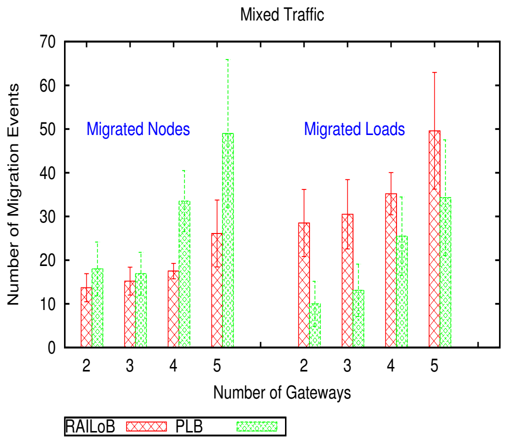

Figure 5.30: Number of Migration Events

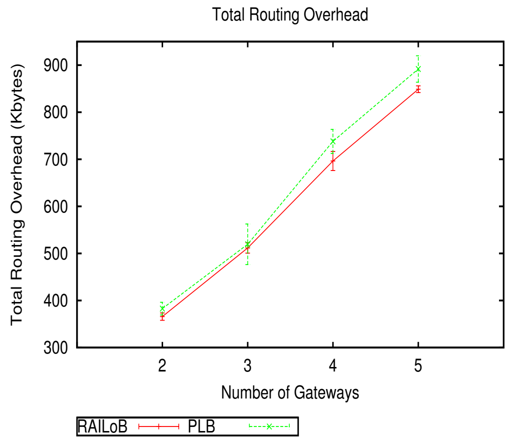

Figure 5.31: Total Routing Overhead

Figure 5.30 also shows that PLB results in a smaller amount of migrated traffic than RAILoB for all scenarios. As expected, in the case of both methods, the number of migrated nodes and amount of migrated traffic increases as the number of gateways rises, since there is also a proportional increase in the traffic load. Figure 5.31 illustrates the total routing overhead that is produced in the OLSR routing protocol when both the algorithms are tested. The total routing overhead takes into account the traditional routing messages of the protocol and the routing messages of the inter-cluster routing approaches (i.e. the defection messages and routes to the gateways). It is worth noting that the routing overhead increases with the number of gateways, when both PLB and RAILoB are used. This can also be expected because of the increase of nodes and flows and consequently, the traffic and mesh router migration also tend to increase. Surprisingly, RAILoB results in a slower routing overhead than PLB in all the network scenarios. In addition, the difference between RAILoB and PLB increases with the increase of the number of gateways. This results from the fact that RAILoB does not require defection messages for the adjacent TSN and only sends defection messages to the candidate node. Hence, RAILoB reduces the forwarded defection messages when the number of gateways increases. Furthermore, this reduction of messages offsets the extra information that RAILoB incurs in the network because of its proactive load balancing strategy (i.e. Cummulative Load of Top-Sub nodes). As a result of this, RAILoB is also a scalable solution for the routing overhead produced by the clustering routing protocols and thus, gains more from the clustering solution than PLB.

5.5.4 Effects of Network and Cluster Size on Inter-cluster Load Balancing Routing

The performance evaluation will also assess a triple play service configuration when varying the network and cluster sizes and the impact of these factors on the inter-cluster routing methods will be analysed. The scenario configuration and traffic model are outlined in sub-section 5.5.4.1. The simulation results are examined in sub-section 5.5.4.2.

5.5.4.1 Simulation Configuration

Table 5.9 shows the configuration of both scenarios used in this sub-section.

Table 5.9: Scenario Setup

Parameter	Value
Simulation Time	300s
Flow Lifetime	275s
Network Sizes	50 and 100
Cluster Sizes	17 and 20
Number of Gateways	3 and 5
Grid Topology Sizes	2000m x 2000m, 2500m x 2500m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

The traffic model is equivalent to that used in sub-section 5.5.2.1. In addition, the tests which were carried out in sub-section 5.5.2.1 (i.e. network size of nodes) are used here to compare the results of the network size of 50 and 100 nodes. This means that the results of load configurations A and D for the RAILoB and PLB approaches, are also used in this sub-section. In addition, new tests are also used in this sub-section. In addition, new tests are for scenarios with different network and cluster sizes (i.e. 100 and 20, respectively), in which two kinds of traffic load are used for both scenarios, for example combination A (20, 8 and 3 flows of VoIP, FTP and video, respectively) and combination D (40, 20 and 10 flows of VoIP, FTP and video, respectively). Both scenarios have the same traffic proportion by gateway, which makes it possible to analyze the impact of the network and cluster size on the inter-cluster routing methods.

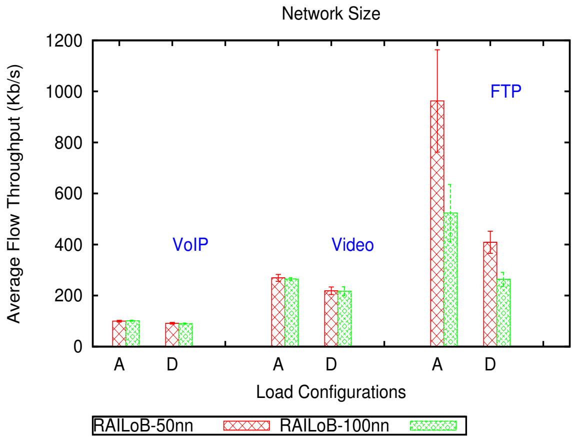

Figure 5.32: Average Flow Throughput of RAILoB

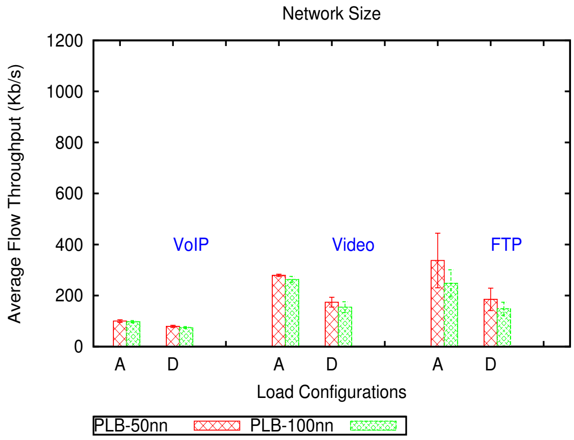

Figure 5.33: Average Flow Throughput of PLB

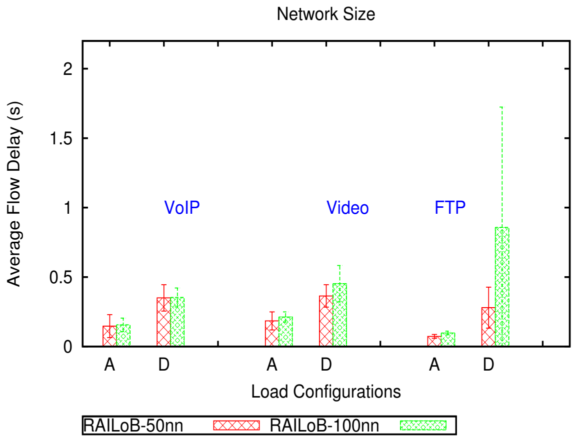

Figure 5.34: Average Flow Delay of RAILoB

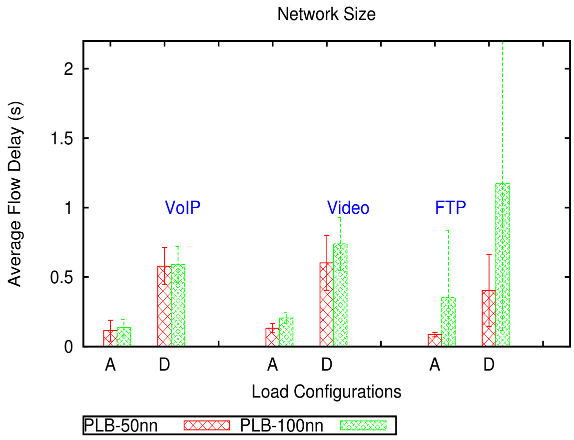

Figure 5.35: Average Flow Delay of PLB

5.5.4.2 Simulation Results

Figures 5.32 to 5.35 show that the network and cluster sizes have little impact on video and VoIP applications of triple play service, whereas these factors have a signifcant effect on FTP application. For example, FTP achieves 408,78 Kb/s and 263,84 Kb/s in high loads in a network size of 50 and 100 nodes respectively, when using RAILoB. This can be explained by the fact that an increase of the network size tends to raise the interference level and traffic load. As described in previous subsections, the transmission rate control policy of the TCP protocol is very sensitive to the packet loss rate which rises to the same extent that the interference and traffic load increase.

Despite the fact that the network and cluster sizes do not significantly influence the VoIP and video throughput, there are some cases where the traffic performance decreases when PLB is employed as a delay parameter. This is illustrated in Figures 5.34 and 5.35 in both load configurations, where it is clear that RAILoB results in a more scalable solution for WMN than PLB, since RAILoB achieves the highest throughput and the lowest delay for most of the cases when it takes into account different load application configurations, as well as network and cluster sizes.

5.5.5 Effects of Topology Scenario on Inter-cluster Load Balancing Routing

The effects of topology types on the inter-cluster routing methods will be investigated in this sub-section where a triple play service configuration is employed. This sub-section is structured as follows 5.5.5.1 shows the scenario configuration and traffic model. The simulation results are described in 5.5.5.2.

5.5.5.1 Simulation Configuration

In a similar way to the previous section, the traffic model is equivalent to that used in subsection 5.5.2.1. These tests are also used here to compare random and grid topologies. The amount of traffic is the same for both topology types. Table 5.10 shows the configuration of both scenarios used in this sub-section.

Table 5.10: Scenario Setup

Parameter	Value
Simulation Time	300s
Flow Lifetime	275s
Network Size	50
Cluster Size	17
Number of Gateways	3
Topology Size	2000m x 2000m
Transmission Range	250m
Interference Range	550m
Propagation Model	TwoRayGround
Network Interface Cards	2
MAC/PHY Specification	IEEE 802.11 b/g
Antenna	Omnidirectional

5.5.5.2 Simulation Results

Figures 5.36 to 5.39 show that the topology type does not have a significant effect on the triple play service, since neither of the inter-cluster routing approaches

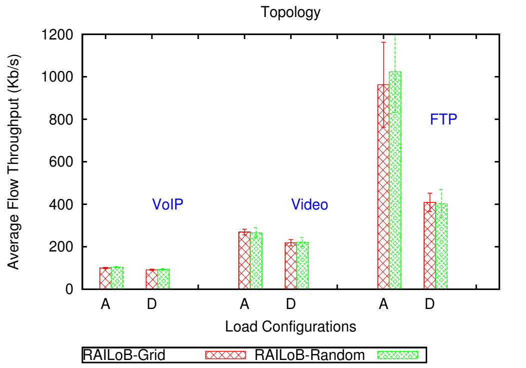

Figure 5.36: Average Flow Throughput of RAILoB

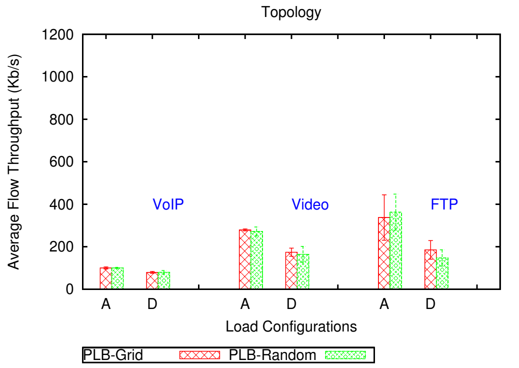

Figure 5.37: Average Flow Throughput of PLB

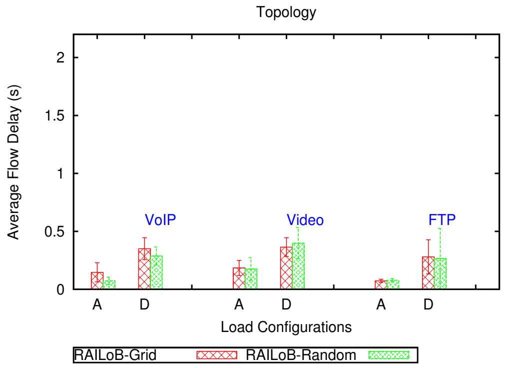

Figure 5.38: Average Flow Delay of RAILoB

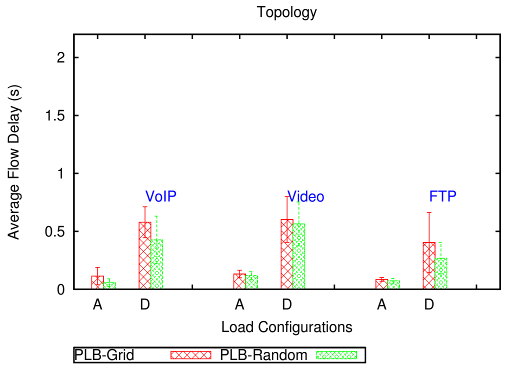

Figure 5.39: Average Flow Delay of PLB

results in a significant increase or decrease of traffic performance for any application. Nevertheless, there are some cases where the traffic performance slightly increases or decreases in a random topology that depends on the inter-cluster routing approach. For example, RAILoB achieves a higher improvement of throughput than PLB for FTP application in low loads, FTP achieves 962,70 Kb/s and 1023,75 Kb/s for grid and random topologies respectively when RAILoB is used, whereas FTP achieves 337,55 Kb/s and 362,93 Kb/s for grid and random topologies respectively, when PLB is used. The reason for this is that the random topology can have a varied number of border nodes for the mesh router migration method (i.e. PLB), including no single border node, since the node placement is not regular. This means that the traffic performance can be affected by slow and inflexible load balancing approaches in this specific case. Nonetheless, RAILoB results in the best traffic performance for most cases, as well as for both of the topology types.

5.6 Summary

WMN play an important role in providing ubiquitous Internet access through a wireless backbone where different kinds of wireless technologies can be integrated. In addition, these networks have the potential to increase their size considerably, since COTS products can be reused to increase the wireless backbone and extend the coverage area. For this reason, solutions that provide scalability for these networks are required. However, there are scalability limitations in the most used routing protocols in WMN since they are unable to exploit the overall capacity of WMN. In the light of this, clustering has been employed to improve the scalability of the existing routing protocols. Although it succeeds in doing this by controlling the routing overhead, it is not able to deal with the huge increase of the traffic load to manage the resource or take advantage of the overall network capacity. This study recommends CARMA as the clustering approach with load balancing that can distribute the traffic load between the neighbour clusters uniformly. CARMA consists of a clustering scheme, mesh traffic migration method, and intra-cluster as well as inter-cluster load balancing routing algorithms. The mesh traffic migration method provides more agility and flexibility for the load balancing routing process to mitigate the overload areas around the gateways. The CoCLuS clustering scheme provides a clustering structure for inter-cluster load balancing routing. The proposed RAILoB inter-cluster routing algorithm selects the candidates for traffic migration. Moreover, the proposed clustering approach contains a synergy with the MIND cross-layer routing metric that allows it to make an intra-cluster routing decision. An in-depth and detailed performance evaluation was undertaken to demonstrate that the proposed clustering approach achieves a better scalability for WMN than the most relevant related work. This evaluation covers a wide range of factors that can influence traffic performance such as topology, applications and network size. RAILoB achieves a better traffic performance than PLB (i.e. the most relevant related work) while it also keeps control of the routing overhead in most of the assessed scenarios, applications and performance parameters analysed in this chapter. Nonetheless, RAILoB has a drawback in dealing with low-load configurations for VoIP and video traffic because of its superfluous migrations. This supports the hypothesis that inter-cluster routing load balancing decisions should be aware of the kind of applications in the selected mesh routers that are used for the migration process. Furthermore, a comparison was also made between a common intra-cluster routing scheme of routing protocols (in which shortest path algorithm was combined with a cross-layer routing metric) and an intra-cluster load balancing routing scheme. These comparisons show that despite the evolution of cross-layer routing metrics, the load balancing intra-cluster scheme just results in a slightly better traffic performance in the highest load.

Chapter 6

Conclusion and Future Work

This thesis has examined the challenge of providing a scalability improvement for Wireless Mesh Networks (WMN) through the routing process. The Conclusion chapter examines what can be learnt from the research study conducted, together with issues that need to be addressed in future work. Section 6.1 describes the value of the findings and the most relevant conclusions that resulted from the performance evaluation carried out and Section 6.2 makes some recommendations for further studies in this area.

6.1 Conclusion

In this thesis, some ideas were put forward on how to enhance the scalability of WMN through the routing process. To achieve this, the proposed solution focused on three key areas of the routing process over WMN, which are as follows: inaccuracy in the route selection process, high routing overhead in large networks and the congestion areas around the gateways. These issues concerning routing approaches were addressed through a survey conducted in Chapter 3 and served to underpin the architecture proposed in this thesis. It was observed that none of the analysed approaches combines solutions for dealing with these questions on their own. As a means of filling this gap, an architectural model based on a top-down approach was employed, called Architecture of Cooperative Routing Management (ACRoMa), which is the main contribution of this thesis. ACRoMa seeks solutions for each of the questions previously discussed, such as a clustering solution to reduce the routing overhead, a load balancing routing algorithm to avoid overload situations at the gateways, and a cross-layer routing metric to improve the accuracy of the route selection process. It should be pointed out that these solutions are coordinated to increase network scalability and thus, improve the overall capacity of WMN.

The first specific contribution of this thesis is the cross-layer routing metric. A new taxonomy for the existing measurements used in the cross-layer routing metrics was proposed, which provides an in-depth knowledge of the main areas of this important subject. In the wide range of cross-layer routing metrics used for WMN, what has been lacking is a system that combines the most accurate measurements that are needed to depict interference and traffic load for a more complete solution. The analysis of the metric characteristics carried out in the taxonomy and survey (shown in Chapter 4) has laid the ground for devising a new cross-layer routing metric system that overcomes this limitation and the Metric for Interference and channel Diversity (MIND) cross-layer routing metric was proposed. MIND combines the most accurate measurements of traffic load, as well as physical and logical interference (i.e. inter-flow interference) while using passive monitoring solutions to obtain the measurements from the MAC and physical layers. Extensive simulation results showed that MIND outperforms several cross-layer routing metrics (e.g. CATT, MIC, iAWARE, INX, CWB and ILA) in WMN with a wide range of features, such as types of scenarios, traffic patterns and applications. It should also be stressed that other important factors were noticed, for instance the influence that the kind of environment (i.e., indoor or outdoor) has on the analysed metrics. In addition, although cross-layer routing metrics affect the performance of both the network and user levels, there are some cases where they have less impact on the user level, because the parameters for user perception are less influenced by the behaviour of the network.

The next specific contribution is the clustering approach, called Clustering Approach for Routing MAnagement. A new taxonomy for the existing clustering approaches used in the in WMN was also proposed. Although clustering is a well-established concept, the clustering approach proposed in this thesis is, as far as we are aware, an innovative system. CARMA consists of a traffic migration method, a clustering scheme and an inter-cluster load balancing routing algorithm. In the first place, the mesh traffic migration which is described as the new traffic migration method, allows a greater degree of agility and flexibility in the load balancing routing process than the mesh router migration method and thus reduces the overload areas around the gateways more efficiently. The proposed clustering scheme and the inter-cluster load balancing routing algorithm provide a routing scheme for the mesh traffic migration. The proposed clustering scheme in this thesis, called Collaborative CLustering Scheme (CoCLuS), seeks to provide a clustering structure that allows efficient inter-cluster load balancing routing. CoCLuS consists of a hybrid routing scheme and new clustering features. Both intra-cluster and inter-cluster load balancing routing were employed; for instance a comparison between a common intra-cluster routing scheme of routing protocols (in which a shortest path algorithm was combined with a cross-layer routing metric) and an intra-cluster load balancing routing scheme, demonstrated that despite the way cross-layer routing metrics have evolved, the load balancing intra-cluster scheme results in a better traffic performance, especially for high loads.

The inter-cluster load balancing routing algorithm, which is called Routing Algorithm for Inter-cluster Load Balancing (RAILoB), has been proposed to deal with the huge increase of the traffic load and act as a mechanism to balance the traffic load between the neighbouring clusters uniformly. Thus, RAILoB is designed to manage the resources so that it can make the most of the overall network capacity. CoCLuS was combined with an inter-cluster load balancing algorithm. In addition, there was an interaction between the RAILoB and MIND metrics to allow an intra-cluster routing decision to be made. Hence, it can be claimed that RAILoB represents the ACRoMa architecture conceptually by combining all the components. A detailed performance evaluation was undertaken which takes into account the main factors that have an impact on the traffic performance (e.g. topology, applications and network size). RAILoB outperforms the Partition Load Balancing (PLB) approach in most of the scenarios, applications and performance parameters that were analysed. Furthermore, it also results in lower routing overhead than PLB. The value of the proposed architecture was validated by simulation and the results obtained showed that it satisfied the objectives of the conception, by improving traffic performance and increasing the level of network utilization while reducing the overhead. Hence, it improves the WMN scalability through the routing process.

6.2 Future Work

The evaluation of the components included in this thesis has achieved some interesting results, but there are still aspects that need further work and other aspects that have risen. The first aspect that needs to be addressed with more detail is the evaluation by experimentation. Experimental work would be interesting to evaluate some components of the contributions of this dissertation. Particularly, we are mainly interested in validating some aspects of the MIND routing metric on a testbed. In addition, it would be interesting to add or design new measurements to MIND. In the long-term, we aim at developing a prototype which integrates all the components of the ACRoMa architecture.

The second aspect that requires more detailed study is the RAILoB algorithm. Although the results show that RAILoB achieves the highest traffic performance in high loads in all applications of triple play service (Sub-Section 5.5.2), RAILoB obtains the worst performance in low loads in VoIP and video, due to the problem of superfluous migrations. This supports the hypothesis that the inter-cluster routing decision should include the kind of applications found in the selected mesh routers for the traffic migration process. Furthermore, a classification of multimedia applications could achieve a significant improvement in these applications, such as class-based differentiation, which is provided by some mechanisms, like IEEE 802.11e.

The third aspect that could be included to supplement ACRoMa is the integration of a cognitive radio-based solution. As a result, ACRoMa will be able to improve the WMN coverage (e.g., up to 50 Km), as cognitive radio approaches enable to locate and use lower vacant frequency bands (i.e., frequency bands with higher ranges), while the cost for the network providers and users does not suffer any increase. Furthermore, this proposed architecture can be extended to the networking environment of Machine-to-Machine (M2M), since the WMN and M2M have similar characteristics such as, infrastructure-based or infrastructure-less wireless carrier network which provides communication between wireless end devices and back-end server in the wired networks.

Bibliography

[802.16d, 2002] 802.16d, I. (2002). IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems. *IEEE Std 802.16-2001*, pages 1–322.

[802.16m, 2011] 802.16m, I. (2011). http://standards.ieee.org/news/2011/80216m.html.

[802.16n, 2011] 802.16n, I. (2011). IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment: Higher Reliability Networks. web site: http://www.ieee802.org/16/pars/p802.16n.pdf. last visit: 12th november 2011.

[Adya et al., 2004] Adya, A., Bahl, P., Padhye, J., Wolman, A., and Zhou, L. (2004). A Multi-radio Unification Protocol for IEEE 802.11 Wireless Networks. In *BROADNETS '04: Proceedings of the First International Conference on Broadband Networks*, pages 344–354, Washington, DC, USA. IEEE Computer Society.

[Aïache et al., 2008] Aïache, H., Conan, V., Lebrun, L., Leguay, J., Rousseau, S., and Thoumin, D. (2008). XIAN Automated Management and Nano-Protocol to Design Cross-Layer Metrics for Ad Hoc Networking. In *NETWORKING'08: Proceedings of the 7th international IFIP-TC6 networking conference on AdHoc and sensor networks, wireless networks, next generation internet*, pages 1–13, Berlin, Heidelberg. Springer-Verlag.

[Aiache et al., 2008] Aiache, H., Conan, V., Lebrun, L., and Rousseau, S. (2008). A Load Dependent Metric for Balancing Internet Traffic in Wireless Mesh Networks. In *Mobile Ad Hoc and Sensor Systems, 2008. MASS 2008. 5th IEEE International Conference on*, pages 629–634.

[Akyildiz and Wang, 2008] Akyildiz, I. and Wang, X. (2008). Cross-Layer Design in Wireless Mesh Networks. *Vehicular Technology, IEEE Transactions on*, 57(2):1061–1076.

[Akyildiz et al., 2009] Akyildiz, I. F., Lee, W.-Y., and Chowdhury, K. R. (2009). CRAHNs: Cognitive Radio Ad Hoc Networks. *Ad Hoc Networks*, 7(5):810 – 836.

[Akyldiz et al., 2005] Akyldiz, I. F., Wang, X., and Wang, W. (2005). Wireless Mesh Networks: A Survey. *Computer Networks*, 47(4):445–487.

[Alsarhan and Agarwal, 2009] Alsarhan, A. and Agarwal, A. (2009). Cluster-Based Spectrum Management Using Cognitive Radios in Wireless Mesh Network. In *Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on*, pages 1 –6.

[Anastasopoulos et al., 2007] Anastasopoulos, M. P., Panagopoulos, A. D., and Cottis, P. G. (2007). A Distributed Routing Protocol for Providing QoS in Wireless Mesh Networks operating above 10 GHz. *Wireless Communications and Mobile Computing*, 9999(9999).

[Aoun et al., 2006] Aoun, B., Boutaba, R., Iraqi, Y., and Kenward, G. W. (2006). Gateway Placement Optimization in Wireless Mesh Networks With QoS Constraints. *IEEE JSAC*, 24(11):2127–2136.

[Ares et al., 2007] Ares, B. Z., Park, P. G., Fischione, C., Speranzon, A., and Johansson, K. H. (2007). On Power Control for Wireless Sensor Networks: System Model, Middleware Component and Experimental Evaluation. In *In European Control Conference*.

[Ashraf, 2010] Ashraf, U. (2010). *Quality of Service and Routing in Wireless Mesh Networks*. PhD thesis, Université de Toulouse. LAAS-CNRS. Available in: http://tel.archives-ouvertes.fr/docs/00/48/09/55/PDF/Thesis_Usman_ASHRAF_2010.pdf. Last visit: 13th november 2011.

[Ashraf et al., 2008] Ashraf, U., Abdellatif, S., and Juanole, G. (2008). An Interference and Link-Quality Aware Routing Metric for Wireless Mesh Networks. In *Proceedings of the 68th IEEE Vehicular Technology Conference, VTC Fall*, pages 1–5.

[Ashraf et al., 2011] Ashraf, U., Abdellatif, S., and Juanole, G. (2011). Route Maintenance in IEEE 802.11 Wireless Mesh Networks. *Computer Communications*, 34(13):1604 – 1621.

[Athanasiou et al., 2009] Athanasiou, G., Korakis, T., Ercetin, O., and Tassiulas, L. (2009). A Cross-layer Framework for Association Control in Wireless Mesh Networks. *Mobile Computing, IEEE Transactions on*, 8(1):65–80.

[Azcorra et al., 2009] Azcorra, A., Banniza, T., Chieng, D., Fitzpatrick, J., Von-Hugo, D., Natkaniec, M., Robitzsch, S., and Zdarsky, F. (2009). Supporting Carrier Grade Services over Wireless Mesh Networks: the Approach of the European FP-7 Strep CARMEN. *Communications Magazine, IEEE*, 47(4):14–16.

[Bacioccola et al., 2007] Bacioccola, A., Cicconetti, C., and Stea, G. (2007). User-Level Performance Evaluation of VoIP using NS-2. In *ValueTools '07: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, pages 1–10, ICST, Brussels, Belgium, Belgium.

[Bahr, 2007] Bahr, M. (2007). Update on the Hybrid Wireless Mesh Protocol of IEEE 802.11s. *IEEE MASS 2007*, pages 1–6.

[Barraca et al., 2008] Barraca, J. P., Fernandes, P., Sargento, S., and Rocha, R. M. (2008). An Architecture for Community Mesh Networking. In *PIMRC*, pages 1–6.

[Baumann et al., 2008] Baumann, R., Heimlicher, S., and Plattner, B. (2008). Routing in Large-Scale Wireless Mesh Networks Using Temperature Fields. *Network, IEEE*, 22(1):25–31.

[Baumann et al., 2007] Baumann, R., Heimlicher, S., Strasser, M., and Weibel, A. (2007). A Survey on Routing Metrics. Available in: www.baumann.info/public/tik262.pdf. Technical report, Computer Engineering and Networks Laboratory, ETH-Zentrum, Switzerland.

[Bejerano, 2004] Bejerano, Y. (2004). Efficient Integration of Multihop Wireless and Wired Networks with QoS Constraints. *IEEE/ACM Trans. Netw.*, 12(6):1064–1078.

[Bejerano et al., 2007] Bejerano, Y., Han, S.-J., and Kumar, A. (2007). Efficient Load-balancing Routing for Wireless Mesh Networks. *Computer Networks*, 51(10):2450–2466.

[Bellman, 1958] Bellman, R. (1958). On a Routing Problem. *Quarterly of Applied Mathematics*, 16(1):87–90.

[Beuster et al., 2008] Beuster, M., Beigl, M., Rohr, D., Riedel, T., Decker, C., and Berchtold, M. (2008). Matrix Routing – An Interference Range Insensitive Routing Protocol for Wireless Sensor Networks. In *Applications and the Internet, 2008. SAINT 2008. International Symposium on*, pages 137–140.

[Biaz and Qi, 2008] Biaz, S. and Qi, B. (2008). iETT: A Quality Routing Metric for Multi-Rate Multi-Hop Networks. In *WCNC*, pages 2729–2734.

[Boccolini et al., 2011] Boccolini, G., Luise, M., Mignone, V., and Vignaroli, L. (2011). Two-way Satellite Platform with ACM/QoS Integration for the Delivery of TV-centric Triple-play Services. *International Journal of Satellite Communications and Networking*, 29(3):293–313.

[Borges et al., 2011a] Borges, V. C., Curado, M., and Monteiro, E. (2011a). Cross-Layer Routing Metrics for Mesh Networks: Current Status and Research Directions. *Computer Communications*, 34(6):681 – 703.

[Borges et al., 2011b] Borges, V. C., Curado, M., and Monteiro, E. (2011b). The Impact of Interference-aware Routing Metrics on Video Streaming in Wireless Mesh Networks. *Ad Hoc Networks*, 9(4):652 – 661.

[Borges et al., 2012a] Borges, V. C., Curado, M., and Monteiro, E. (2012a). AC-RoMa Architecture for Cooperative Routing Management in Wireless Mesh Networks. *Submitted Paper to the Wireless Network Springer*.

[Borges et al., 2010] Borges, V. C. M., Curado, M., and Monteiro, E. (2010). A Cross-Layer Routing Scheme for Scalable Triple Play Service in Wireless Mesh Networks. In *Proceedings of the 19th IEEE ICCCN 2010, Zürich, Switzerland, August 2-5, 2010*, pages 1–6.

[Borges et al., 2012b] Borges, V. C. M., Dimitrov, E., Curado, M., and Monteiro, E. (2012b). Performance Assessment of Cluster Load Balancing Routing Methods for Triple Play Services in Wireless Mesh Networks. In *Proceedings of the 13th IEEE/IFIP NOMS 2012*, pages 974–980.

[Borges et al., 2012c] Borges, V. C. M., Dimitrov, E., Curado, M., and Monteiro, E. (2012c). RAILoB - A Routing Algorithm for Inter-cluster Load Balancing in Wireless Mesh Networks. In *Proceedings of the 9th IEEE CCNC 2012*, pages 904–909.

[Borges and Pereira, 2009] Borges, V. C. M. and Pereira, D. (2009). Cross-Layer Routing Metrics for OLSR in NS-2.31. available in: http://eden.dei.uc.pt/ vc-martins/resources.html.

[Borges et al., 2009] Borges, V. C. M., Vieira, D., Curado, M., and Monteiro, E. (2009). Routing Metric for Interference and Channel Diversity in Multi-Radio Wireless Mesh Network. *8th ADHOCNOW'09, Springer LNCS*, 5793:55–68.

[Brakmo and Peterson, 1995] Brakmo, L. and Peterson, L. (1995). TCP Vegas: end to end congestion avoidance on a global internet. *Selected Areas in Communications, IEEE Journal on*, 13(8):1465 –1480.

[Campista et al., 2008] Campista, M., Esposito, P., Moraes, I., Costa, L., Duarte, O., Passos, D., de Albuquerque, C., Saade, D., and Rubinstein, M. (2008). Routing Metrics and Protocols for Wireless Mesh Networks. *Network, IEEE*, 22(1):6–12.

[Capone and Martignon, 2007] Capone, A. and Martignon, F. (2007). A Multi-Commodity Flow Model for Optimal Routing in Wireless MESH Networks. *Journal of Networks (JNW)*, 2(3):1–5.

[Castro, 2011] Castro, M. C. (2011). *Enhancing P2P Systems over Wireless Mesh Networks*. PhD thesis, Karlstad University (Sweden). Available in: http://kau.diva-portal.org/smash/get/diva2:453220/FULLTEXT01. Last visit: 22th April 2012.

[Chakeres and Perkins, 2007] Chakeres, I. and Perkins, C. (2007). Dynamic manet on-demand (dymo) routing. Internet Draft, Internet Engineering Task Force. URL http://www.ietf.org/internet-drafts/draft-ietf-manet-dymo-11.txt. Work in progress.

[Chen and Heinzelman, 2007] Chen, L. and Heinzelman, W. (2007). A survey of routing protocols that support qos in mobile ad hoc networks. *Network, IEEE*, 21(6):30 –38.

[Chen et al., 2005] Chen, L.-J., Sun, T., Yang, G., Sanadidi, M., and Gerla, M. (2005). Ad Hoc Probe: Path Capacity Probing in Wireless Ad Hoc Networks. In *Proceedings of the WICON*, pages 156–163.

[Chen et al., 2009] Chen, L.-J., Sun, T., Yang, G., Sanadidi, M. Y., and Gerla, M. (2009). AdHoc Probe: End-to-End Capacity Probing in Wireless Ad Hoc Networks. *Wirel. Netw.*, 15(1):111–126.

[Chiang, 2005] Chiang, M. (2005). Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control. *Selected Areas in Communications, IEEE Journal on*, 23(1):104 – 116.

[Choi and Han, 2008] Choi, H.-G. and Han, S.-J. (2008). Load Balancing Routing for Wireless Mesh Networks: An Adaptive Partitioning Approach. In *CCNC 2008. 5th IEEE*, pages 1–5.

[Choi and Han, 2010] Choi, H.-G. and Han, S.-J. (2010). Domain Load Balancing Routing for Multi-gateway Wireless Mesh Networks. *Journal of Wireless Networks. Springer Kluwer Academic Publishers*, 16:2105–2122.

[Chroboczek, 2011] Chroboczek, J. (2011). The Babel Routing Protocol. .

[Chu and Wang, 2010] Chu, S. and Wang, X. (2010). MIMO-Aware Routing in Wireless Mesh Networks. In *INFOCOM, 2010 Proceedings IEEE*, pages 1 –9.

[Chuah and Katz, 2002] Chuah, C.-N. and Katz, R. H. (2002). Characterizing Packet Audio Streams from Internet Multimedia Applications. In *ICC 2002*, pages 1199–2203.

[Cordeiro et al., 2007] Cordeiro, W., Aguiar, E., Moreira, W., Abelem, A., and Stanton, M. (2007). Providing Quality of Service for Mesh Networks Using Link Delay Measurements. *16th IEEE ICCCN 2007*, pages 991–996.

[Corson and Macker, 1999] Corson, S. and Macker, J. (1999). Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations. RFC 2501.

[Couto et al., 2003] Couto, D. S. J. D., Aguayo, D., Bicket, J., and Morris, R. (2003). A high-throughput path metric for multi-hop wireless routing. In *9th MobiCom '03*, pages 134–146, New York, NY, USA. ACM.

[Crichigno et al., 2008] Crichigno, J., Wu, M.-Y., and Shu, W. (2008). Protocols and Architectures for Channel Assignment in Wireless Mesh Networks. *Ad Hoc Networks*, 6(7):1051–1077.

[Dai and Han, 2003] Dai, H. and Han, R. (2003). A Node-centric Load Balancing Algorithm for Wireless Sensor Networks. In *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, volume 1, pages 548 – 552 Vol.1.

[Davis and Raimondi, 2005] Davis, M. and Raimondi, T. (2005). A Novel Framework for Radio Resource Management in IEEE 802.11 Wireless LANs. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2005. WIOPT 2005. Third International Symposium on*, pages 139–147.

[Dely et al., 2010] Dely, P., Kassler, A., Bayer, N., and Sivchenko, D. (2010). An Experimental Comparison of Burst Packet Transmission Schemes in IEEE 802.11-Based Wireless Mesh Networks. In *GLOBECOM*, pages 1–5.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A Note on Two Problems in Connexion With Graphs. *Numerische Mathematik*, 1:269–271.

[Doyle and Carroll, 2005] Doyle, J. and Carroll, J. (2005). *Routing TCP/IP.* Cisco Press, second edition edition.

[Draves et al., 2004a] Draves, R., Padhye, J., and Zill, B. (2004a). Comparison of Routing Metrics for Static Multi-hop Wireless Networks. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 133–144, New York, NY, USA. ACM.

[Draves et al., 2004b] Draves, R., Padhye, J., and Zill, B. (2004b). Routing in Multi-radio, Multi-hop Wireless Mesh Networks. In *MobiCom '04*, pages 114–128, New York, NY, USA. ACM.

[Ekling et al., 2007] Ekling, J., Gidlund, M., and Flodin, R. (2007). Performance of Triple Play Services in Wireless Meshed Networks. IEEE ISCE. pages 1 –5.

[Ergin et al., 2008] Ergin, M. A., Gruteser, M., Luo, L., Raychaudhuri, D., and Liu, H. (2008). Available Bandwidth Estimation and Admission Control for QoS Routing in Wireless Mesh Networks. *Comput. Commun.*, 31(7):1301–1317.

[Etemad, 2008] Etemad, K. (2008). Overview of Mobile WiMAX Technology and Evolution. *Communications Magazine, IEEE*, 46(10):31 –40.

[EvalVid, 2012] EvalVid (2012). EvalVid - A Video Quality Evaluation Tool-set.

[Fiore, 2009] Fiore, M. (2009). Wireless Patch for Ns-2.29.

[G.114, 2003] G.114, I.-T. (2003). ITU-T G.114 : One-way Transmission Time.

[Garlington et al., 2005] Garlington, T., Babbitt, J., and Long, G. (2005). Analysis of Free Space Optics as a Transmission Technology. White paper, WP No. AMSEL-IE-TS-05001. U.S. Army Information Systems Engineering Command. p. 3. URL WP No. AMSEL-IE-TS-05001. U.S. Army Information Systems Engineering Command. p. 3 URL http://www.hqisec.army.mil/isec/publications/Analysis_of_Free_Space_Optics_as_a_Transmission_Technology_Mar05.pdf. Retrieved June 28, 2011.

[Genetzakis and Siris, 2008] Genetzakis, M. and Siris, V. (2008). A Contention-Aware Routing Metric for Multi-Rate Multi-Radio Mesh Networks. In *5th IEEE SECON '08*, pages 242–250.

[Gerla and Tsai, 1995] Gerla, M. and Tsai, J. T.-C. (1995). Multicluster, Mobile, Multimedia Radio Network. *Wirel. Netw.*, 1(3):255–265.

[Gomes et al., 2011] Gomes, R. L., Junior, W. M., Cerqueira, E., and Abelém, A. J. (2011). Using fuzzy link cost and dynamic choice of link quality metrics to achieve qos and qoe in wireless mesh networks. *Elsevier Journal of Network and Computer Applications*, 34(2):506 – 516.

[Guerin et al., 2007] Guerin, J., Portmann, M., and Pirzada, A. (2007). Routing metrics for multi-radio wireless mesh networks. In *Telecommunication Networks and Applications Conference, 2007. ATNAC 2007. Australasian*, pages 343 –348.

[Gupta and Kumar, 2000] Gupta, P. and Kumar, P. (2000). The Capacity of Wireless Networks. *Information Theory, IEEE Transactions on*, 46(2):388–404.

[Haas et al., 2002] Haas, Z. J., Pearlman, M. R., and Samar, P. (2002). The Zone Routing Protocol (ZRP) for Ad Hoc Networks. Available in: http://people.ece.cornell.edu/haas/wnl/Publications/draft-ietf-manet-zone-zrp-04.txt. Technical report.

[He et al., 2009] He, B., Sun, D., and Agrawal, D. (2009). Diffusion Based Distributed Internet Gateway Load Balancing in a Wireless Mesh Network. In *GLOBECOM 2009. IEEE*, pages 1 –6.

[He et al., 2008] He, B., Xie, B., and Agrawal, D. P. (2008). Optimizing Deployment of Internet Gateway in Wireless Mesh Networks. *Comput. Commun.*, 31(7):1259–1275.

[Held, 2005] Held, G. (2005). *Wireless Mesh Networks.* Auerbach Publications, Boston, MA, USA.

[Hiertz et al., 2007] Hiertz, G., Max, S., Zhao, R., Denteneer, D., and Berlemann, L. (2007). Principles of IEEE 802.11s. In *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on*, pages 1002 –1007.

[Hsiao et al., 2001] Hsiao, P.-H., Hwang, A., Kung, H., and Vlah, D. (2001). Load-balancing Routing for Wireless Access Networks. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 986 –995 vol.2.

[IBM, 2006] IBM (2006). Triple Play Service Management. Last visit: 29th january 2012.

[IEEE, 1999] IEEE (1999). Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Standard 802.11.

[IEEE, 2005] IEEE (2005). IEEE 802.16-2005 Standard for Local and Metropolitan Area Networks Part 16: Air Interface to fixed and Mobile Broadband Wireless Access Systems. IEEE.

[Ishmael et al., 2008] Ishmael, J., Bury, S., Pezaros, D., and Race, N. (2008). Deploying Rural Community Wireless Mesh Networks. *IEEE Internet Computing*, 12(4):22–29.

[Islam et al., 2010] Islam, M., Alam, M., Hamid, M., Hong, C., and Lee, S. (2010). EFT: A High Throughput Routing Metric for IEEE 802.11s Wireless Mesh Networks. *Annals of Telecommunications*, 65:247–262.

[ITU, 1994] ITU (1994). X.200: Information technology - open systems interconnection - basic reference model: The basic model,.

[Jacquet et al., 2001] Jacquet, P., Mühlethaler, P., Clausen, T., Laouiti, A., Qayyum, A., and Viennot, L. (2001). Optimized Link State Routing Protocol for Ad Hoc Networks. In *IEEE INMIC 2001.*

[Jain et al., 2005] Jain, K., Padhye, J., Padmanabhan, V. N., and Qiu, L. (2005). Impact of interference on multi-hop wireless network performance. *Wirel. Netw.*, 11(4):471–487.

[Jiang et al., 2007] Jiang, W., Liu, S., Zhu, Y., and Zhang, Z. (2007). Optimizing Routing Metrics for Large-Scale Multi-Radio Mesh Networks. In *WiCom 2007*, pages 1550–1553.

[Johnson et al., 2008] Johnson, D., Ntlatlapa, N., and Aichele, C. (2008). Simple pragmatic approach to mesh routing using batman. In *2nd IFIP International Symposium on Wireless Communications and Information Technology in Developing Countries*.

[Johnson and Maltz, 1996] Johnson, D. B. and Maltz, D. A. (1996). Dynamic source routing in ad hoc wireless networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers.

[Karrer and Pescape, 2007] Karrer, R. P. and Pescape, A. (2007). 2nd Generation Wireless Mesh Networks: Technical, Economical and Social Challenges. *Future Generation Communication and Networking*, 1:262–267.

[Kashyap et al., 2007a] Kashyap, A., Ganguly, S., Das, S. R., and Banerjee, S. (2007a). Voip on wireless meshes: Models, algorithms and evaluation. In *INFOCOM*, pages 2036–2044.

[Kashyap et al., 2007b] Kashyap, A., Lee, K., Kalantari, M., Khuller, S., and Shayman, M. (2007b). Integrated Topology Control and Routing in Wireless Optical Mesh Networks. *Comput. Netw.*, 51:4237–4251.

[Khabazian and Aissa, 2010] Khabazian, M. and Aissa, S. (2010). A Load-distributive QoS Routing Protocol for Multi-service Wireless Mesh Networks. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on*, pages 233 –238.

[Kim et al., 2010] Kim, C., Ko, Y.-B., and Vaidya, N. H. (2010). Link-state routing without broadcast storming for multichannel mesh networks. *Computer Networks*, 54(2):330 – 340.

[Kim et al., 2008] Kim, H., Claffy, K., Fomenkov, M., Barman, D., Faloutsos, M., and Lee, K. (2008). Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices. In *CONEXT '08: Proceedings of the 2008 ACM CoNEXT Conference*, pages 1–12, New York, NY, USA. ACM.

[Koksal et al., 2006] Koksal, C. E., Jamieson, K., Telatar, E., and Thiran, P. (2006). Impacts of Channel Variability on Link-level Throughput in Wireless Networks. *SIGMETRICS Perform. Eval. Rev.*, 34(1):51–62.

[Kone et al., 2007] Kone, V., Das, S., Zhao, B. Y., and Zheng, H. (2007). QUORUM - QUality Of service RoUting in wireless Mesh networks. *ACM Mobile Networking and Applications (MONET)*, 12(5–6):358–369.

[Kong et al., 2009] Kong, P.-Y., Pathmasuntharam, J., Wang, H., Ge, Y., Ang, C.-W., Su, W., Zhou, M.-T., and Harada, H. (2009). A Routing Protocol for WiMAX Based Maritime Wireless Mesh Networks. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1 –5.

[Kowalik et al., 2007] Kowalik, K., Keegan, B., and Davis, M. (2007). RARE - Resource Aware Routing for mEsh. In *IEEE ICC '07*, pages 4931–4936.

[Kuran et al., 2007] Kuran, M. S., Gür, G., Tugcu, T., and Alagöz, F. (2007). Cross-layer Routing-scheduling in IEEE 802.16 Mesh Networks. In *Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications*, MOBILWARE '08, pages 47:1–47:6, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[Kyasanur and Vaidya, 2006] Kyasanur, P. and Vaidya, N. H. (2006). Routing and Link-layer Protocols for Multi-channel Multi-interface Ad Hoc Wireless Networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 10(1):31–43.

[Lacage et al., 2004] Lacage, M., Manshaei, M. H., and Turletti, T. (2004). IEEE 802.11 Rate Adaptation: A Practical Approach. In *MSWiM '04: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 126–134, New York, NY, USA. ACM.

[Langar et al., 2009] Langar, R., Bouabdallah, N., and Boutaba, R. (2009). Mobility-aware clustering algorithms with interference constraints in wireless mesh networks. *Comput. Netw.*, 53(1):25–44.

[Lie and Klaue, 2008] Lie, A. and Klaue, J. (2008). Evalvid-RA: Trace Driven Simulation of Rate Adaptive MPEG-4 VBR video. *Multimedia Syst.*, 14(1):33–50.

[Liu et al., 2009] Liu, S., Feng, S., Ye, W., and Zhuang, H. (2009). Slot allocation algorithms in centralized scheduling scheme for ieee 802.16 based wireless mesh networks. *Computer Communications*, 32(5):943 – 953.

[Liu and Liao, 2008] Liu, T. and Liao, W. (2008). On routing in multichannel wireless mesh networks: Challenges and solutions. *Network, IEEE*, 22(1):13–18.

[Ma and Denko, 2007] Ma, L. and Denko, M. (2007). A Routing Metric for Load-Balancing in Wireless Mesh Networks. In *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, volume 2, pages 409–414.

[Manikantan Shila and Anjali, 2008] Manikantan Shila, D. and Anjali, T. (2008). Load Aware Traffic Engineering for Mesh Networks. *Comput. Commun.*, 31(7):1460–1469.

[Marina et al., 2010] Marina, M. K., Das, S. R., and Subramanian, A. P. (2010). A topology control approach for utilizing multiple channels in multi-radio wireless mesh networks. *Computer Networks*, 54(2):241 – 256.

[Marinho and Monteiro, 2011] Marinho, J. and Monteiro, E. (2011). Cognitive Radio: Survey on Communication Protocols, Spectrum Decision Issues, and Future Research Directions. *Wireless Networks*, pages 1–18.

[Moradi et al., 2010] Moradi, H., Falahpour, M., Reafi, H., LoPresti, P., and Atiquzzaman, M. (2010). Availability Modeling of FSO/RF Mesh Networks through Turbulence-Induced Fading Channels. In *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*, pages 1–5.

[Motorola, 2011] Motorola (2011). Long Term Evolution (LTE). White paper, Motorola. Available in http://www.motorola.com/web/Business/Solutions/Industry%20Solutions/Service %20Providers/Wireless%20Operators/LTE/_Document/Static%20Files/ 6833_MotDoc_New.pdf. Work in progress.

[Nachtigall et al., 2008] Nachtigall, J., Zubow, A., and Redlich, J. P. (2008). The Impact of Adjacent Channel Interference in Multi-Radio Systems using IEEE 802.11. In *Proc. International Wireless Communications and Mobile Computing Conference IWCMC '08*, pages 874–881.

[Nee and Prasad, 2000] Nee, R. V. and Prasad, R. (2000). *OFDM for Wireless Multimedia Communications*. Artech House Publisher.

[Nguyen et al., 2008] Nguyen, L. T., Beuran, R., and Shinoda, Y. (2008). A Load-aware Routing Metric for Wireless Mesh Networks. In *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on*, pages 429–435.

[NS-2, 2012] NS-2 (2012). Ns-2. available in: http://www.isi.edu/nsnam/ns. last visit: 29th january 2012.

[Oestges and Clerckx, 2007] Oestges, C. and Clerckx, B. (2007). *MIMO Wireless Communications: From Real-World Propagation to Space-Time Code Design*. Academic Press, Inc., Orlando, FL, USA.

[Olszewski, 2007] Olszewski, K. J. (2007). SINR Measurement Method for OFDM Communications Systems.

[Ouni et al., 2011] Ouni, A., Rivano, H., and Valois, F. (2011). On the Capacity and Energy Trade-Off in LTE-Like Network. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pages 1 –5.

[p.800.1, 2006] p.800.1, I.-T. (2006). ITU-T p.800.1 : Mean Opinion Score (MOS) terminology .

[Pal and Nasipuri, 2011] Pal, A. and Nasipuri, A. (2011). A Quality Based Routing Protocol for Wireless Mesh Networks. *Pervasive and Mobile Computing Elsevier Journal*, 7(5):611 – 626.

[Park and Han, 2008] Park, J.-S. and Han, S.-J. (2008). Load balancing for video streaming services in hierarchical wireless networks. *Comput. Netw.*, 52(1):259–274.

[Penrose, 1999] Penrose, M. D. (1999). On k-connectivity for a geometric random graph. *Random Struct. Algorithms*, 15(2):145–164.

[Perkins and Bhagwat, 1994] Perkins, C. E. and Bhagwat, P. (1994). Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. *SIGCOMM Comput. Commun. Rev.*, 24(4):234–244.

[Perkins, 1999] Perkins, C.E.; Royer, E. (25-26 Feb 1999). Ad-hoc On-demand Distance Vector Routing. *Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA '99. Second IEEE Workshop on*, pages 90–100.

[Postel, 1980] Postel, J. (1980). User datagram protocol. Request For Comments 768, IETF.

[Postel, 1981] Postel, J. (1981). Transmission control protocol: DARPA internet program protocol specification. Request For Comments 793, IETF.

[Prehofer and Bettstetter, 2005] Prehofer, C. and Bettstetter, C. (2005). Self-Organization in Communication Networks: Principles and Design Paradigms. In *Advances in Self-Organizing Networks*. IEEE Communications Magazine.

[Quintero et al., 2004] Quintero, A., Elalamy, Y., and Pierre, S. (2004). Performance Evaluation of a Broadband Wireless Access System Subjected to Heavy Load. *Computer Communications*, 27(9):781 – 791.

[Ramach et al., 2005] Ramach, K. N., Buddhikot, R. M. M., Ch, G., Miller, S., Belding-royer, E. M., and Almeroth, K. C. (2005). On the Design and Implementation of Infrastructure Mesh Networks. Technical report, in IEEE Workshop on Wireless Mesh Netorks (WiMesh.

[Ramachandran et al., 2007] Ramachandran, K. N., Sheriff, I., Belding, E. M., and Almeroth, K. C. (2007). Routing Stability in Static Wireless Mesh Networks. *8th PAM 2007, Louvain-la-neuve, Belgium, April 5-6, 2007, Springer LNCS*, 4427.

[Rappaport, 2001] Rappaport, T. (2001). *Wireless Communications: Principles and Practice*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

[Reis et al., 2006] Reis, C., Mahajan, R., Rodrig, M., Wetherall, D., and Zahorjan, J. (2006). Measurement-based Models of Delivery and Interference in Static Wireless Networks. *SIGCOMM Comput. Commun. Rev.*, 36(4):51–62.

[Rodrigues et al., 2011] Rodrigues, D., Cerqueira, E., and Monteiro, E. (2011). Quality of experience adaptation controllers for voice and video in wireless networks. In *Proceedings of the 9th IFIP TC 6 international conference on Wired/wireless internet communications*, WWIC'11, pages 350–361, Berlin, Heidelberg. Springer-Verlag.

[Ros and Ruiz, 2007] Ros, F. J. and Ruiz, P. M. (2007). Cluster-based OLSR Extensions to Reduce Control Overhead in Mobile Ad hoc Networks. In *Proceedings of IWCMC*, pages 202–207, New York, NY, USA. ACM.

[Rouse and Hemami, 2008] Rouse, D. M. and Hemami, S. S. (2008). Understanding and simplifying the structural similarity metric. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1188–1191, San Diego, CA, USA.

[Schulzrinne et al., 2003] Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V. (2003). RTP: A Transport Protocol for Real-Time Applications.

[Schwarz et al., 2007] Schwarz, H., Marpe, D., and Wiegand, T. (2007). Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(9):1103–1120.

[Smadi et al., 2009] Smadi, M., Ghosh, S., Farid, A., Todd, T., and Hranilovic, S. (2009). Free-Space Optical Gateway Placement in Hybrid Wireless Mesh Networks. *Lightwave Technology, Journal of*, 27(14):2688 –2697.

[Sobrinho, 2002] Sobrinho, J. L. (2002). Algebra and Algorithms for QoS Path Computation and Hop-by-hop Routing in the Internet. *IEEE/ACM Trans. Netw.*, 10(4):541–550.

[Stevenson et al., 2009] Stevenson, C., Chouinard, G., Lei, Z., Hu, W., Shellhammer, S., and Caldwell, W. (2009). IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard. *Communications Magazine, IEEE*, 47(1):130 –138.

[Subramanian et al., 2006] Subramanian, A., Buddhikot, M., and Miller, S. (2006). Interference Aware Routing in Multi-radio Wireless Mesh Networks. *Wireless Mesh Networks, 2006. WiMesh 2006. 2nd IEEE Workshop on*, pages 55–63.

[Tang and Tianfield, 2006] Tang, H. and Tianfield, H. (2006). Self-Organizing Networks of Communications and Computing. In *International Transactions on Systems Science and Applications*, volume 1, pages 421–431.

[Tang et al., 2005] Tang, J., Xue, G., and Zhang, W. (2005). Interference-aware Topology Control and QoS Routing in Multi-channel Wireless Mesh Networks. In *MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pages 68–77, New York, NY, USA. ACM.

[Tschudin et al., 2004] Tschudin, C., Gold, R., Rensfelt, O., and Wibling, O. (2004). LUNAR: a Lightweight Underlay Network Ad-hoc Routing Protocol and Implementation. In *in Proceedings of Next Generation Teletraffic and Wired/Wireless Advanced Networking (NEW2AN). St. Petersburg, 2004.*

[Vlavianos et al., 2008] Vlavianos, A., Law, L., Broustis, I., Krishnamurthy, S., and Faloutsos, M. (2008). Assessing Link Quality in IEEE 802.11 Wireless Networks: Which is the right metric? In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–6.

[Waharte et al., 2006] Waharte, S., Boutaba, R., Iraqi, Y., and Ishibashi, B. (2006). Routing Protocols in Wireless Mesh Networks: Challenges and Design Considerations. *Multimedia Tools Applications*, 29(3):285–303.

[Waharte et al., 2008a] Waharte, S., Ishibashi, B., Boutaba, R., and Meddour, D. (2008a). Performance Study of Wireless Mesh Networks Routing Metrics. *Computer Systems and Applications, ACS/IEEE International Conference on*, 0:1100–1106.

[Waharte et al., 2008b] Waharte, S., Ishibashi, B., Boutaba, R., and Meddour, D.-E. (2008b). Interference-Aware Routing Metric for Improved Load Balancing in Wireless Mesh Networks. In *ICC*, pages 2979–2983.

[Wang et al., 2004a] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004a). Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.

[Wang et al., 2004b] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004b). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

[Wi-Fi, 2007] Wi-Fi (2007). The Wi-Fi Alliance. Available from: http://www.wimaxforum.org/home. Last visit: 12th november 2011.

[WiMAX, 2007] WiMAX (2007). The wimax forum. available from: http://www.wimaxforum.org/home. last visit: 12th november 2011.

[WiMAX, 2011] WiMAX (2011). The ieee 802.16 working group on broadband wireless access standards.

[Woo and Singh, 2001] Woo, S.-C. M. and Singh, S. (2001). Scalable Routing Protocol for Ad Hoc Networks. *Wirel. Netw.*, 7(5):513–529.

[Wu and cker Chiueh, 2007] Wu, G. and cker Chiueh, T. (2007). Passive and Accurate Traffic Load Estimation for Infrastructure-mode Wireless LAN. In *ACM MSWiM '07*, pages 109–116, New York, NY, USA. ACM.

[Wu et al., 2006] Wu, Z., Ganu, S., and Raychaudhuri, D. (2006). IRMA: Integrated Routing and MAC Scheduling in Multi-hop Wireless Mesh Networks. In *Second IEEE Workshop on Wireless Mesh Networks*, Reston, Virginia, USA. IEEE.

[Xiao, 2005] Xiao, Y. (2005). Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs. *Wireless Communications, IEEE Transactions on*, 4(4):1506 – 1515.

[Xie et al., 2008] Xie, B., Yu, Y., Kumar, A., and Agrawal, D. P. (2008). Load-balanced Mesh Router Migration for Wireless Mesh Networks. *Elsevier Journal of Parallel and Distributed Computing*, 68(6):825 – 839.

[Xiuchao, 2004] Xiuchao, W. (2004). Simulate 802.11b Channel Within NS2. Available in: http://www.comp.nus.edu.sg/w̃uxiucha/research/reactive/publication/Simulate80211ChannelWit Technical report, School of Computing, National University of Singapore.

[Yang et al., 2005a] Yang, Y., Wang, J., and Kravets, R. (2005a). Designing Routing Metrics for Mesh Networks. In *Proceedings of the IEEE Workshop on Wireless Mesh Networks (WiMesh). IEEE Press.*

[Yang et al., 2005b] Yang, Y., Wang, J., and Kravets, R. (2005b). Interference-aware Load Balancing for Multihop Wireless Networks. Available in: http://www.cs.uiuc.edu/research/techreports.php?report=UIUCDCS-R-2005-2526. Technical report, Tech. Rep. UIUCDCS-R-2005-2526. Department of Computer Science. University of Illinois at Urbana-Champaign.

[Yick et al., 2008] Yick, J., Mukherjee, B., and Ghosal, D. (2008). Wireless Sensor Network Survey. *Computer Networks*, 52(12):2292 – 2330.

[Yousefi et al., 2006] Yousefi, S., Mousavi, M. S., and Fathy, M. (2006). Vehicular Ad Hoc Networks (VANETs): Challenges and Perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761–766.

[Yu and Chong, 2005] Yu, J. and Chong, P. (2005). A Survey of Clustering Schemes for Mobile Ad Hoc Networks. *Communications Surveys and Tutorials, IEEE*, 7(1):32–48.

[Zhang and Davis, 2008] Zhang, C. and Davis, M. (2008). An Investigation of the Correlation between Passive Monitoring and Activie Probing Techniques used in WLAN Mesh Networks. In *Information Technology and Telecommunications Conference 2008 (ITT 2008).*

[Zhang and Zhang, 2008] Zhang, Q. and Zhang, Y.-Q. (2008). Cross-Layer Design for QoS Support in Multihop Wireless Networks. *Proceedings of the IEEE*, 96(1):64–76.

[Zhang et al., 2007] Zhang, Y., fei Wei, Y., kun Zhang, L., and de Song, J. (2007). Simulation and Analysis of Node Throughput using Smart Antenna in Wireless Mesh Networks. *The Journal of China Universities of Posts and Telecommunications (Elsevier)*, 14(2):28 – 32, 50.

[Zhao et al., 2010] Zhao, L., Al-Dubai, A., and Romdhani, I. (2010). Novel QoS-Aware Gateway Centralized Multi-hop Routing for Wireless Mesh Networks. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, pages 336 –342.

[Zhao et al., 2005] Zhao, S., Wu, Z., Acharya, A., and Raychaudhuri, D. (2005). PARMA: A PHY/MAC Aware Routing Metric for Ad-Hoc Wireless Networks with Multi-Rate Radios. In *WOWMOM '05: Proceedings of the Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks*, pages 286–292, Washington, DC, USA. IEEE Computer Society.

Coimbra

2012