



UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA



Analysis & Design of Cognitive Wireless Multihop Networks

TESIS PRESENTADA A LA FACULTAD DE INGENIERÍA DE LA
UNIVERSIDAD DE LA REPÚBLICA POR

Germán Capdehourat Longres

EN CUMPLIMIENTO PARCIAL DE LOS REQUERIMIENTOS
PARA LA OBTENCIÓN DEL TÍTULO DE
DOCTOR EN INGENIERÍA ELÉCTRICA.

DIRECTORES DE TESIS

Dr. Pablo Belzarena Universidad de la República
Dr. Federico La Rocca Universidad de la República

TRIBUNAL

Dr. Eduardo Grampín Universidad de la República
Dr. Daniel Kofman Telecom ParisTech
Prof. María Simon Universidad de la República

REVISORES EXTERNOS:

Dr. Fernando Paganini Universidad ORT
Dr. Francisco Javier Simó Reigadas ... Universidad Rey Juan Carlos

DIRECTOR ACADÉMICO

Dr. Pablo Belzarena Universidad de la República

Montevideo
domingo 2 agosto, 2015

Analysis & Design of Cognitive Wireless Multihop Networks, Germán Capdehourat Longres

ISSN 1688-2784

Esta tesis fue preparada en L^AT_EX usando la clase iietesis (v1.0).

Contiene un total de 168 páginas.

Compilada el domingo 2 agosto, 2015.

<http://iie.fing.edu.uy/~gcapde/tesis.pdf>

Learning never exhausts the mind.

LEONARDO DA VINCI

This page was intentionally left blank.

Acknowledgements

This work was partially supported by Plan Ceibal (Centro Ceibal para el Apoyo a la Educación de la Niñez y la Adolescencia), Comisión Sectorial de Investigación Científica de la Udelar (CSIC) with the R&D project “Algoritmos de control de acceso al medio en redes inalámbricas”, Agencia Nacional de Investigación e Innovación del Uruguay (ANII) with the PhD grant POS_2011.1_3525, and French National Research Agency (ANR) under two projects: the MAITRE project of the STIC AmSud program, and the RESCUE project of the VERSO program.

Luego de los agradecimientos formales, paso a los agradecimientos personales, donde la lista es larga y probablemente incompleta. El orden es meramente accidental y no refleja necesariamente la importancia relativa de cada uno.

En primer lugar debo agradecer a la familia, por el apoyo fundamental para realizar el doctorado. Sin la compañía de Belén, y en estos últimos meses también de la pequeña Inés, nada de esto hubiera sido posible.

Seguidamente debo hacer mención a Fiorella Haim, quien tuvo la idea original de que yo realizara mi doctorado relacionado con el Plan Ceibal, así como propició el apoyo brindado por la institución para llevarlo adelante.

Luego, agradezco a mis tutores Pablo Belzarena y Federico Larroca, quienes me brindaron su apoyo incondicional durante todos estos años, así como la guía necesaria para llevar este trabajo a buen puerto.

A lo largo de este recorrido, que no empezó ni termina con esta tesis, la familia del IIE ocupa un lugar más que relevante. La historia empieza junto a Mauricio Delbracio, dando nuestros primeros pasos guiados por Alicia Fernández y Federico Lecumberry. Más acá en el tiempo tuve la suerte de trabajar varios años con Pablo Musé, con quien aprendí mucho del *savoir-faire* de la investigación académica. La lista incompleta incluye a Ignacio Ramírez, Juan Cardelino, Álvaro Gómez, Gabriel Gómez, Paola Bermolen, Laura Aspirot, Juan Pechiar, Pablo Cancela, Pablo Sprechmann, Pablo Monzón, Álvaro Giusto, Andrés Corez, Rodrigo Alonso, Martín Rocamora, Gregory Randall, Isabel Amigo, Cecilia Aguerrebere, Marcelo Fiori, Ignacio Irigaray, Fernanda Rodríguez, Víctor González Barbone, Claudina Rattaro, y tantos otros. Además en esta lista, si bien no es posible poner a todos por un tema de espacio, no pueden faltar los nombres de Maxi, Laura y María.

Agradezco también a los compañeros del Plan Ceibal, en particular del área técnica, quienes directa o indirectamente, colaboraron con este trabajo.

Por último agradezco a los integrantes del tribunal, quienes amablemente aceptaron evaluar esta tesis.

This page was intentionally left blank.

A la memoria de la Abuela Susi.

This page was intentionally left blank.

Abstract

In the last decade, the deployment of wireless local area networks (WLANs) based on the IEEE 802.11 standard [1] has grown significantly, mainly due to the use of unlicensed frequency bands. It also became the de facto standard for the so-called wireless mesh networks (WMNs) [2–4]. In this context, a typical application of this type of wireless solutions is to provide Internet access in suburban and rural areas. This kind of networks was useful for the deployment of Plan Ceibal [5], the nationwide implementation in Uruguay of the novel one-to-one educational model, which was the main motivation to carry on this thesis. The issue is how to optimize these networks to meet the growing traffic demands and the new requirements imposed by the future applications. In this regard, we start this thesis presenting a characterization and statistical model of a WLAN, on the one hand analyzing the demand in the educational context of Plan Ceibal, and on the other hand estimating the capacity of 802.11-based wireless links from measurements of the physical layer.

In the second part of this work, we propose a new routing and forwarding scheme for multihop wireless networks, based on the development of a statistical model of the links' queues, learned from live network measurements. We address the problem of deciding the most suitable path for the packet flows between each source-destination pair. We seek an optimal solution, capable of balancing the traffic load based on the resources available in each link. We present a suitable algorithm that solves the optimization problem posed in a distributed manner. Several simulations in different scenarios were performed to verify the performance of the proposed method, and also to compare with other schemes. In all the simulations, independently of the topology size, we observed a quick adaptation of the proposed algorithm to traffic changes and also a stable operation, avoiding the routing oscillations of the routing method included in the 802.11s standard, already noticed before by [6, 7].

Finally, the last part of the thesis is founded on the cognitive radio networks (CRNs) paradigm [8]. In this case we propose a novel robust spectrum allocation mechanism that takes advantage of the free spaces in licensed bands to expand the network resources (e.g. spectrum holes in TV bands). The introduced robust method was evaluated by several simulations for different network topologies. The results show that in all cases the robust approach ensures compliance with the effective capacity required on each wireless link with high probability. The additional spectrum for this robustness is below 35 % more than the optimality bound, given by the case of knowing in advance the primary users activity.

This page was intentionally left blank.

Resumen

Hace ya un buen tiempo que las redes inalámbricas constituyen uno de los temas de investigación más estudiados en el área de las telecomunicaciones. Un reflejo del avance logrado en el tema es el gran número y variedad de redes de estas características que encontramos desplegadas en el mundo hoy día. Sin embargo, el crecimiento de las demandas de tráfico y los tipos de servicios que dichas redes deben ser capaces de manejar hacen que la investigación en este tópico esté muy lejos de su fin. En la última década el despliegue de redes inalámbricas de área local (*Wireless Local Area Networks* o WLANs) basadas en el estándar IEEE 802.11 [1] ha crecido en gran forma, principalmente debido al uso de bandas de frecuencia no licenciadas. Esta solución tecnológica se ha convertido en la preferida, tanto para redes pequeñas y medianas como hogares, empresas, hoteles, así como redes de gran escala como campus universitarios y despliegues municipales. Esto ha generado un gran interés en la comunidad científica, que ha realizado diversos trabajos de investigación en este tipo de redes, entre los que se destacan estudios de modelado, estimación de la capacidad y análisis de la interacción con las capas superiores [9–13].

Este gran crecimiento en el despliegue de redes basadas en 802.11 trajo como efecto secundario, debido principalmente a los bajos costos del equipamiento de radio, que el uso de esta tecnología se extendiera a otras aplicaciones para las que no había sido pensada originalmente, como enlaces de largas distancias o escenarios con gran concentración de usuarios. Además se convirtió en el estándar de facto para las denominadas redes inalámbricas en malla (*Wireless Mesh Networks* o WMNs) [2–4], redes en las cuales todos los nodos pueden comunicarse directamente con otros nodos (siempre que se escuchen mutuamente). Esta solución ha cobrado interés para extender la cobertura de las WLANs, así como para la implementación de redes de sensores y otro tipo de aplicaciones. Este nuevo desafío llevó a que una tecnología que originalmente fue diseñada para un solo salto (*single-hop*) inalámbrico, el que va del cliente al punto de acceso, ahora sea utilizada para redes inalámbricas con múltiples saltos (*multi-hop*).

En este marco, una aplicación de este tipo de soluciones inalámbricas es el acceso a Internet en zonas suburbanas y rurales, un tema de particular interés para el despliegue del Plan Ceibal [5], lo que constituyó la principal motivación para el desarrollo de esta tesis. El Plan Ceibal corresponde a la implementación en Uruguay del modelo educativo uno a uno a nivel nacional, lo que implica la necesidad de brindar acceso a Internet en la totalidad de los centros educativos del país, tanto a nivel de enseñanza primaria como secundaria. En ese contexto las redes

inalámbricas con múltiples saltos son de gran importancia, puesto que constituyen la tecnología de acceso para un número importante de centros educativos ubicados fuera de las zonas urbanas. Además, este tipo de solución inalámbrica también se utiliza como red de acceso para brindar conectividad en espacios públicos, tanto interiores como exteriores.

En base a lo anterior, esta tesis se desarrolla sobre la base de dos ejes fundamentales. Por un lado, como motivación principal de la investigación y potencial aplicación objetivo de los algoritmos propuestos, se busca una solución de acceso inalámbrico de bajo costo, orientada particularmente a zonas suburbanas y rurales. Esta solución debe cumplir los requerimientos necesarios para ser la infraestructura de soporte que permita el desarrollo de un programa socio-educativo como el Plan Ceibal. Por otro lado, la tesis se enfoca particularmente en redes inalámbricas con múltiples saltos, basadas en el estándar 802.11 y operando en bandas no licenciadas. Estas soluciones con una tecnología de bajo costo y con diversos despliegues ya existentes, se destacan como una alternativa factible para cumplir con los requerimientos planteados y con buenas perspectivas a futuro.

A lo largo de esta tesis, se destacan tres grandes áreas de trabajo. En primer lugar tenemos la caracterización y modelado estadístico de redes WLAN, por un lado analizando la demanda en el contexto educativo del Plan Ceibal, y por otro lado estimando la capacidad de los enlaces inalámbricos a partir de medidas de capa física. Luego, se desarrolló un nuevo esquema de ruteo y encaminamiento de paquetes para una red inalámbrica con múltiples saltos, basado en un modelo estadístico de las colas de los enlaces inalámbricos, obtenido mediante aprendizaje automático con medidas de la propia red. Por último, en base al nuevo paradigma de redes cognitivas, se propuso una asignación de espectro robusta, que tiene en cuenta tanto el espectro no licenciado como los huecos disponibles en bandas licenciadas (ej: espacios libres en bandas de televisión).

Entrando más en profundidad en cada uno de los puntos abordados, en primer lugar tenemos el análisis de una red inalámbrica como la del Plan Ceibal, infraestructura que soporta el modelo educativo uno a uno a escala nacional. El objetivo en esta primera parte de la tesis es la descripción del caso de uso que motiva el resto del trabajo, así como identificar algunos puntos importantes en lo que hace a la demanda que debe soportar la red. En particular, se destacan dos resultados relevantes en este análisis. Por un lado la constatación de que la demanda crece muy rápidamente, lo que hace necesario aumentar la capacidad de la red para poder hacer frente a los requerimientos futuros. Por otro lado, si bien es cierto que la demanda va en aumento, y por lo tanto no es estacionaria, si se observan los datos a una escala de tiempo menor (ej: un día o una semana) el comportamiento sí es bastante estable, lo que hace posible caracterizar con buena precisión la demanda en base a datos estadísticos. Esto es importante, ya que permite el desarrollo de algoritmos de asignación de recursos que estimen la demanda en base a datos de la propia red y que luego hagan una asignación óptima en base a los recursos disponibles.

Luego, para completar la primera parte de la tesis, donde por un lado vimos la posibilidad de caracterizar la demanda de tráfico, en la segunda parte se trabaja

en la caracterización de la capacidad de una red inalámbrica. La idea en este caso es poder estimar y predecir, a partir de medidas de la capa física de la red, cuál es la capacidad resultante a nivel de capas superiores. En base a los resultados obtenidos, se verifica que es posible realizar una inferencia estadística de la capacidad efectiva con tráfico TCP de un enlace inalámbrico 802.11, a partir de medidas de capa física como la potencia de la señal recibida (RSS) o la relación señal a ruido resultante (SNR). Esta posibilidad de estimar en forma adecuada la capacidad en capas superiores de los enlaces inalámbricos nos abre las puertas al desarrollo de algoritmos de asignación de recursos que tengan en cuenta, ya no solamente los recursos a nivel de cantidad de espectro asignado a un cierto enlace, sino directamente la capacidad que es posible obtener dadas las características físicas en un lugar y momento determinado. De esa forma es posible hacer una asignación de recursos óptima, estimando los recursos disponibles para cada enlace, los cuales se infieren a partir de medidas obtenidas de los enlaces de la propia red.

La segunda parte de la tesis se centra en el problema del ruteo y encaminamiento de los paquetes en una red inalámbrica con múltiples saltos, buscando además una solución con capacidad de balancear la carga por los distintos caminos posibles entre cada par origen-destino. Si bien la versión 802.11s del estándar [14] define un protocolo de ruteo que considera este nuevo escenario de múltiples saltos e incorpora funcionalidades para resolver el encaminamiento de los paquetes, el problema aún sigue abierto y es una tema de investigación importante en el área. En particular existen diferencias importantes respecto al caso del ruteo en redes cableadas, que hacen que los paradigmas utilizados anteriormente no sean válidos en entornos inalámbricos. Por ejemplo se plantea la posibilidad de que la interacción entre las capas de acceso al medio y la capa de red, así como capas superiores, sea tomada en cuenta en este caso. Algunos trabajos presentan nuevas métricas que tienen en cuenta las capas inferiores para resolver el ruteo [15, 16], mientras que otros exploran la manera óptima de establecer los enlaces para resolver el problema [17–20] y otras alternativas buscan mecanismos de optimización *cross-layer* [21].

Por otro lado, los avances tecnológicos en materia de equipamiento permiten día a día nuevas posibilidades y cambian el escenario de análisis para este tipo de redes. En particular, estudios realizados para nodos que cuentan con una única interfaz de radio, que por lo tanto pueden hablar en un solo canal por vez, deben extenderse al caso con nodos que cuentan con múltiples interfaces de radio y por lo tanto pueden operar en diferentes canales de frecuencia simultáneamente [22, 23]. Esto permite mejorar el desempeño al aumentar sustancialmente la capacidad la red, pero a su vez cambia el escenario de análisis para resolver la asignación de recursos, por lo que nuevos algoritmos deben desarrollarse. Nuestro trabajo considera nodos de estas características, es decir, que cuentan con múltiples interfaces de radio, ya que es lo habitual en el equipamiento para redes de acceso que se fabrica en la actualidad.

En esta línea, se formuló el problema de la asignación de caminos para los paquetes entre cada par origen-destino de la red, buscando una solución capaz de balancear la carga en forma óptima en base a los recursos disponibles en cada

enlace. Para ello, se desarrolló un modelo de la cola de paquetes de cada enlace inalámbrico, basado en aprendizaje estadístico a partir de medidas tomadas de los propios enlaces de la red. Mediante este modelo, se planteó un problema de optimización definiendo como función objetivo a minimizar, a la suma de la cola de paquetes en todos los enlaces de la red. De esta forma, la asignación de caminos óptima será aquella que lleve la red a operar con la mínima congestión media.

A partir del problema de optimización planteado se presentó un algoritmo que permite resolver el problema en forma distribuida. Se realizaron diversas simulaciones en distintos escenarios para verificar el funcionamiento del método propuesto, así como la comparación con otros esquemas, como por ejemplo el ruteo del estándar 802.11s. Los resultados fueron analizados teniendo en cuenta diversas medidas de desempeño como el retardo y sus fluctuaciones para tráfico UDP y la capacidad efectiva promedio para tráfico TCP. En los experimentos realizados se observan claramente las ventajas del método propuesto frente a un esquema de ruteo con métricas dinámicas como el que se incluye en el estándar 802.11s. En todas las simulaciones, independientemente del tamaño de la red, se logra una rápida adaptación del algoritmo desarrollado frente a cambios en el tráfico, mostrando una operación estable, evitando así las oscilaciones de ruteo que presenta 802.11s, ya advertidas anteriormente por [6, 7].

En la tercera y última parte de esta tesis, se trabaja en otro problema relevante para las redes de acceso inalámbricas con múltiples saltos, que corresponde a la asignación de espectro para cada uno de los enlaces de la red [24]. Mientras que en la segunda parte de la tesis se desarrolló un método para aprovechar de la mejor forma posible los recursos disponibles de la red, en ese caso asignando en forma óptima los caminos posibles a cada uno de los flujos de la red, ahora se busca ampliar los recursos disponibles, apelando en este caso al nuevo paradigma de redes cognitivas. Esta idea, que apareció hace ya varios años [8], busca romper con el esquema tradicional de asignación de espectro de bandas licenciadas para uso exclusivo, permitiendo que los usuarios secundarios puedan usar aquellas bandas licenciadas cuando los usuarios primarios no están presentes.

El éxito de diversos estándares en bandas libres, destacándose la relevancia mundial de 802.11 en estos días, ha llevado a la necesidad de contar con más espectro no licenciado. Una alternativa para ello, es justamente el desarrollo de radios cognitivos, que sean capaces de operar también en bandas licenciadas cuando ellas no están en uso. Un ejemplo de los esfuerzos en esta línea son las tecnologías que usan los huecos disponibles en bandas de televisión, para lo cual ya existen hoy varios países con la regulación que lo ampara y estándares como el 802.11af para su implementación. En esta línea, nuestro trabajo aborda el problema de la asignación de espectro para una red cognitiva con múltiples saltos.

Para la formulación del problema se asume una asignación de frecuencias periódica para cada enlace, cada cierto intervalo de tiempo fijo predefinido. Se define como función objetivo la suma de los costos asociados a cada una de las bandas de frecuencia asignadas para cada enlace. Además, se asume la posibilidad de estimar, de manera similar a lo realizado en la primera parte de la tesis, las capacidades efectivas que se logran en cada banda de frecuencia para cada enlace inalámbrico,

a partir de medidas recolectadas en los propios nodos de la red. Por último, se impone como restricción la necesidad de cumplir con cierta demanda de tráfico a cumplir en cada uno de los enlaces. De esta forma, se llega a un problema de optimización estocástico, debido al componente aleatorio que significa la disponibilidad variable que ofrecen las bandas licenciadas, ya que dependen de la actividad de los usuarios primarios en dichas bandas.

Para resolver este problema estocástico se utiliza un equivalente determinístico que es robusto a la distribución de probabilidad [25]. Esto implica que la solución propuesta tiene en cuenta solamente la media y la varianza de la actividad de los usuarios primarios, valores que se estiman a partir de las medidas recolectadas de la actividad en cada banda de frecuencia. De esa forma es posible, mediante un esquema probabilístico, encontrar la mejor solución para un valor del riesgo predefinido. Además, el esquema propuesto permite encontrar la asignación óptima mediante un algoritmo descentralizado, y se presenta una arquitectura adecuada para su implementación.

El método robusto propuesto fue evaluado mediante diversas simulaciones para distintas topologías de red. Además, se comparó el desempeño con otras alternativas como considerar únicamente las bandas no licenciadas o utilizar una solución basada en el valor medio. También se tuvo en cuenta en los experimentos una cota de optimalidad dada por la asignación correspondiente al caso de saber la actividad de los usuarios primarios con antelación a su ocurrencia. Los resultados muestran que la solución propuesta tiene mucho mejor desempeño que un esquema basado en el valor medio, sin asignar mucho más espectro. El enfoque robusto garantiza el cumplimiento de la capacidad efectiva necesaria en cada enlace inalámbrico con alta probabilidad. El espectro adicional necesario para esta robustez está por debajo del 35 %, con respecto a la cota de optimalidad dada por la asignación que se logra sabiendo con antelación la actividad de los usuarios primarios.

Como se dijo anteriormente, el foco de esta tesis está centrado en tecnologías de acceso para escenarios suburbanos y rurales, donde las redes inalámbricas con múltiples saltos ya han demostrado ser una solución factible a través de varios despliegues existentes en la actualidad. El problema planteado es cómo optimizar estas redes para hacer frente a las crecientes demandas de tráfico y cumplir con los requerimientos que el futuro nos impone. En este sentido, dos aspectos fueron abordados en nuestro trabajo. Por un lado, se trabajó en el desarrollo de herramientas de ingeniería de tráfico para maximizar la explotación de los recursos de la red, a través de un encaminamiento óptimo de los paquetes considerando los flujos origen-destino. Por otro lado, en base al paradigma de redes cognitivas, se desarrolló un nuevo marco de trabajo para la asignación de espectro y un algoritmo robusto que permite aprovechar los espacios libres en bandas licenciadas para ampliar los recursos de la red.

Hace algunas décadas, la llegada de las redes basadas en la conmutación de paquetes llevaban a predecir un futuro de convergencia, donde todos los servicios (datos, voz, video) serían transmitidos por una única red. Este proceso llevó un tiempo, pero la realidad es que hoy podemos decir que finalmente la convergencia ha llegado y el nuevo paradigma dejó atrás las viejas redes basadas en conmutación

de circuitos. Estas redes permitían mucho menor flexibilidad para administrar los recursos, pero por otro lado garantizaban con certeza los requerimientos de extremo a extremo. La flexibilidad que trajeron las redes de paquetes, si bien fue exitosa desde los inicios para servicios de datos, muchos esfuerzos debieron llevarse adelante para implementar mecanismos adecuados de asignación de recursos, que permitieran cumplir con los requerimientos necesarios de los distintos servicios que confluyen en la misma red.

Actualmente, con el nuevo paradigma de las redes cognitivas, vemos cierta analogía con la situación antes descrita. Claramente, este nuevo enfoque permite mucho mayor flexibilidad para asignar los recursos, incorporando el espectro como una nueva variable. Sin embargo, también es claro que para manejar este nuevo grado de libertad, es necesario desarrollar los mecanismos adecuados para mantener el cumplimiento de calidad de servicio necesario, en particular para aplicaciones con requerimientos exigentes. Esta tesis es una contribución en esta dirección, y probablemente, si bien el camino puede ser largo y duro, las redes cognitivas van a convertirse algún día en realidad, de la misma forma que las redes convergentes lo son hoy en día.

Contents

Acknowledgements	III
Abstract	VII
Resumen	IX
1. Introduction	1
1.1. Context and Evolution	1
1.2. Wireless Mesh Networks	2
1.3. Cognitive Radio Paradigm	5
1.4. Summary and Main Contributions	6
1.5. Structure of the Thesis	9
I Motivation and WiFi Performance Basics	11
2. The Plan Ceibal Use Case Scenario	13
2.1. Introduction	13
2.2. Plan Ceibal's Description	15
2.2.1. Plan Ceibal's Wireless Network	16
2.3. Network Measurements	18
2.4. Statistical Data Analysis	19
2.4.1. Long and Mid-term Analysis	19
2.4.2. Analysis by Type of Institution	21
2.4.3. Outdoor Usage Analysis	25
2.5. Final Comments and Reflections	26
3. WiFi Performance Estimation from PHY Layer Measurements	29
3.1. Introduction	29
3.2. Machine Learning Algorithms	31
3.2.1. Support Vector Regression	31
3.2.2. k -Nearest Neighbors	32
3.3. Experimental Platform and Dataset	32
3.3.1. Wireless Testbed Description	33
3.3.2. Experimental Protocol	34
3.3.3. Measures Defined as Features	36

Contents

3.4. Throughput Estimation from PHY Measurements	37
3.4.1. Machine Learning-based Methodology	37
3.4.2. Estimation Accuracy	38
3.4.3. Experimental Results	38
3.4.4. Final Remarks	41
Conclusions of Part I	43
II Dynamic Load-Balancing in Wireless Networks	45
4. Routing and Forwarding in Wireless Mesh Networks	47
4.1. Introduction	48
4.2. Related Work	50
4.3. Network Model and Problem Formulation	52
4.4. Learning the Wireless Link Dynamics	54
4.4.1. Convexity Assumption	55
4.4.2. Average Queue Regression Example	57
5. Minimum Queue Length Load-Balancing	61
5.1. Convex Optimization Problem Formulation	61
5.2. Distributed Algorithm Proposal	64
5.3. Discussion on Implementation Issues	66
5.4. Simulation Experiments	68
5.4.1. Multipath Forwarding: 3-nodes Topology	69
5.4.2. Gateway Selection Problem	74
5.4.3. Multipath Forwarding: 4-nodes Topology	77
5.4.4. Gateway Selection: 25-nodes Topology	81
Conclusions of Part II	83
III Spectrum Assignment in Cognitive Networks	85
6. Spectrum Allocation in Cognitive Radio Networks	87
6.1. Introduction	88
6.2. Related Work	90
6.3. Network Model and Problem Formulation	92
6.3.1. Single Collision Domain	93
6.3.2. Model Extension for a Wireless Mesh Network	94
7. Chance Constrained Optimization	97
7.1. Introduction	97
7.2. Problem Formulation and Probabilistic Constraints	98
7.3. Random Data with Gaussian Distribution	99

7.4. Distributionally Robust Approach	100
8. Robust Spectrum Allocation	103
8.1. Distributionally Robust Solution Proposed	103
8.1.1. Single Domain Spectrum Allocation	104
8.1.2. Extension for the Multiple Domain Case	106
8.2. Distributed Algorithm Architecture	106
8.2.1. Distributed Optimization	107
8.2.2. Proposed Algorithm Architecture	109
8.2.3. Implementation Issues	110
8.3. Simulation Experiments	112
8.3.1. Single Domain Spectrum Allocation	114
8.3.2. Multi-Domain Case: 3-links Network Topology	120
8.3.3. Real World Network Topology	123
Conclusions of Part III	129
9. General Conclusions and Perspectives	131
Bibliography	146

This page was intentionally left blank.

Chapter 1

Introduction

1.1. Context and Evolution

About 50 years ago, the first research steps on packet switched networks were being carried out, probably without knowing the global impact they would have in the world. Nowadays, the importance acquired by the Internet is unquestionable, and beyond some economic discussions, virtually everyone recognizes the relevance it gained in the various aspects of the lives of people around the planet. Many studies have found a direct relationship between the broadband access and the investment in information and communication technology (ICT) and the job creation, economic growth and human development index of countries [26–28]. Moreover, in many cases ICT applications are promoted as a way to make significant progress in developing countries [29, 30]. In this sense, education emerges as one of the most important verticals for incorporating ICT, with various initiatives already underway in various countries and regions (e.g. Plan Ceibal [5] in Uruguay, Conectar Igualdad [31] in Argentina, ConnectED [32] in the US and many more).

In this context, the progress in broadband deployment is key in advancing the reduction of the digital divide [33]. The deployment, maintenance and optimization of the necessary infrastructure is a tough challenge, and an additional complexity is added, as bandwidth requirements evolve every day. For example, the Federal Communications Commission (communications regulator in the US, which implies a great influence worldwide) includes as part of its 2015 Broadband Progress Report [34] the redefinition of the term *broadband*, by raising the minimum download speeds needed from 4 Mbps to 25 Mbps, and the minimum upload speed from 1 Mbps to 3 Mbps. This simple redefinition tripled the number of US households without broadband access, now reaching a total of 17% of the population. This situation is much worse in rural areas where the percentage reaches 53% of the people lacking access to 25 Mbps/3 Mbps services. The previous definition of broadband (4 Mbps/1 Mbps) was from 2010, just five years ago, a clear sign of the challenge of keeping up with the necessary infrastructure, with requirements that change so much and so fast.

In those suburban or rural areas, where less people have access to broadband,

Chapter 1. Introduction

is where it becomes more difficult to afford the necessary infrastructure. The low population density and large swathes of territory to cover, make infeasible the deployment of optical fiber due to the high costs, particularly for developing countries with lower budgets. For those cases, the only affordable alternative to build the necessary infrastructure is to use wireless technologies [35]. This advantage from the economic point of view becomes a major challenge from the technical point of view, since it is necessary to develop wireless solutions, capable to cope with the high bandwidth requirements, but which are based on a finite resource: the shared spectrum available.

Wireless technologies have evolved dramatically in recent years. On the one hand we have the emergence of mobile telephony over 20 years ago and its massive deployment worldwide, mainly since the success of GSM, today extended to 3G and 4G, providing also broadband data services. On the other hand, the regulatory decisions taken many years ago today see their fruits, given the success and consolidation of various standards in unlicensed bands, particularly IEEE 802.11, popularly known as WiFi. The latter has had an impact not only on the last hop connectivity for which it was originally intended, but also as last-mile access technology. This alternative low-cost access technology has even led several research efforts focused on variations to the standard to enable and improve the operation on long distance links [36–38]. These works have not been limited to laboratory testing, but have also deployed the technologies developed, enabling applications such as telemedicine in rural areas [39].

In this ongoing evolution of wireless technologies, much research effort has been recently dedicated to what is known as mesh networks [2]. More recently, many research works focused in this type of mesh solutions, have also incorporated the cognitive radio paradigm, introduced by Mitola 15 years ago [8]. In the next sections we briefly describe both concepts, wireless mesh networks and the cognitive radio paradigm, which provide the general framework for the research to be presented in this thesis. Our work will focus on these novel wireless solutions, inspired on the motivation previously raised about the need of infrastructure to provide ICT access to bridge the digital divide, particularly in education. With that inspiration in mind, the two main research areas of our work are: firstly, to exploit as much as possible the available resources using traffic engineering strategies; and secondly, to take advantage of idle resources to get more capacity, based on dynamic spectrum allocation under the cognitive radio paradigm.

1.2. Wireless Mesh Networks

The typical architecture of a wireless mesh network (WMN) is depicted in Figure 1.1. As we can see, it consists of several intermediate nodes (typically stationary) called wireless mesh routers. These nodes enable the interconnection between the end users (mesh clients), also often called stations, which can be mobile or stationary, with one or more Internet gateways, which are usually equipped with a (typically wired) broadband Internet access. A broad literature in the area can be found in this book [40], and also in the previously cited survey [2].

1.2. Wireless Mesh Networks

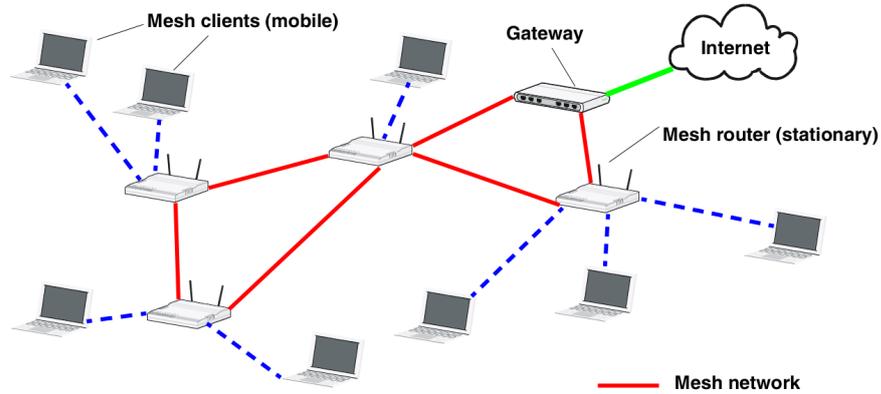


Figure 1.1: Wireless Mesh Network.

A brief historical perspective on WMNs, taken from the Communication Systems Group at ETH Zurich, mentioned as the first antecedents to the WMNs the research done by the DoD on battlefield communications in infrastructureless hostile environments, during the 70s and 80s. The main characteristics were a MAC layer mechanism combining Aloha & CSMA, distance vector routing, operating in the frequency range 1.78 - 1.84 GHz and with data rates ranging from 100 to 400 kbps. More recently, during the 90s, another project from the DoD focused on multimedia communications with hand-held devices in the office environment, with more advanced characteristics: CSMA/CA and TDMA, several routing and topology control schemes, operating frequency range 225-450 MHz and a data rate of 300 kbps.

In the standardization field, we can mention the creation of the IETF - Mobile Ad Hoc Networking (MANET) working group in 1997 and the IEEE - 802.11s working group in 2004. The great success of the IEEE 802.11 standard [41] in the WLAN environment, brought as a side effect, mainly due to lower prices of radio equipment, to extend its usage to WMNs, becoming the de facto standard for this kind of networks [2-4]. This fact imposed a new challenge to this technology, that was originally designed for a single hop wireless network (between clients and access points), and now it is being used in multihop wireless networks. The new version of the IEEE standard, the 802.11s [14], addressed this new multihop scenario, particularly focused on new features to solve the packet routing and forwarding.

While the new version of the standard defines a routing protocol, this problem in such networks is still open and is an important research topic in the area. In a WMN, routing has significant differences with the wired case, that make the old paradigms used before not valid in wireless environments (e.g. the interaction between the MAC layer and the network and transport layer should be taken into account in this case). Some people have proposed new metrics that take into account the lower layers to solve the routing [15,16], while others explore which is

Chapter 1. Introduction

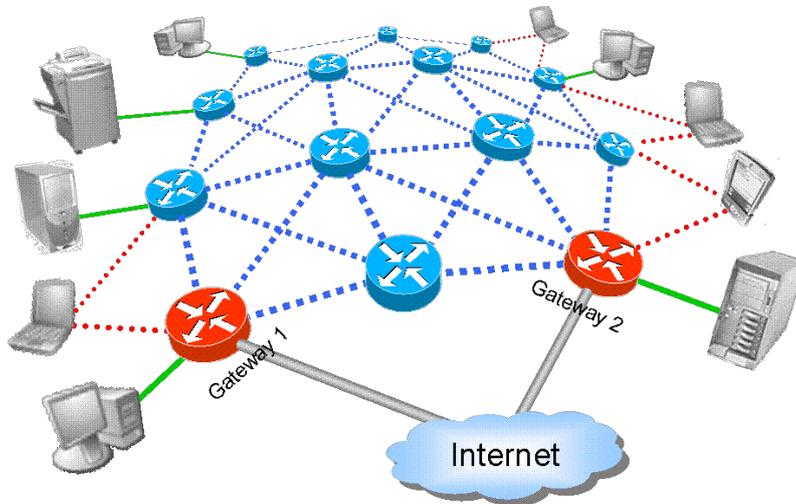


Figure 1.2: Gateway selection example in a wireless mesh network.

the best way to establish the links to solve the problem [17–20] or seek alternative cross-layer optimization mechanisms [21].

Furthermore, the technology advances fast, and allows new possibilities every day, which changes the scenario of analysis that should be considered. For example, previous studies analyzed the case of a WMN where nodes have a single radio interface, which implies that each node can talk on only one channel at a time. Then, typical WMNs equipment improved to nodes with multiple radio interfaces, which allows them to operate in different frequency channels simultaneously [22, 23]. The multi-radio multi-channel (MR-MC) networks presented a substantial performance improvement because of the great network capacity increase, but also changed the scenario of analysis to solve the resource allocation. This implies that, even for many previously well studied issues, new research should be done in order to consider the updated scenario.

Among several open issues in the development of WMNs we can highlight the following, which are of particular interest in this thesis. First, we have the problem (illustrated in Figure 1.2) of how to optimally choose the gateway dynamically [42]. Then, we already mentioned the routing and forwarding, for which traffic engineering tools should be developed, in order to enable load-balancing in real time, depending on the actual traffic demand [43,44]. Another issue is how to ensure a fair access for the multiple clients to the shared resources [45,46]. Finally, we are interested in finding algorithms that dynamically optimize resource allocation and maximize the utility of the network [47,48].

1.3. Cognitive Radio Paradigm

The cognitive radio (CR) paradigm was introduced by J. Mitola in his PhD dissertation in 2000 [8]. He described the characteristics of a novel radio interface, that used information from the environment in order to intelligently choose the best parameters to use for communications. Although unrealizable in the short term, it has converted in a unifying vision of how a future radio device might behave. However, several key factors point in the direction that the underlying technology for this (once utopic) idea of such a system is every day closer to become a reality.

The first proof that we are getting closer to see a real cognitive radio network are the recent advances of the research on the subject [49]. This new paradigm has generated an increasing research interest, with much industrial and academic efforts, which is reflected by the number of publications and conferences on the topic. In addition, the importance of wireless networks today is also increasing, with a demand of ubiquity access and also novel access technologies as WMNs. As an example, two novel standards were approved in recent years: the IEEE 802.22 standard [50] in 2011, which defines a Wireless Regional Area Network (WRAN), and the 802.11af amendment [51] in 2014, which enables the operation of 802.11-based networks in TV bands. Finally, software-defined radio (SDR) is becoming possible and increasingly capable, which seems to be the most suitable technology to support cognitive radios.

At the same time this new paradigm was being introduced, the regulatory agencies were acknowledging the need for a new way of spectrum allocation. The scarcity in some frequency bands, while others are unused, opens the door for a more dynamic system. This flexibility might be very helpful for WMNs, typically based on unlicensed spectrum, to expand the available resources through idle licensed bands. The main idea is to have two types of users; licensed or primary users (PUs), which have the preferential right to use the band; and unlicensed or secondary users (SUs), which can use the band only in the absence of the PUs. This type of spectrum allocation contributes to a more efficient use compared to traditional static assignments, as testified by some recent FCC rulings [52].

As described in depth in [53], there is a kind of symbiosis between WMNs and CR, which is closely related with the focus of this thesis. In particular, the ability to use any spectrum that is not being used could greatly improve WMNs performance, especially in rural areas. TV white spaces or bands assigned to cellular networks with low usage could be exploited by the WMNs, providing it with valuable additional bandwidth. However, as frequency use for these bands could be transient, the WMNs would result in a highly variable network capacity. In this thesis we address this issue through a novel robust approach, which aims to give greater compliance guarantees with the network requirements.

1.4. Summary and Main Contributions

As previously mentioned, the main focus of this thesis is on low-cost Internet access technologies for suburban and rural settings, particularly concerning on education and e-learning applications, where multihop wireless networks have already proven to be a feasible solution. The issue is how to optimize these networks to meet the growing traffic demands and requirements imposed by the future applications. In this regard, two important aspects were addressed in our work. On the one hand, we worked on novel traffic engineering tools, which seek to maximize the exploitation of network resources through optimal packet forwarding considering origin-destination flows. On the other hand, based on the cognitive radio networks paradigm, we introduced an original spectrum assignment framework, and a robust algorithm was proposed to solve the problem posed, which takes advantage of the free spaces in licensed bands to expand the network resources.

The thesis is divided in three main parts, each one dedicated on the different topics addressed in this work. The first one deals with the characterization and statistical modeling of a WLAN, firstly analyzing the demand in the educational context of Plan Ceibal, and secondly estimating the capacity of 802.11-based wireless links from measurements of the physical layer. Then, we propose a new decentralized scheme for routing and forwarding in multihop wireless networks, which optimally distributes the end-to-end traffic flows among all the possible paths of all origin-destination pairs. The algorithm is based on the development of a statistical model of the links' packet queues, obtained using machine learning techniques from live network measurements. Finally, founded on the new paradigm of cognitive networks, we propose a novel robust spectrum allocation mechanism, that takes into account both, the unlicensed spectrum, and the free spaces available in licensed bands (e.g. spectrum holes in TV bands).

The thesis starts with the description and analysis of the wireless network from Plan Ceibal, which is the supporting infrastructure to the one-to-one educational model nationwide in Uruguay. The aim of this first part is to present the use case that motivates the rest of the work, and also identify the key points in regard to the requirements that the network must support. In particular, two important findings are highlighted from the analysis. On the one hand, the confirmation that the traffic demand is growing very rapidly, making it necessary to increase the network capacity to cope with the future requirements. On the other hand, even in this scenario of an increasing demand, which is therefore not stationary, if we look at the data at a shorter time scale (e.g. a day or a week) the observed behavior is fairly steady, making it possible to characterize the demand with good accuracy based on statistical data. This is important because it allows the development of resource allocation algorithms that learn the demand based on data from the live network and then are able to perform an optimal allocation based on the available network resources.

While in the first half of this part of the thesis we confirm that is possible to characterize the traffic demand at short timescales, in the second half we analyze how to compute the capacity of a wireless network. The idea here is to estimate

1.4. Summary and Main Contributions

and predict, from measurements of the physical layer, what is the resulting capacity at higher layers. Based on the results obtained, we verify that it is possible to perform a statistical inference of the maximum achievable TCP throughput on a 802.11-based wireless link, from physical layer measurements of the received signal strength (RSS) or the resulting signal to noise ratio (SNR). This ability to properly estimate the capacity in upper layers of the wireless links opens the door to developing resource allocation algorithms that take into account, not only the resources considering the amount of spectrum assigned to a certain link, but directly knowing the capacity that is possible to obtain given the physical conditions in a certain time and location. Thus, it is possible to make an optimal resource allocation, considering the resources available for each link, which are learned from measurements obtained from the links of the live network.

The second part of the thesis focuses on the problem of routing and forwarding packets in a multihop wireless network. The goal is to find a distributed algorithm to decide the most suitable path for every packet flow between each origin-destination pair. We look for a solution capable of optimally balancing the traffic load, based on the available resources in each network link. For this purpose, a model of the packet queue for a 802.11-based wireless link was developed, based on statistical learning from measurements taken from the live network. Using this model, we pose an optimization problem which seeks to minimize the sum of the packet queues on all network links. This way, the optimal packet forwarding will conduct the network to operate at the minimum average congestion, thus exploiting in the best possible way the available resources.

Next, a suitable algorithm that solves the optimization problem posed in a distributed manner is presented. Several simulations were performed in different scenarios to verify the performance of the proposed method, and also for comparison with other schemes, such as the routing method of the 802.11s standard. The results were analyzed considering several performance measures such as delay and jitter for UDP traffic and average goodput for TCP traffic. The results in the experiments clearly show the advantages of the proposed algorithm compared with routing schemes based on a dynamic metric, as the one included in the 802.11s standard routing method. In all the simulations, independently of the topology size, we observed a quick adaptation of the proposed algorithm to traffic changes and also an stable operation, avoiding the routing oscillations of 802.11s, already noticed before by [6, 7].

In the third and final part of this thesis, our work is focused on another important issue for wireless access networks with multiple hops, which corresponds to the spectrum allocation for each of the links in the network. Now, the goal is to expand the available resources, based in this case on the cognitive radio networks paradigm, using both unlicensed bands and free spaces in licensed bands. This novel paradigm, allows secondary users to use licensed bands when the primary users are not present. The success of various standards in unlicensed bands, with a special mention to the global relevance of 802.11 these days, has led to the need for more unlicensed spectrum. An alternative to solve this situation, is precisely the development of cognitive radios, that are able to operate in licensed bands

Chapter 1. Introduction

as well, when they are not in use. An example of the efforts in this line are the technologies that use the free spaces in TV bands, for which today we already have standards like 802.11af for the implementation and several countries with the proper regulation to allow the operation.

For the formulation of the problem we consider a periodic assignment of frequency bands for each link, with a certain predefined fixed time interval. The objective function we seek to minimize is defined as the sum of the assignment costs associated with each frequency bands allocated for each link. Furthermore, we assume it is possible to estimate the effective capacity on each frequency band for each wireless link, similarly to what was introduced in the first part of the thesis, using measurements from the live network collected at each node. Finally, we impose the constraint that the effective capacity assigned should be enough to comply with certain traffic demand in each of the links. This way, we reach a stochastic optimization problem, due to the random variable component related to the variable availability of licensed bands, since it depends on the primary users activity. To solve this problem we seek a deterministic equivalent, that is robust to the probability distribution. This implies that the proposed solution takes into account only the mean and variance of the primary users activity, values estimated at each node from measurements collected on each frequency band. Thus, it is possible, using a probabilistic scheme, to find the best solution for a predefined value at risk.

The proposed robust method was evaluated by several simulations for different network topologies. In addition, the performance was compared with alternatives such as only considering unlicensed bands or an approach based on the average value. It was also taken into account in the experiments an optimality bound given by the solution corresponding to the case of knowing in advance the primary users activity. The results show that the proposed solution have much better performance than a scheme based on the average value, without assigning much more spectrum. The robust approach ensures compliance with the effective capacity required on each wireless link with high probability. The additional spectrum for this robustness is below 35% more than the optimality bound, given by the case of knowing in advance the primary users activity. Finally, we highlight that the proposed scheme enables to find the optimal allocation with a decentralized algorithm, and a proper architecture for its implementation is also presented.

The ultimate goal of the whole research carried out is to deepen in the analysis and design of suitable low-cost solutions for wireless access to the Internet in rural areas, aimed at supporting the traffic demands imposed by the new applications of the future, in particular the ones related to the education system under the novel one-to-one model. With this in mind, our line of work is particularly focused in resource allocation mechanisms that take advantage of live network measurements, in order to find suitable models based on statistical learning from the collected data. Then, using a problem formulation based on the NUM (Network Utility Maximization) framework, we look for the proper optimization schemes to reach the desired goals on each case, and we base on convex optimization and decomposition theory to find distributed solutions, which enables to reach the optimal network

operation point with decentralized algorithms.

1.5. Structure of the Thesis

The thesis is divided into three parts and is organized as follows. Part I relates to the use case and main motivation behind this work which is Plan Ceibal, as well as a basic statistic learning perspective to analyze 802.11-based wireless networks. First, Chapter 2, based on paper [CP4], introduces the Plan Ceibal one-to-one educational model approach and describes the wireless network infrastructure deployed. Then, in Chapter 3, which is based on paper [CP2], the performance estimation of a 802.11-based network is presented, using machine learning techniques to estimate the capacity in these kind of networks.

Secondly, Part II deals with the problem of finding an optimal forwarding in wireless mesh networks and is divided into two chapters. Chapter 4 covers the presentation of the problem, a brief literature review and introduces the nomenclature used. It also includes the developed model, based on the statistical learning of the average queue size function at each wireless link of the network. Then, Chapter 5 describes the proposed solution and illustrates the operation with several simulation experiments. The content of these chapters is based on the paper [CP1] and the article [JA1].

In Part III, another major problem addressed in this thesis is presented, which is the spectrum assignment in a cognitive network environment. The part is divided into three chapters. Chapter 6 describes the problem, presents a related work review and develops the resource allocation model to use. In Chapter 7 a brief review of stochastic optimization with chance constrained programming is presented, and introduces a particular distributionally robust approach to find a deterministic equivalent problem, which will be used in the proposed spectrum allocation solution. Then, Chapter 8 introduces the algorithm developed, describes the distributed architecture for a proper implementation and presents several simulation experiments to validate the proposed framework. The content of these chapters is based on the paper [CP3] and the article [JA2].

In order to close the thesis, in Chapter 9 the general conclusions of this work are presented and future prospects are discussed as well.

List of publications

International Conference Papers

[CP1] “Minimum queue length load-balancing in planned Wireless Mesh Networks”, **Germán Capdehourat**, Federico Larroca and Pablo Belzarena, *9th International Symposium on Wireless Communication Systems (ISWCS)*, August 28-31th 2012, Paris, France.

[CP2] “Predictive Estimation of Wireless Link Performance from Medium Physical Parameters Using Support Vector Regression and k-Nearest Neighbors”,

Chapter 1. Introduction

Guillaume Kremer, Philippe Owezarski, Pascal Berthou and **Germán Capdehourat**, *6th International Workshop on Traffic Monitoring and Analysis (TMA)*, April 14th 2014, London, UK.

[CP3] “Robust spectrum allocation for cognitive radio networks”, **Germán Capdehourat**, Federico Larroca and Pablo Belzarena, *11th International Symposium on Wireless Communication Systems (ISWCS)*, August 26-29th 2014, Barcelona, Spain.

[CP4] “Plan Ceibal’s Wireless Network for the Implementation of the 1:1 Educational Model”, **Germán Capdehourat**, Gonzalo Marín and Ana Rodríguez, *8th edition of the Latin America Networking Conference (LANC)*, September 18-19th 2014, Montevideo, Uruguay.

Journal Articles

[JA1] “Optimal multipath forwarding in planned wireless mesh networks”, **Germán Capdehourat**, Federico Larroca and Pablo Belzarena, *Computer Communications*, 38:36-49, 2014 (online October 2013, published January 2014).

[JA2] “Decentralized Robust Spectrum Allocation for Cognitive Radio Wireless Mesh Networks”, **Germán Capdehourat**, Federico Larroca and Pablo Belzarena, *recently accepted for publication in Ad Hoc Networks* (preprint submitted in December 2014).

Part I

Motivation and WiFi Performance Basics

Chapter 2

The Plan Ceibal Use Case Scenario

As discussed in the introductory chapter, wireless networks have experienced significant growth in recent years. This importance gained by wireless solutions is twofold. On the one hand, it is today the most common connectivity technology for a variety of end user devices, including laptops, tables, smartphones and even watches. On the other hand, it has also become a suitable last-mile access technology, particularly important to provide Internet access in remote rural areas. In this context, one of the most important applications of wireless networks, and a major vertical of the WiFi industry today, is education.

This chapter presents a real use case scenario which is the deployment of Plan Ceibal [5]. This wireless network corresponds to the supporting infrastructure of the nationwide implementation of the novel one-to-one educational model. We also present a usage analysis based on measurements collected from the wired side. This study is intended to characterize the traffic demand in this scenario and validate the hypothesis that this demand can be estimated accurately considering a proper time scale. This assumption will be used later on this thesis, for the development of optimal resource allocation methods based on live measurements from the operating network.

2.1. Introduction

During the past decade, the advances in wireless technologies suggest the possibility of a near future world where everyone is connected to the Internet everywhere. This fact has already been a reality in some places for the case of mobile telephony. Lately, it has spread to broadband services, mainly because of the massive deployment experienced by the 802.11 standard, better known as the popular WiFi. The number of places with wireless Internet access is increasing worldwide, either through *hotspots*, community networks [54] or municipal deployments [55]. In parallel to this, the world of end user devices has evolved as well, and smartphones, tablets and a huge variety of new devices have joined the traditional laptops.

Within this context, being education an important aspect of our society, it is in line with these developments and is being deeply impacted by technology.

Chapter 2. The Plan Ceibal Use Case Scenario

Regarding laptops, their low costs have cleared the way for the consideration of one-to-one educational models. Although prices are still a bit higher than those of the original Nicholas Negroponte's idea (introduced in the World Economic Forum in Davos in 2005) of getting USD 100 educational laptops, these have dropped significantly. Nowadays it is viable to endorse this kind of programs by using a small percentage of the national GDP (the case of Plan Ceibal is less than 0,2% of Uruguay's GDP, which represents less than 5% of the total public expenses in primary and intermediate basic education [56]).

Over the last years the one-to-one educational model has been gaining importance within different regions of the world. Aside from delivering a laptop for each student, providing wireless Internet access (first within educational centers and then expanded to public places as squares and clubs) has also changed the educational paradigm completely. The access to information is no longer limited to educational centers and technology empowers students to have all the knowledge available, all the time, everywhere. In this respect, the Uruguayan Plan Ceibal has been a pioneer, beginning its deployment in 2007 and being the first project of its kind in having a national scope. This new educational model, along with the technological breakthroughs of the last decade, have made of education one of the main applications of wireless networks nowadays. This is clear from the products portfolio of the main wireless industry providers, where education always takes an important place.

The deployment of WiFi wireless networks keeps growing. In spite of the development of wideband technologies for cellular networks such as LTE, it is expected that within the next years 802.11 networks will continue to grow and expand around the world [57, 58]. As a consequence, every important wireless industry supplier has 802.11 equipments among their portfolios. The relevance acquired by the impact generated at the educational sector makes this application one of the most important within solutions offered by manufacturers. Education also has an important technical challenge regarding wireless solutions, as it presents scenarios where technology reaches its limits. The high density of users in a reduced area, also with high traffic demands, generates situations only comparable with events such as large conferences or sport events [59].

Plan Ceibal has been a pioneer facing these kinds of scenarios, not only on a national level but also worldwide. It has one of the largest wireless networks in the country, comparable to that of a local ISP, with over 10.000 access points. It covers every public educational center along with some private schools that have joined the project, public spaces such as squares and clubs and also housing projects, both rural and urban. On a worldwide level, Uruguay has been the first country to encourage the implementation of the one-to-one educational model with a national scope. This has placed the experience as a reference for several programs that start every year around the world. New challenges are currently emerging, such as the deployment of a videoconference network with over 1000 points. This network would allow universalizing English teaching on a primary level through distance learning programs, among other planned activities.

Finally, in addition to the presentation of the use case that motivates this the-

2.2. Plan Ceibal's Description

sis, we present a traffic demand analysis with data collected from the wired side of the network. This particular kind of measurement campaigns, over operative networks and under uncontrolled environments, have been called *in the wild* measurements, in contrast to those studies conducted under laboratory conditions, in absence of outside interference and with controlled users behavior. There are several previous works within different contexts, such as companies or conferences, where those which resemble the most to ours are the measurements campaigns taken on university campuses [60–62]. The variety of results and their contradictions within these studies are explained by the constant changing of the wireless traffic over recent years [63]. A couple of innovative aspects of this work stand out from the previous. First, measurements have been obtained from an environment where the one-to-one educational model is implemented, which implies that every child and teenager has a personal computer. Besides, because of the national scale of the study, the amount of data makes conclusions stronger.

Two important results emerge from this first analysis and should be highlighted. On the one hand, we confirm that the traffic demand is growing very rapidly, and we observe that the higher the network capacity, the greater the traffic demand. This fact shows how difficult it will be to cope with the future network requirements. On the other hand, the upward trend in the traffic demand implies that its evolution is not stationary, but if we look at the data at a shorter timescale (e.g. a day or a week) the observed behavior is fairly steady, making it possible to accurately characterize the demand based on statistical data. This result is interesting, as it enables the development of resource allocation algorithms that may learn the demand based on data from the live network. Thus, it is possible to find optimal resource assignments based on the network capabilities, a possibility that will be exploited by the algorithms proposed later in this thesis.

2.2. Plan Ceibal's Description

Plan Ceibal is a governmental project which goal is to promote digital inclusion, aiming to reduce the gap both with other countries and between the citizens of Uruguay, in order to ensure a larger and better access to education and culture. Uruguay kick-started Plan Ceibal with the idea of Nicholas Negroponte [64] as a starting point and through Presidential Decree of April 18, 2007 [65]. This decree stipulated that studies, evaluations and actions necessary to provide each children and teacher from public education a portable computer must be held, that teachers must be trained in the use of the new tools and that the elaboration of educational proposals must be promoted. The initial goal included, besides giving laptops to students and teachers, providing wireless connectivity to the Internet at every public educational center and other open public spaces.

The connectivity requirement of Plan Ceibal set the challenge of providing Internet access to every educational center in the country (over 2000 sites between primary and middle schools, plus technical high schools among others). Open spaces and social centers like squares and clubs were gradually added up. Different

Chapter 2. The Plan Ceibal Use Case Scenario

Internet access technologies such as optical fiber (OF from now on), DSL, EDGE, 3G and satellite were used in the attempt of reaching the connectivity goal, all of them provided by Uruguay's national public ISP (Antel).

2.2.1. Plan Ceibal's Wireless Network

This section contains a brief description of the wireless network deployed by Plan Ceibal (illustrated in Figure 2.1). The list below summarizes all the Internet access technologies used within the network:

- Wired access: DSL (3 Mbps downlink / 512 kbps uplink) and OF (10 Mbps downlink / 2 Mbps uplink).
- Cellular network access: EDGE (256 kbps downlink) and 3G (1 Mbps downlink).
- Satellite (4 Mbps downlink shared by all services).

The wireless access technology used by Ceibal is WiFi (802.11b/g) for all the provided laptops. The solution deployed in educational centers basically consists, on the one hand, of a server in each school with Internet access (through any of the aforementioned technologies). On the other hand, access points, connected to the server *via* Ethernet, are distributed along the site to get a proper area coverage. The design criteria used for the indoor solution states that:

- Every classroom must be covered with WiFi. Besides, the coverage of recreational areas, yards and lunchrooms should be attempted.
- The design is aimed to support a simultaneity factor of 80 % (80 % of the students must have service simultaneously).

The design goal for public spaces coverage was to maximize the coverage area. In this case the equipments were installed over facades and rooftops, again connected to the server through Ethernet. 802.11 radio links at 5.8 GHz were also implemented in order to extend the network by installing special APs over high points, such as telecommunications towers and masts, water tanks or private buildings. This extension made it possible to spread wireless coverage in every town. In addition, these radio-links have proved to be useful to ensure connectivity to those educational centers where no other technology was available in order to provide a good quality of service. There are several examples of the use of 802.11 unlicensed bands for the deployment of access networks in rural zones of developing countries [66,67].

The group of sites to which Ceibal brings services can be divided into two main categories: indoor service, mostly deployed in educational centers, and outdoor coverage, concerning different type of sites. A detailed list of them is presented next:

- Educational centers (indoor):

2.2. Plan Ceibal's Description

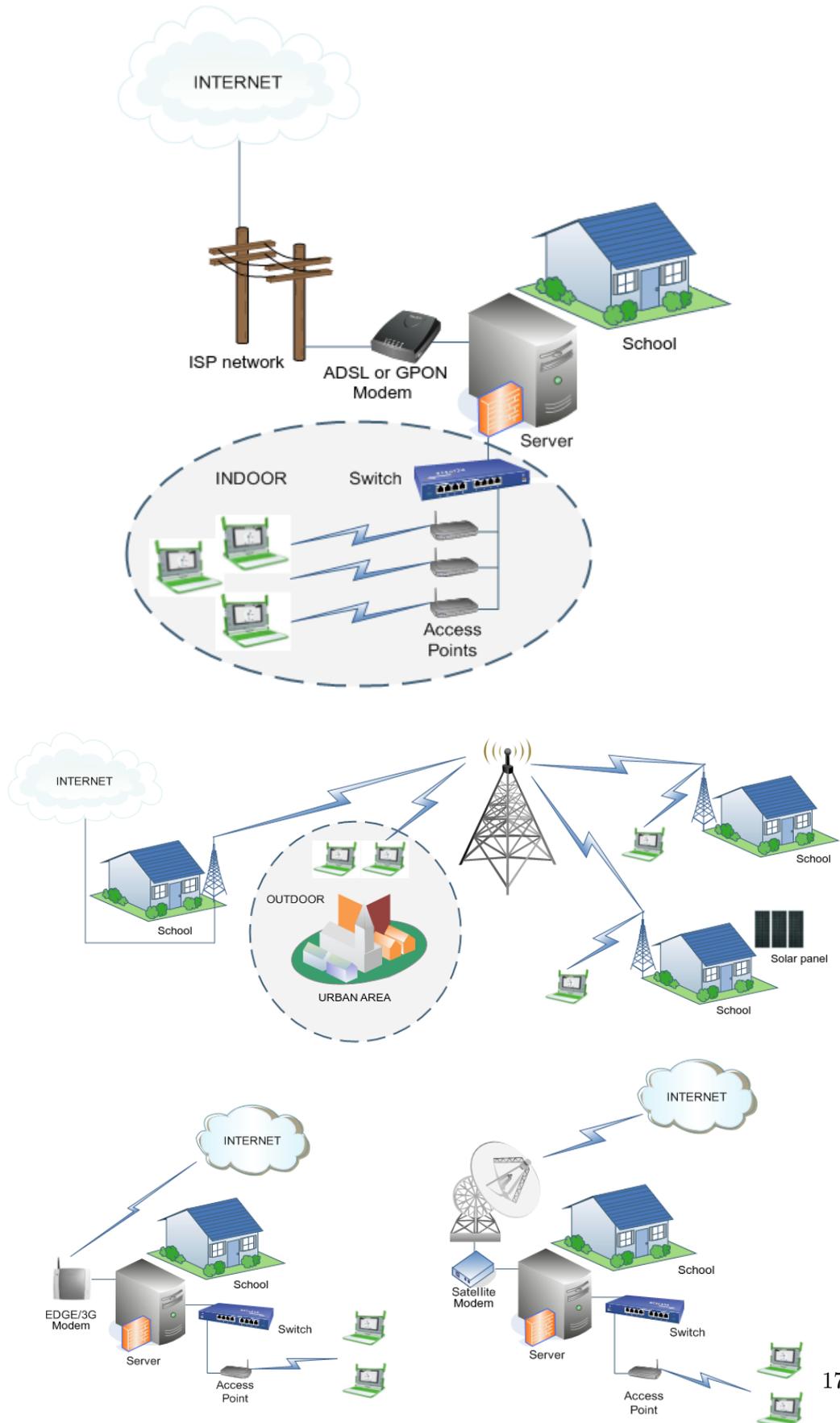


Figure 2.1: Plan Ceibal wireless network infrastructure.

Chapter 2. The Plan Ceibal Use Case Scenario

- Primary schools.
 - Middle schools.
 - Technical high schools.
 - Teaching training centers.
 - Centers for disabled persons.
 - Other educational centers.
- Outdoor coverage:
- High priority neighborhoods (HPN).
 - Squares and sports and social clubs.
 - Urban housing developments.
 - Rural housing developments.
 - High points (telco towers and water tanks).
 - Other sites (communitary classrooms, companies, detention facilities).

It is also important to mention that there were differences in the connectivity strategy for public areas between the capital city (Montevideo), and the rest of the country. The design criteria used for the deployment along the country was to cover those urban areas which counted with an educational center with more than 100 students, aiming to avoid coverage holes with a diameter larger than a 300 meters. Being Montevideo the most populous city in the country, and the one where Internet has got into more homes, the strategy there was based on *hotspots*. The goal was to cover key points such as squares and clubs, instead of having a large coverage area, which implied higher deployment and maintenance costs.

2.3. Network Measurements

The previously described connectivity solution counts with more than 3500 servers and 10.000 access points. Network management involves a permanent monitoring system to register if the service is active or not in every point. This system also allows to register indicators about the user activity at each site. These measurements are fundamental when it comes to optimizing the network because they contain information about the performance of the solution. They are also useful to get to know the needs of the different users and this way provide the necessary resources to every specific case.

As it has been mentioned in the introduction, this kind of data collected from real world operating networks is known as *in the wild* measurements [68]. In this particular case they are gathered from the wired side of the network. This means that the equipments are periodically queried *via* SNMP and data is obtained regarding the number of clients and the carried traffic. This data is then aggregated to different scales either temporary, to identify average behaviors through a day or a week, or spatial, to characterize usage within a town or a region.

2.4. Statistical Data Analysis

Data is also analyzed according to other characteristics such as the kind of access technology of the site, or if it belongs to a primary school, or to a middle or high school. This classification is also used for sites that are not educational centers like squares, clubs and rural or urban housing projects. This way it is possible to analyze where a new Internet connection would impact most in the community, and so, establish deployment policies in order to maximize the benefits users would get from the available resources.

The measures considered are: downlink traffic rate (from the Internet to the access point), what we called demand, and the amount of connected users (called clients). These metrics are then aggregated in different ways, either considering averages or maximums depending on the case. The data was collected during five months, starting in July 2013 (including the winter break) and ending in November 2013, the month prior to the end of the school year. In addition, a temporary filter to keep only school hours was introduced, for which only those measurements taken from Monday to Friday between 8 am and 5 pm are considered.

For the two measures considered (demand and number of clients) we worked with a granularity of one hour. This implies that the minimum timescale available is one hour and the value considered for each measure is the average during the hour.

2.4. Statistical Data Analysis

In this section the collected data is analyzed in several different ways. First, a long-term evolution of the data is considered, comparing the different Internet access technologies. Then, a mid and short-term analysis is held, particularly focusing on those sites with OF Internet access. An indoor usage comparison has been made considering the different type of schools. Finally, outdoor usage has been looked at, identifying which sites have the most impact in the community, based on the number of clients observed on each deployed access point.

2.4.1. Long and Mid-term Analysis

Concerning only schools, Figure 2.2 shows the evolution of the average demand within school hours. It is differentiated according to the type of Internet access, where it can be noticed an important difference between OF and DSL and the rest. There is a noteworthy transient during the first two weeks of July, as they correspond to the winter break. For the following months there is a remarkable upward trend in the traffic demand, except for special days when there are no classes (official holidays). There are two main factors that could explain this evolution: a general increase observed in the popular use of Internet and the fact that near the end of the year school activities might demand more use. A fact to note is that the maximum peak was on November 20th, the same day when Uruguay qualified to the world cup Brazil 2014, something not rare in such a football fan country!

Now, looking at a shorter time scale, we can see that the data has a stationary evolution. Figure 2.3 shows the average demand and number of clients during

Chapter 2. The Plan Ceibal Use Case Scenario

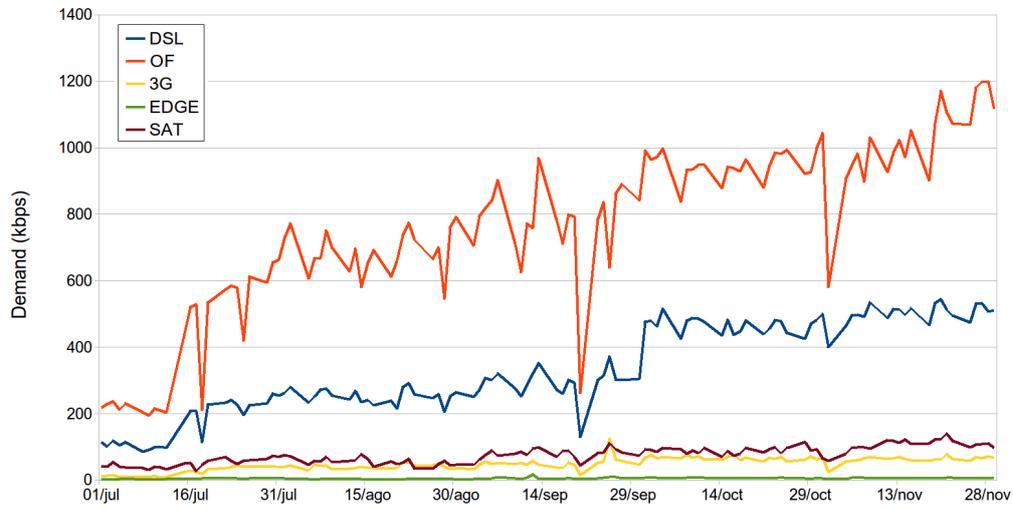
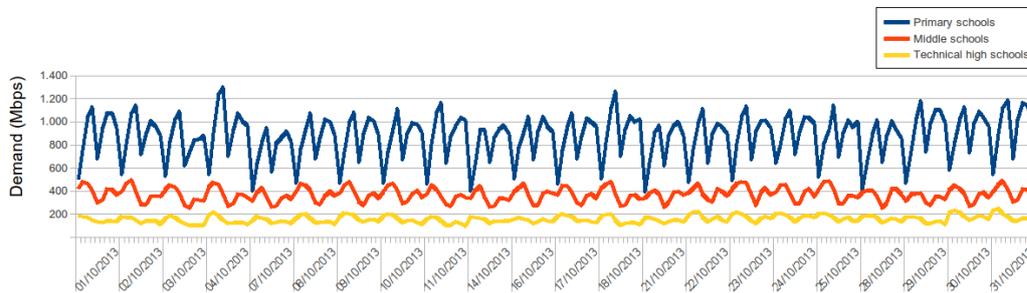
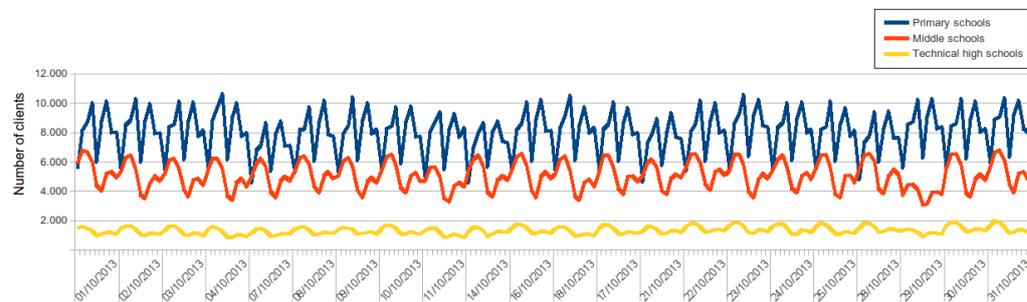


Figure 2.2: Indoor average demand evolution at class time during Jul-Nov 2013.



(a) Demand (Mbps).



(b) Number of clients.

Figure 2.3: Usage evolution for schools during school hours in October 2013.

2.4. Statistical Data Analysis

Internet Access	Number of schools	Demand (kbps) per school	Clients per school	Demand (kbps) per client
OF	862	1.802,7	14,1	127,4
DSL	418	497,0	7,4	67,4
3G	349	57,4	1,0	57,6
Satellite	45	60,4	1,7	34,6
EDGE	328	7,0	0,2	30,7

Table 2.1: Indoor usage at schools between 10 am and 11am (overall peak hour) during October 2013.

October 2013 within school hours. It can be noticed that the daily typical behavior is different depending on the type of institution, something that will be analyzed in detail in the next section.

Table 2.1 summarizes the analysis performed looking at the usage between 10 am and 11am, which is the peak hour. It can firstly be noticed that the largest demand is observed for those sites with OF access, with an average that is more than three times the one observed for sites with DSL access. An interesting insight from the results is that the demand per client is also larger for these sites (twice the one for DSL), which implies that usage is greater where the access technology is better. On the other side, a similar demand is observed in those sites with 3G and satellite access, which is explained because of the context (mostly rural schools with few students). Finally, the poor EDGE connection has the lowest usage, although the context also influences in this case (very few students per school, typically less than 20).

The following part of the study takes into account only the data corresponding to schools with OF Internet access. These schools are the ones which have the greatest part of both the demand and the number of clients, which is clear by looking at Figure 2.4. The results show that at the overall peak hour, sites with OF access take 87% of the demand and 77% of the clients. Similar results were obtained looking at the average during school hours, instead of the peak hour.

2.4.2. Analysis by Type of Institution

Analyzing the data by type of institution, the results are in line with those expected, with middle and high schools with larger demand than primary schools. Table 2.2 shows the maximum usage summary for the schools with OF access during October and November 2013. Both the demand and the number of clients is larger for middle and high schools. Looking at the demand per client, it is unusual to have a smaller one for middle schools than for primary schools, which suggests access saturation problems within these schools. Although primary schools have a lower demand per site, they still have the major part of the overall demand, because of the greater number of sites. In Figure 2.5 it can be seen that they take more than 60% of the demand and almost 57% of the clients, when the maximum usage measured during October and November 2013 is considered.

Chapter 2. The Plan Ceibal Use Case Scenario

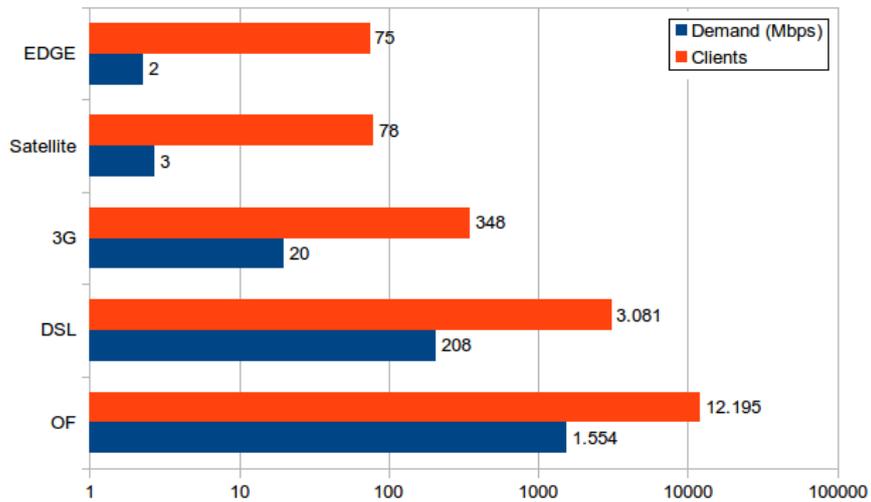


Figure 2.4: Indoor usage distribution at schools between 10 am and 11am (overall peak hour) during October 2013.

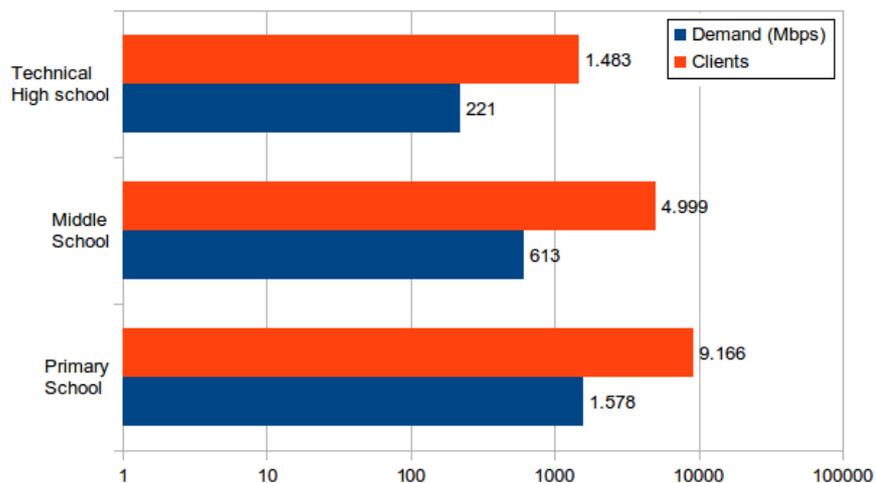
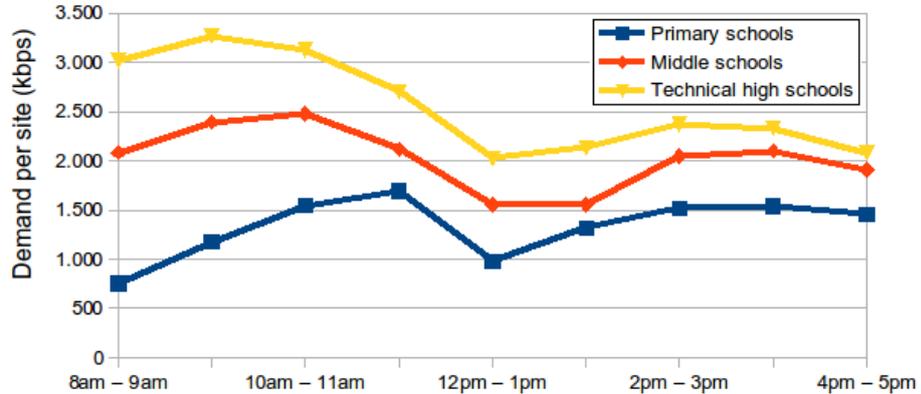


Figure 2.5: Indoor maximum usage distribution for schools with OF access during October and November 2013.

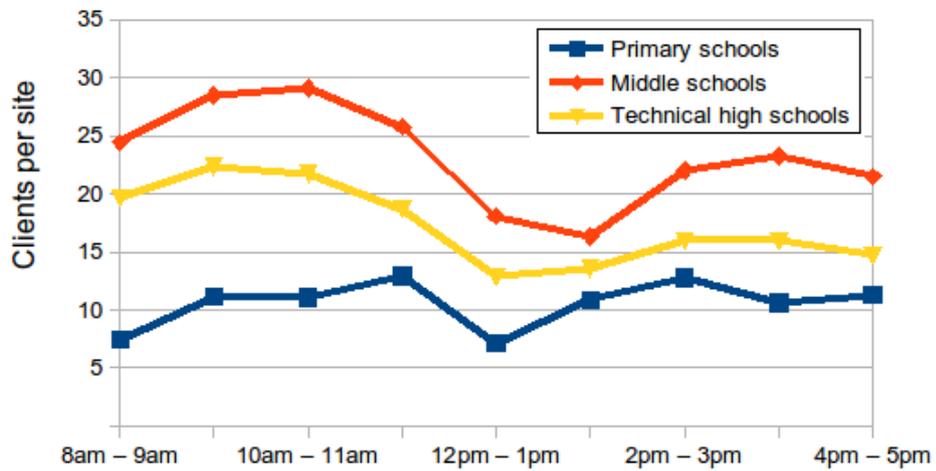
Type of institution	Number of schools	Demand (kbps) per school	Clients per school	Demand (kbps) per client
Primary school	574	2.749	16	172
Middle school	159	3.854	31	123
Tech high school	53	4.179	28	149

Table 2.2: Indoor maximum usage for schools with OF access during October and November 2013.

2.4. Statistical Data Analysis



(a) Average demand per site (kbps).



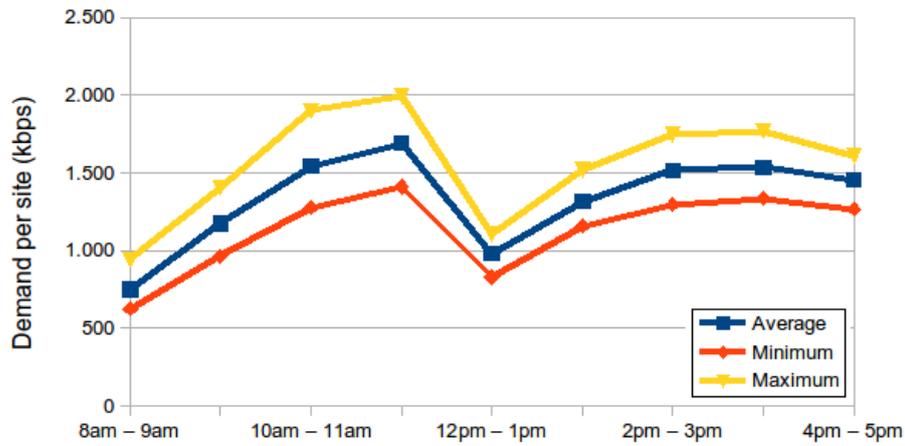
(b) Average number of clients per site.

Figure 2.6: Typical day usage for schools with OF access during class time in October 2013.

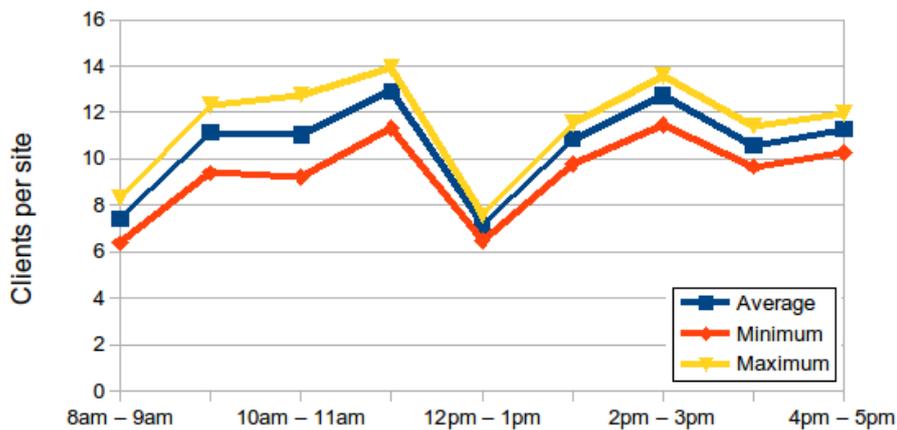
Next, data corresponding to October are again analyzed, but this time with a focus on daily averages, in order to study the typical behavior during a school day. There is a noticeable difference between the dynamics of primary schools with the one observed in middle and tech high schools (see Figures 2.6(a) and 2.6(b)). On one side, the former have two comparable peaks during the day, which are registered in the morning, close to noon, and in the mid-afternoon, respectively. On the other side, the latter, while also has two peaks, the one in the morning is earlier and well above the one in the afternoon.

The observed behavior can be explained by the dynamics of a typical double shift primary school, where the drop at midday is obviously because of lunch time. The usage at primary schools looks more uniform during the afternoon, while in the morning it increases towards noon. Some particular schools start their school hours

Chapter 2. The Plan Ceibal Use Case Scenario



(a) Average demand per site (kbps).



(b) Average number of clients per site.

Figure 2.7: Typical day usage for primary schools with OF access during class time in October 2013.

later (full time or rural schools), which could explain the difference. Concerning middle and high schools, evolution is similar during the morning and afternoon, but usage is higher in the morning. In that case, the number of students per shift could be the reason, as usually the morning shift is the busiest.

Finally, the aforementioned usage stationarity looking at a short scale deserves special attention. In Figures 2.7(a) and 2.7(b) the average daily behavior during school hours for primary schools is shown. While looking at the minimum and maximum curves it can be seen that the observed variation is quite small. In fact, the maximum standard deviation observed in October was 150 kbps (between 10 am and 11 am) and 0.83 for the clients (between 11 am and 12 pm). This is a nice

2.4. Statistical Data Analysis

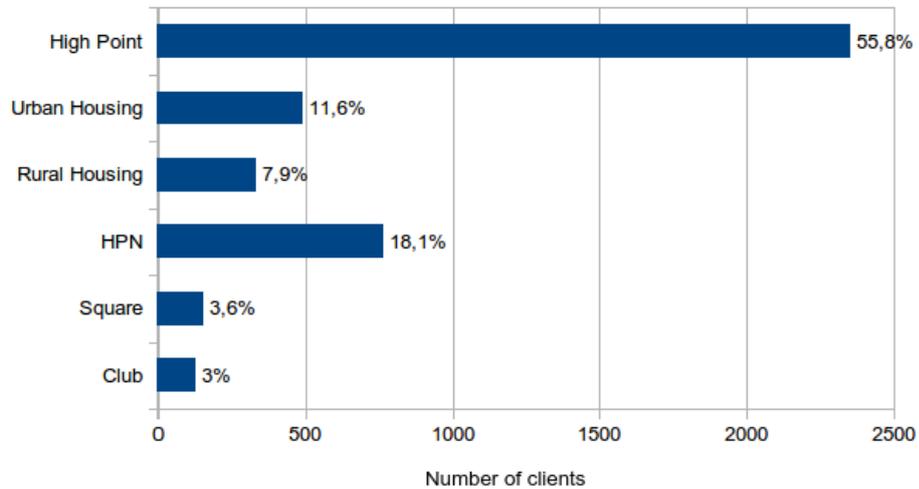


Figure 2.8: Outdoor clients distribution at maximum usage during October and November 2013.

property, which enables the development of an accurate model of the daily usage dynamics.

2.4.3. Outdoor Usage Analysis

To complete the study, data from the outdoor access points were analyzed. In this case only data of the number of clients were considered, as the goal was to measure the impact of the outdoor deployments taking into account how many people actually use it.

Figure 2.8 shows the clients distribution at the maximum usage seen during the months of October and November. The total number of clients is over 4000, which is quite a significant usage, considering that is slightly above than a quarter of the indoor usage. Now, looking at the distribution, as the number of access points installed in high points is the largest one, when we look at the totals, the largest number of clients connects to these access points.

Next, to get a better measure of the impact of each access point, the number of clients per site is analyzed. Table 2.3 summarizes the results. It can be seen that the greater impact is observed in rural housing and high priority neighborhoods. For the former, the lack of good quality Internet connections in rural areas is probably the cause of a higher usage. It is worth noting that not only the students benefits from Plan Ceibal's network, but also their families, what also appears to be an important reason of the high usage in this context. In deprived neighborhoods the situation is different. These are typically sites within urban areas where high quality Internet access is available, but these families may not be able to pay for the service. This is one of the reasons that could explain the observed high use of Plan Ceibal resources. Added to this, it is a fact that these families do not have

Chapter 2. The Plan Ceibal Use Case Scenario

Type of institution	Number of clients	Number of APs	Clients per AP
Club	124,8	95	1,3
Square	151,0	118	1,3
HPN	764,2	154	5,0
Rural Housing	331,3	64	5,2
Urban Housing	489,3	139	3,5
High Point	2352,7	533	4,4

Table 2.3: Average number of clients at maximum outdoor usage during October and November 2013.

access to computers other than the ones provided by Plan Ceibal. Within these neighborhoods, many activities are carried out working with the community, which promotes the appropriation and use of the available resources.

High points also present quite high usage, mostly because they are installed in high population density areas with a wide area coverage. On the other side are squares and clubs, where the least used access points are located. In these areas, a larger number of Internet access in households is probably the reason of a lower usage. In between are the urban housing developments with a middle usage.

Figure 2.9 shows the average daily evolution for outdoor access points. The typical observed dynamics looks complementary to the one observed within the schools. It can be seen that the higher use takes place after the end of school hours, with most of the peak usages after 7 pm (the overall peak usage is between 8 pm and 9 pm). While it is reasonable to have a higher outdoor usage outside school hours, families usage seems to be again another reason to explain this behavior. During school hours there is a lower usage in the morning than in the afternoon. This might indicate that the outdoor usage of morning shift students during the afternoon is higher than the outdoor usage of the afternoon shift students during the morning.

2.5. Final Comments and Reflections

While as marked in the introduction the Plan Ceibal has pioneered the massive deployment of technology and Internet access in education, other projects have continued this line at several places in the world. Examples are the project Conectar Igualdad in Argentina [31] since 2010, and the recent program ConnectED in the United States [32], which was launched in 2013. The latter aims to achieve 99 % of American school students to have access to next-generation broadband by 2018. That means a lot of money will be used to provide Internet connectivity and educational technology into classrooms, and of course, it is going to be wireless.

That is to say, the picture we are taking with this study is just the beginning of the story. It is clear that the number of wireless networks will continue to grow, both to provide connectivity to end-user devices, as well as to deploy access

2.5. Final Comments and Reflections

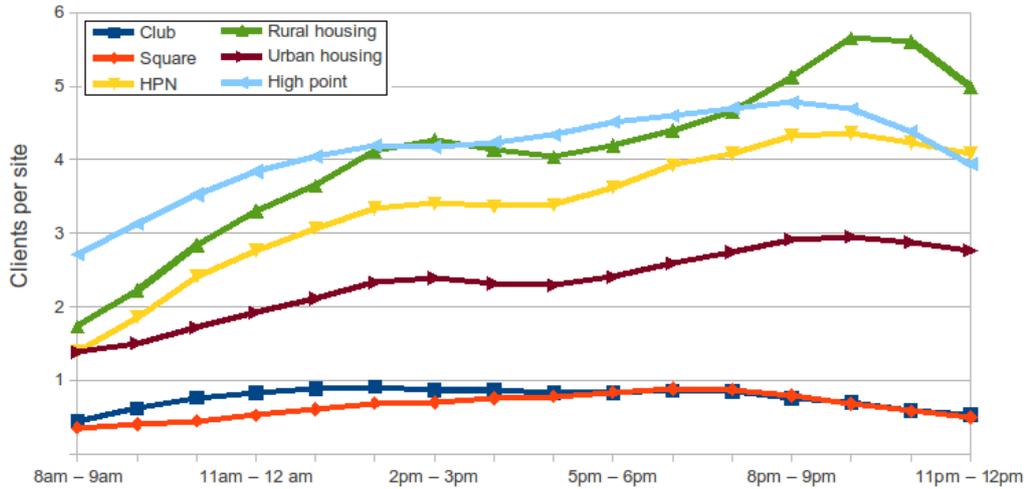


Figure 2.9: Typical day usage for outdoor during October and November 2013.

networks that reach suburban and rural areas. There is also no doubt that the traffic demand will increase as well, something that we already evidenced in this first study. Adaptation and habituation of teachers with technology tools and the growing number of platforms and educational software, will also help this traffic demand to continue increasing. It will be a great challenge for technology developers and providers of wireless solutions to live up to support these requirements.

Finally, another important point that emerges from this study, and that could help to meet the requirements adequately, is the possibility to characterize the traffic demand properly. As we saw, if the timescale is suitably chosen, statistical modeling captures the demand behavior correctly, something which could be used to optimize the network resources dynamically. This idea of optimizing the network with resource allocation mechanisms based on a statistical model, which is learned from actual measurements of the wireless network under operation, is something that will be exploited in all the proposed solutions to the different problems addressed throughout this thesis.

This page was intentionally left blank.

Chapter 3

WiFi Performance Estimation from PHY Layer Measurements

In the previous chapter we validated the assumption that it is possible to estimate the traffic demand from measurements collected from the operating network. Now, we will concentrate in other randomness we typically have in wireless networks, which is the effective capacity achievable for each link. A proper method to estimate this capacity would be a leap ahead in the optimization of wireless communications, as it will enable the possibility to develop optimal resource allocation mechanisms which evolve dynamically, taking into account the variations in both the traffic demand and the effective capacity of wireless links. Solutions of this kind are the ones proposed in Part II and Part III of this thesis.

In this chapter we address the following question: is it possible to infer, given certain physical medium conditions, what is the performance obtained at the upper layers, i.e. network and transport layer. For this purpose, we designed a measurement bench that allows us to accurately control the noise level on an unidirectional WiFi communication link in the RF-protected-environment of an anechoic room. This way, we generated different medium conditions and collected several measurements for various physical layer (PHY) parameters on that link. Then, using the collected data, we estimate the throughput performance from the measured PHY parameters by means of different machine learning methods. In particular, our work concentrates on the performance analysis of a 802.11 link.

3.1. Introduction

In recent decades the development of statistical learning has come a long way. On the one hand, much research effort has been lately devoted to the subject. On the other hand, the increase of computing power has enabled efficient implementations of various methods that existed long ago, which today are used for many different applications. Among others we can list medical diagnosis [69], fraud detection [70], as well as problems in telecommunications networks, such as automatic traffic classification [71]. These techniques offer great potential to cha-

Chapter 3. WiFi Performance Estimation from PHY Layer Measurements

racterize systems with complex variability such as wireless networks, where the physical medium is the cause of most of the errors and performance drops.

Because of the randomness of the wireless shared medium conditions, monitoring wireless networks is very difficult, and the performance analysis is then more complex than in the wired case. Monitoring such networks at the IP layer is only useful for traffic demand characterization, as presented in the previous chapter, but it is not suitable for performance estimation. Some previous work tried to include MAC level information in the monitoring of wireless networks [72], but none integrates the full monitoring of the network from physical to network layers. We nevertheless argue that this is the direction to follow, and our proposal is to estimate the relations between the physical signal parameters and the performance at the network level. Physicists are doing very strong studies on the signal level, but do not study the impact on upper layers [73]. In the work presented in this chapter, it is intended to bridge the gap between the physical signal characteristics and the resulting performance in the digital world of wireless communication networks.

In order to answer the question posed, first we designed and built a platform for benchmarking wireless communications. Many wireless testbeds identified in the literature already exist for that purpose. They typically consist of a large grid of wireless nodes which can be programmed individually to transmit, receive and/or measure data. Custom topologies can be made out of the grid by switching on and off nodes. For example, Orbits [74] follows this approach. However, these platforms are built in open environments and lack the isolation and environmental control required to conduct an accurate cross-layer study on wireless networks. Contrary to these works, our testbed is built in an anechoic chamber to fully control the experimental environment, and avoid external signals to disturb the behavior of the communicating devices and the quality of the measurements. We used on this platform the common digital communications devices that are widely used (laptops, tablets, smartphones), as well as dedicated signal measurement tools specifically designed for physicists.

Our study is focused in the analysis of the relations between the PHY parameters of a WiFi connection, and the performance parameters on top of the IP layer. It aims at demonstrating that, at the opposite of wired networks, the monitoring of wireless networks cannot avoid monitoring the physical level. It is shown that using a very limited number of signal parameters (one or two), it is possible to very accurately estimate the communication performance (considering the network level throughput). With a carefully selected and set machine learning (ML) algorithm, it is even possible to predict performance drops at the scale of one second. For this purpose, we rely on two different kinds of supervised machine learning algorithms, namely Support Vector Regression (SVR) [75] and k -Nearest Neighbors (k -NN) [76]. Both of them are known to have good prediction capabilities and to succeed in many domains, as long as these domains can provide accurate time series [77]. However, their operational characteristics are very different, making them more prone to different usage and applications. For example, SVR algorithms are strong learners whereas k -NN's learning is weak, thus making SVR unable to assimilate training data on the fly, because of the huge compu-

tational complexity. However, SVR algorithms are more sophisticated than k -NN and so are more efficient to generalize data and usually more accurate on the estimations [76]. Therefore, we will compare the relative estimation performances and execution time (learning and estimation delays) corresponding to SVR and k -NN, in order to study the trade-off between complexity and estimation accuracy.

3.2. Machine Learning Algorithms

Two different classes of machine learning algorithms were considered, Support Vector Regression (SVR) and k -Nearest Neighbors (k -NN). In this section we present the basics of both methods. More details can be found in [78] for SVR and in [79] for k -NN.

3.2.1. Support Vector Regression

Given a set of training data $\{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{X} \times \mathbb{R}$ with \mathbb{X} the input space, the purpose of SVR algorithm is to estimate a function $f(x)$ with the requirements of having at most ϵ deviations from the targets y_i . Equations (3.1) and (3.2) show respectively SVR approximation for linear and non-linear form, with $\langle \cdot, \cdot \rangle$ the notation for the dot product in \mathbb{X} . In the linear case, SVR performs a linear regression in the input space. In the non-linear case, no regression can be done in the input space. Therefore, on the one hand, the SVR algorithm has to map the data into some feature space \mathbb{F} via the function $\phi : \mathbb{X} \rightarrow \mathbb{F}$. On the other hand, the classical SV regression algorithm is applied in the new feature space.

$$f(x, w) = \langle w, x \rangle + b \text{ with } w \in \mathbb{X} \text{ and } b \in \mathbb{R}. \quad (3.1)$$

$$f(x, w) = \langle w, \phi(x) \rangle + b \text{ with } w \in \mathbb{X} \text{ and } b \in \mathbb{R}. \quad (3.2)$$

The second requirement for the regression is to maximize the *flatness* of the weights, here measured by $\|w\|^2$. Hence, in the non-linear case both coefficients w and b are estimated by minimizing the regularized risk function given in (3.3). In this equation, C is a user-defined constant which controls the trade-off between the training error and the model flatness. L_ϵ is the ϵ -insensitive loss function defined by equation (3.4). This function allows the SVR algorithm to only penalize estimation errors greater than ϵ .

$$R(f, C) = C \sum_{i=1}^n L_\epsilon(y_i, f(x(i), w)) + \frac{1}{2} \|w\|^2. \quad (3.3)$$

$$L_\epsilon(y_i, f(x(i), w)) = \begin{cases} |y_i - f(x(i), w)| - \epsilon & \text{if } |y_i - f(x(i), w)| \geq \epsilon. \\ 0 & \text{otherwise.} \end{cases} \quad (3.4)$$

To complete the regression we need to solve a convex optimization problem, which is more easily done by maximizing its dual form and introducing the Lagrange multipliers (α_i, α_j^*) . The resulting optimization problem is:

$$\begin{aligned}
 \text{maximize} \quad & -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \phi(x_i), \phi(x_j) \rangle \\
 & - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y(i)(\alpha_i - \alpha_i^*), \\
 \text{subject to} \quad & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\
 & \alpha_i^* \in [0, C].
 \end{aligned} \tag{3.5}$$

Solving this leads to a new definition of (3.2) as

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b.$$

At this point, this definition shows that the solution can be found by only knowing $\langle \phi(x_i), \phi(x) \rangle$ instead of explicitly knowing ϕ . A function $k(x, x')$ which corresponds to a dot product in some feature space \mathbb{F} as defined by $k(x, x') = \langle \phi(x), \phi(x') \rangle$ is called a kernel. This kernel function can be any symmetric function satisfying Mercer's condition¹ such as the Gaussian Radial Basis (RBF) which is defined by $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$. The Gaussian kernel is parametrized by γ ($\gamma > 0$) which impacts the generalization capability of the regressor among other things.

3.2.2. k -Nearest Neighbors

The learning approach of k -NN is to memorize the entire training set. As so, the algorithm belongs to the class of the so-called lazy learners as [80,81] for instance. Given a set of training data $D = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{X} \times \mathbb{R}$, with $\mathbb{X} \subseteq \mathbb{R}$, the process followed by k -NN to estimate an object $z = (x', y')$ can be easily summed-up in three steps. Firstly, the algorithm computes the distance $d(x', x)$ between z and every object $(x_i, y_i) \in D$. Secondly, the set F of the k closest neighbors to z is selected. Thirdly, k -NN computes the estimation as $y' = \frac{1}{k} \sum_{i=1}^k y_i$ with $(x_i, y_i) \in F$. Variants exist and concern essentially the method used to compute the distance $d(x, x')$ such as the Manhattan, Euclidean or Minkowski distance. The p -order Minkowski distance for two points $x^1 = (x_1^1, \dots, x_n^1)$ and $x^2 = (x_1^2, \dots, x_n^2) \in \mathbb{R}^n$ is defined by $(\sum_{i=1}^n |x_i^1 - x_i^2|)^{\frac{1}{p}}$.

3.3. Experimental Platform and Dataset

The implementation of a dedicated wireless testbed is a major requirement for our work. First of all, experimentations must be reproducible, allowing comparison

¹A real-valued function $K(x, y)$ is said to fulfill Mercer's condition if for all square integrable functions $g(x)$ one has $\iint K(x, y)g(x)g(y)dxdy \geq 0$.

3.3. Experimental Platform and Dataset

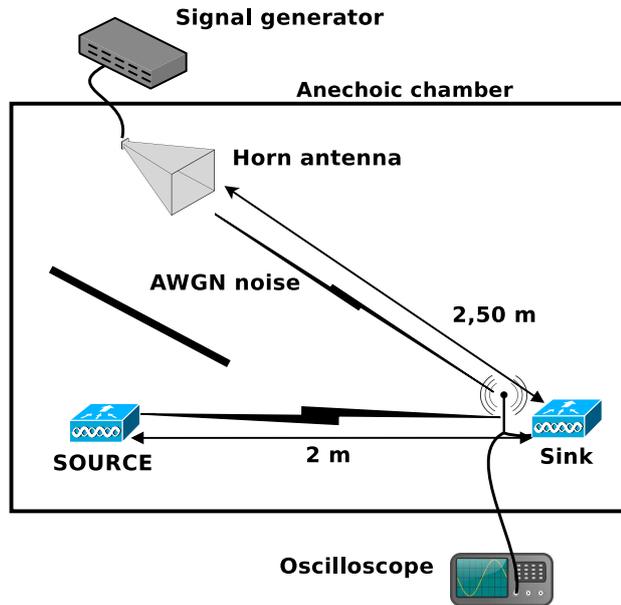


Figure 3.1: Disposition of the different equipments in the wireless testbed. The cable from the receiver antenna is connected to a power splitter which enables the signal from the antenna to be dispensed similarly to the WiFi sink and to the Lecroy oscilloscope with negligible signal alteration.

between different sets of measurements and algorithms. This point is not trivial when using wireless networks as the environment factors have a high impact on the network performances. Secondly, part of the originality of this work comes from the combination of measurements made at multiple network layers, using electronics instruments and software tools. This was also a strong requirement to be able to monitor the physical layer (the wireless transmission), and compare it to the higher layers, from the MAC layer information given by the network cards to the end-to-end layers, as transport throughput for instance. The hardware introspection requirement has an impact on the components choice as explained below. Thirdly, the synchronization of all of these datasets was a sticky point, but absolutely required to ensure a good behavior of the learning algorithms.

3.3.1. Wireless Testbed Description

Our wireless testbed was designed inside an anechoic room (a diagram is shown in Figure 3.1). An anechoic room is a protected RF room which simulates free space conditions. Our chamber model is 4,10 meters long for 2,50 meters wide. Inside, walls are covered of microwave absorbers materials that break and scatter any wireless signal that would come from an inside source. The chamber is then free of any multipath propagation. There are different types of absorbers, each of them defined for a specific frequency range that allows us to use the anechoic chamber for different purposes and frequencies. The absorbers protect also the inner environment of the room from outside perturbations. This protected context

Chapter 3. WiFi Performance Estimation from PHY Layer Measurements

minimizes the uncontrolled parameters of our communication.

Inside the anechoic chamber we placed two WiFi nodes. The nodes are controlled through a wired network to avoid interference with the wireless communication. The nodes are Avila-GW2348-4 gateway platforms [82] and run a Linux OpenWrt OS [83]. The boxes have an Intel Xscale processor, 64 MB of SDRAM and 16MBytes of Flash memory. The WiFi network controllers are based on the AR5414 chipset from Atheros which uses the `ath5k` driver and are attached to an omnidirectional antenna. The choice of the WiFi chipset and its driver was crucial because they define the amount of metrics and the accuracy that it is possible to obtain.

The `ath5k` driver is open source and well documented thanks to an active online community support [84]. It has also a good integration within the OpenWrt OS. The OpenWrt OS is flexible enough to allow the implementation of new functionalities so that it accelerates the upgrade of the bench. In addition, we used an oscilloscope connected to the receiver antenna to capture the noise level and the received signal strength, by recording the amplitude of the received signal. The oscilloscope chosen was a fast Lecroy WaveRunner which allows us to capture a maximum number of frame signal with little loss and to record them on internal memory. The precision of this instrument gives us the ground truth required by the training methods used. It also embeds a large library of filters, and operators which can be applied on the input signals.

As we used several equipments to get measurements, it is needed to have their clock very accurately synchronized. This was done with NTP by using a dedicated wired connection to a remote NTP server (accuracy with a shared network bus is not sufficient).

3.3.2. Experimental Protocol

The noise and the interferences significantly impact the performance in wireless communications. One of the objectives of our environment is to minimize the presence of uncontrolled parameters during the experiments. On the other hand, another objective is to generate and control selected parameters that will affect the performance. For this purpose, we then inject noise in the environment using a signal generator to perturb the communication. The signal generator is a device which emits RF signals and it can be configured to generate realistic noise.

Among the parameters of the generated noise, two important elements have a crucial impact. First, the modulation used defines the main characteristics of the noise signal in the time and frequency domains (i.e. it determines the spectral occupancy of the generated signal, its fading or narrowness). Second, the amplitude of the signal also affects the measured level of noise on the receiver side. We found that the AWGN (Additive White Gaussian Noise) noise was a good choice for our preliminary studies because of its simplicity. Moreover it can be used to impact the entire bandwidth of a 802.11g channel, contrary to most other modulation schemes which produce narrow band noise. The noise level was determined empirically by testing the effects on the communication.

3.3. Experimental Platform and Dataset

Training set	Dataset definition
Notation	{Tx Power (dBm); Noise Power (dBm)}; {sample 2};...
<i>Dataset1</i> (5323 vectors)	{10;-20};{10;-17};{10;-15};{10;-13};{10;-10}; {10;-7};{10;-5}; {20;-20};{20;-17};{20;-15}; {20;-13};{20;-10};{20;-7};{20;-5}
<i>Dataset2</i> (2661 vectors)	{10;-20};{10;-17};{10;-15};{10;-13};{20;-20}; {20;-17};{20;-15};{20;-13}
<i>Dataset3</i> (1330 vectors)	{10;-20};{10;-17};{10;-15};{10;-7};{10;-5}; {20;-20}

Table 3.1: Constitutions and characteristics of our training sets. Each vector represents 1 second of measurements

Another important element that affects the noise generated in the anechoic chamber is the antenna. It characterizes the waveform, the direction and the amplitude of the noise wave. In order to perturb only one side of the communication we used a very directional antenna pointed to the receiving station. The Lecroy oscilloscope was set to capture and flush the data as soon as a frame is detected on the input cable. This happens when the amplitude of the sensed signal is above a specific threshold, set to be in between the current noise floor and the minimal amplitude value of a frame. This threshold has to be set in a way to prevent exceptional high noise values that could be incorrectly detected as a frame.

The configuration of the network interfaces is done in promiscuous mode to capture any packets sensed by their antenna. The packets are captured at the MAC layer using the PCAP library and tools when they arrive at the kernel interface. The packets contain data from link to application layers, such as the 802.11 channel number, the type of frame at the MAC layer, or packet size at the network layer. Additionally, a packet also contains a RADIOTAP header which gives radio level information such as the received signal strength (RSS) reported by the `ath5k` driver. We modified the `ath5k` drivers of the OpenWrt OS to permit, when possible, the propagation of packets with frame check sequence (FCS) errors to the upper layers, while on the original kernel they were discarded. The propagation is only possible if the error corrupted the data but not the header fields. Following this modification the RADIOTAP header now contains a flag specifying whether a FCS error was detected when decoding the packet. Finally, we use Iperf [85] to generate traffic between the two peers. The traffic is a TCP flow over a 802.11 link, with fixed PHY layer rate of 24 Mbps. The size of the packets is set to 1470 bytes.

We generated different samples with different noise levels and different transmission powers. All the samples have the same duration of 5 minutes and will be used to constitute our training datasets. Table 3.1 sums up the characteristics of the different samples. The same experimental settings (transmission power and noise) are used for training and testing. Therefore a training dataset which contains all these samples will be considered as having full knowledge about the possible use cases met in the test dataset. Hence, to test the generalization ca-

Chapter 3. WiFi Performance Estimation from PHY Layer Measurements

capacity of our algorithm, we built three different training datasets as described in Table 3.1. These datasets differ by the quantities of samples they are made of, and consequently by the level of knowledge they represent.

3.3.3. Measures Defined as Features

Throughput

In this study we are considering the throughput as the performance metric of the communication. It is computed from the PCAP captured at the receiver side of the transmission. It is defined as the computed throughput at second i , given by:

$$BW_i = \sum_{k=1}^n L(p_k) \text{ with } k \in \mathbb{N},$$

where $L(p_k)$ is the length of the payload at the network layer for packet p_k . Thus, it must hold that $p_k \in P_i$, where P_i is the set of received packets without FCS error during second i : $P_i = \{p_1, \dots, p_n\}$.

Received Signal Strength @Atheros

The Atheros RSS is extracted from the RSS field in the radiotap headers of the packets captured, so included in the PCAP files. Given that $RSS(p_k)$ is the RSS of packet p_k such as $p_k \in P_i$, and R_i is the set of RSS extracted from packets captured during second i , it is defined as:

$$ATH_RSS_i = \overline{R_i} \text{ with } R_i = \{RSS(p_1), \dots, RSS(p_n)\}.$$

In addition to the Atheros values, we extract different metrics from the Lecroy collected data. These values are computed from the Root Mean Square (RMS) values of the raw data. These RMS values can be split into three parts, which are the data that are before, during and after the frame. The part of the data before and after the frame corresponds to the noise level values and therefore can be used to extract the noise floor during the reception of that frame. We consider A and C , the sets of these points. Therefore we compute the average noise floor of the data during the reception of frame f with $N_f = \overline{A \cup C}$.

Noise level @Lecroy

With M_i the set of noise levels extracted from the frames captured by the Lecroy oscilloscope during second i , we compute the feature for the noise floor at second i as:

$$LECR_NOISE_i = \overline{M_i} \text{ with } M_i = \{N_{p_1}, \dots, N_{p_n}\} \text{ and } p_k \in P_i.$$

3.4. Throughput Estimation from PHY Measurements

Received Signal Strength @Lecroy

The RSS of the received frame is computed on the first 8 symbols to comply with the IEEE 802.11 standard [1, 41]. These points constitute the set D . Thus, similarly to previous equations, the RSS for a frame f is given by $R_f = \overline{D}$, and then the resulting measure is:

$$LECR_RSS_i = \overline{R}_i \text{ with } R_i = \{R_{p_1}, \dots, R_{p_n}\} \text{ and } p_k \in P_i.$$

Signal to Noise Ratio @Lecroy

Finally we compute the signal to noise ratio (SNR) S_f for frame f as the difference between the noise floor and the RSS of the frame P , using both values corresponding to the Lecroy oscilloscope measurements. Therefore, the SNR is obtained using the previous formulas as $S_f = R_f - N_f$, and then the resulting SNR measure is:

$$LECR_SNR_i = \overline{W}_i \text{ with } W_i = \{S_{p_1}, \dots, S_{p_n}\} \text{ and } p_k \in P_i.$$

3.4. Throughput Estimation from PHY Measurements

3.4.1. Machine Learning-based Methodology

The presented study considers two different machine learning algorithms, SVR and k -NN, to analyze the tradoff between complexity and estimation accuracy. The SVR algorithm has been used with RBF as a kernel function. As pointed out in section 3.2.1, in our configuration SVR requires three user-defined parameters (C , γ and ϵ) which can impact performance, and therefore must be carefully selected with regard to the application. For our estimations, we used a grid search to select these SVR parameters. It is a common empirical method which consists in an exhaustive test run of SVR training using generated settings combinations. We then select the best combination of C , γ and ϵ among the results.

For the performance of k -NN, the value of k must be carefully selected. Therefore, after several tests on the different datasets, we chose a value which allows a good tradeoff between the estimation accuracy and the generalization results. In the presented experimentation we set the value of k to 3. The distance used is Minkowski of order 2, which corresponds to the Euclidean distance, recommended in the traditional version of the algorithm [79].

One part of the analysis of the machine learning estimations concerns the computational time associated with the training and estimations process. Our learning setup uses Python scikit-learn implementation [86] of SVR and k -NN. The delays are computed by reading the current clock using the 'time' function. The clock is read twice: before and after the measured process. The difference of the two measures constitutes the delay for the measured process. For each estimation, we made 100 runs and then computed the average and standard deviation of the delays. The CPU used to conduct the measures is a 64 bits Intel Core 2 Duo

(2x2.53 GHz) with 6 MB of cache memory. The computer disposes of 4 GB of RAM memory. The operating system is Debian Linux.

3.4.2. Estimation Accuracy

To evaluate the accuracy of the estimations, two methods are used. On one side, the well-known Mean Squared Error (MSE) is considered. Given that $\hat{Y}_i, \dots, \hat{Y}_n$ are estimations and Y_i, \dots, Y_n are the real values, the MSE is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2.$$

We also use the percentage of correct estimations noted $P(e < \lambda)$ and defined by:

$$P(e < \lambda) = \frac{1}{n} \sum_{i=1}^n d(\hat{Y}_i, Y_i, \lambda).$$

This value is the percentage of estimations which differ from the corresponding real values by less than a predefined threshold λ , as shown on equation (3.6). These estimations are then considered “correct”. Given the PHY layer rate of 24 Mbps and the size of the packets defined to be 1470 bytes, we set the value of the threshold λ to 1 Mbps. Indeed, this threshold corresponds to an error in the estimation of 4% (89 packets over 2139 transmitted during one second). By considering the preliminary measured performance of the algorithms this value could be considered to be fair to assess the goodness of the algorithms.

$$d(\hat{Y}_i, Y_i, \lambda) = \begin{cases} 1 & \text{if } |\hat{Y}_i - Y_i| < \lambda. \\ 0 & \text{if } |\hat{Y}_i - Y_i| \geq \lambda. \end{cases} \quad (3.6)$$

3.4.3. Experimental Results

Table 3.2(a) contains the results of the throughput estimation based on 6 different PHY or combinations of PHY parameters for *Dataset1*, *Dataset2*, and *Dataset3* respectively. The first column quotes the PHY parameters that have been used for the SVR estimation of the IP throughput. The rest of the table shows the figures obtained for the MSE and the probability $P(e < 1 \text{ Mbps})$ for both machine learning algorithms. Finally, Table 3.2(b) presents the ranking for the PHY parameters according to their ability to estimate the throughput accurately. A ranking of 1 corresponds to the best estimation result among the 6 PHY parameters considered.

For *Dataset1*, i.e. the full one, the best result is obtained with the parameters *LECR_RSS+LECR_NOISE* for both machine learning algorithms. The estimations for SVR are plotted in Figure 3.2. The results exhibit impressive matching between the real and estimated values of the throughput, with just very few outliers appearing (75% matchings considering the 1 Mbps threshold). We also got quite good results for *Dataset2*, and *Dataset3*, but this time, the best results for SVR

3.4. Throughput Estimation from PHY Measurements

Table 3.2: Results of the estimations using physical layer metrics. $D1$, $D2$ and $D3$ stands respectively for *Dataset1*, *Dataset2* and *Dataset3*.

(a) Scores of the estimations.

n°	Physical layer parameter(s)	MSE (Mbps ²)						P($e < 1Mbps$) (%)					
		SVR			k -NN			SVR			k -NN		
		$D1$	$D2$	$D3$	$D1$	$D2$	$D3$	$D1$	$D2$	$D3$	$D1$	$D2$	$D3$
1	<i>ATH_RSS</i>	11.24	11	10.17	23	33	34	35	33	34	24	22	14
2	<i>LECR_RSS</i>	4.42	3.9	4.5	27	7.1	10	51	59	32	18	35	31
3	<i>LECR_NOISE</i>	2.28	5.4	5.8	5	2.8	4.2	69	55	44	50	44	24
4	<i>LECR_SNR</i>	1.69	1.6	1.6	4	2.3	2.8	64	66	62	48	50	45
5	<i>ATH_RSS + LECR_NOISE</i>	1.02	2.3	3.3	4	1.3	1.7	70	49	41	54	60	50
6	<i>LECR_RSS + LECR_NOISE</i>	0.88	2.0	2.53	2	1.2	2.2	75	57	49	64	63	46

(b) Pertinence of the estimations.

n°	Physical layer parameter(s)	SVR Pertinence ranking						k -NN Pertinence ranking					
		MSE			P($e < 1Mbps$)			MSE			P($e < 1Mbps$)		
		$D1$	$D2$	$D3$	$D1$	$D2$	$D3$	$D1$	$D2$	$D3$	$D1$	$D2$	$D3$
1	<i>ATH_RSS</i>	6	6	6	6	6	5	6	6	6	5	6	6
2	<i>LECR_RSS</i>	5	4	4	5	2	6	5	5	5	6	5	4
3	<i>LECR_NOISE</i>	4	5	5	3	4	3	4	4	4	3	4	5
4	<i>LECR_SNR</i>	3	1	1	4	1	1	2	3	3	4	3	3
5	<i>ATH_RSS + LECR_NOISE</i>	2	3	3	2	5	4	2	2	1	2	2	1
6	<i>LECR_RSS + LECR_NOISE</i>	1	2	2	1	3	2	1	1	2	1	1	2

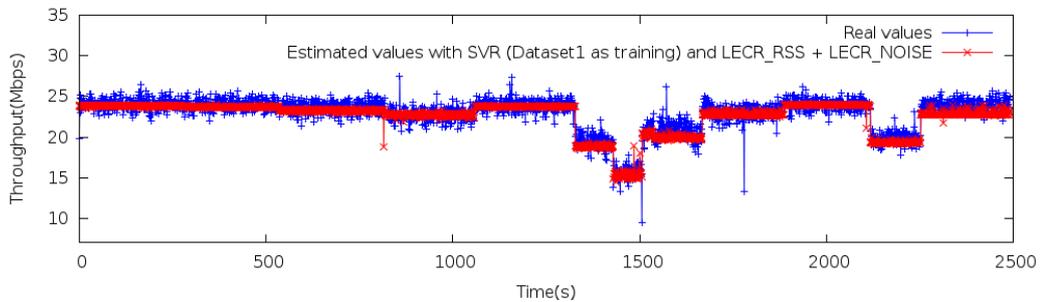


Figure 3.2: Throughput estimation results obtained with the *LECR_RSS + LECR_NOISE* metric compared to the real throughput.

were obtained with the *LECR_SNR* parameter (60% matchings). The difference of the results when using a full trace for the training compared to a sampled one exhibits the non empty intersection between PHY parameters as SNR, RSS and NOISE. The results for k -NN improve with the use of *Dataset2*. Contrary to SVR, the best estimations are obtained with the features 5 and 6 for every training dataset. Generally speaking, SVR performs better than k -NN excepts for *Dataset2* where k -NN outperforms SVR in terms of MSE.

Table 3.3 presents the results of the measured delays for training and estimations using SVR and k -NN. According to these numbers, the time taken by SVR

Chapter 3. WiFi Performance Estimation from PHY Layer Measurements

to train can be very high. Hence, with *Dataset1* and the RSS metrics, the delays goes up to the tens of seconds. Then the time decreases with the use of smaller training sets. In the case of k -NN, no model are computed, the data are simply memorized. Therefore the training is very fast and essentially depends on the size of the training sets. As a consequence, k -NN values decrease geometrically by a factor of 2 when changing from *Dataset1* to *Dataset2* and then from *Dataset2* to *Dataset3*. According to section 3.2.1, SVR forces the estimated function to be within a distance ϵ from the averaged data, a requirement that can be tough for the algorithm to fulfill. Hence, the high time consumption for the SVR model training is due largely to the chosen value of the parameter ϵ , which affects greatly the training accuracy as well as the computational delay. The time delays consumed for the estimation are higher when using SVR, than when using k -NN, which is a point in favor of the latter in the comparison. By observing the global results, we see that k -NN can largely compete with SVR when it comes to accuracy while at the same time being slightly faster.

Table 3.3: Results of the measured delays for training and estimations using physical layer metrics. $D1$, $D2$ and $D3$ stands respectively for *Dataset1*, *Dataset2* and *Dataset3*.

(a) Average delays observed for the training processes on 100 runs (values into brackets are the standard deviation of the distributions. Due to space limitation, standard deviation values are given in 10^3 unit).

Physical layer parameter(s)		Time used for training (s)					
		SVR			k -NN		
		$D1$	$D2$	$D3$	$D1$	$D2$	$D3$
1	<i>ATH_RSS</i>	5.39 (40)	1.40 (2)	0.36 (0.4)	0.048 (2)	0.023 (0.1)	0.012 (0.1)
2	<i>LECR_RSS</i>	41.27 (300)	11.71 (7)	3.12 (2)	0.048 (1)	0.023 (0.2)	0.012 (0.1)
3	<i>LECR_NOISE</i>	5.17 (10)	1.38 (1)	0.36 (0.8)	0.051 (6)	0.023 (0.2)	0.012 (0.1)
4	<i>LECR_SNR</i>	11.54 (6)	3.87 (4)	1.35 (2)	0.048 (4)	0.023 (0.2)	0.012 (0.1)
5	<i>ATH_RSS + LECR_NOISE</i>	4.50 (4)	1.15 (2)	0.30 (0.2)	0.048 (0.6)	0.023 (0.1)	0.012 (0.1)
6	<i>LECR_RSS + LECR_NOISE</i>	4.72 (9)	1.23 (2)	0.31 (3)	0.046 (0.5)	0.023 (0.1)	0.012 (0.08)

(b) Average delays observed for the estimations processes on 100 runs (values into brackets are the standard deviation of the distributions. Due to space limitation, standard deviation values are given in 10^3 unit).

Physical layer parameter(s)		Time used for estimation (s)					
		SVR			k -NN		
		$D1$	$D2$	$D3$	$D1$	$D2$	$D3$
1	<i>ATH_RSS</i>	1.52 (10)	0.78 (6)	0.40 (3)	1.13 (10)	0.63 (1)	0.39 (0.6)
2	<i>LECR_RSS</i>	1.58 (20)	0.81 (10)	0.42 (3)	0.61 (3)	0.44 (0.8)	0.29 (0.6)
3	<i>LECR_NOISE</i>	1.47 (20)	0.76 (10)	0.42 (4)	0.96 (30)	0.35 (0.9)	0.08 (0.3)
4	<i>LECR_SNR</i>	1.38 (30)	0.70 (10)	0.36 (3)	0.74 (60)	0.45 (0.8)	0.19 (0.3)
5	<i>ATH_RSS + LECR_NOISE</i>	1.34 (9)	0.67 (10)	0.35 (8)	0.32 (10)	0.15 (0.3)	0.06 (0.1)
6	<i>LECR_RSS + LECR_NOISE</i>	1.38 (1)	0.70 (4)	0.36 (3)	0.27 (0.4)	0.15 (0.2)	0.08 (0.1)

3.4. Throughput Estimation from PHY Measurements

3.4.4. Final Remarks

In the analysis presented in this chapter, we focus on estimating and predicting the performance of a wireless network considering performance metrics at the network and transport layer level. For this purpose we collected data from the physical layer, which were the inputs of the machine learning algorithms considered (SVR and k -NN). The wireless testbed was set in the RF-protected-environment of an anechoic chamber, allowing us to control the air conditions during the experiments.

From the results obtained in the experiments, it appears that is perfectly possible to estimate and predict (on a one second scale) the performance of the wireless network at layers 3 and 4, using information collected from layer 1 measurements. Thus, we can answer the question posed, confirming that it is in fact possible to infer the performance at the upper layers, from measurements of the physical medium conditions. Nevertheless, a deeper analysis on larger datasets would allow a more accurate characterization of the link between PHY parameters and the network performance at higher layers.

As mentioned in the introduction the machine learning tools offer great potential for various engineering applications, in particular for telecommunications networks. Wireless networks present particular problems that can be addressed with learning techniques, due to the natural randomness that has the propagation of radio signals in the air. The final goal would be to build a generic platform for monitoring and analyzing wireless networks. This way, it would be possible to have real time information of the effective capacity at each link of the network, estimated from measurements of the physical conditions at each link of the operating network.

This page was intentionally left blank.

Conclusions of Part I

The first part of this thesis is clearly divided into two different chapters. One aspect that appears in both chapters and will be common thread throughout this thesis, is the incorporation of statistical analysis tools and machine learning to the analysis and design of wireless networks. In this case we saw, on one hand, the possibility to characterize the traffic in different schools in order to develop a predictive model of the demand, presented in Chapter 2. On the other hand, in Chapter 3, we introduced a way for estimating the effective capacity (at the upper layers of the network) from measurements made at the physical layer. In the next part of this thesis we will see how the use of measurements from a live operating network and statistical learning tools can be useful for the development of optimal algorithms for traffic engineering.

In particular, Chapter 2 focuses on the main motivation of this thesis, which is the national wireless infrastructure from Plan Ceibal. We also presented in this chapter, a basic characterization of the traffic demand of this wireless network, which supports the one-to-one educational model at a nationwide scale. Typical behaviors were identified within educational centers depending on their type. Unsurprisingly, a larger demand was observed for middle and higher education in every case. The demand correlation with the access technology has also been analyzed, where it has been observed an increasing demand when the access quality is better. This fact, along with the growing use of the Internet in Uruguay [87], and also adding the new resources and activities available for the users of Plan Ceibal (such as virtual library and educational platforms) suggest that demand will keep growing through the next years.

For those sites outside educational centers, the usage level is diverse, with deprived neighborhoods of urban areas and housing developments in rural areas registering the higher values. There is also a notorious difference in the traffic evolution during the day in contrast to what is observed in educational centers. This information is useful in order to define policies that contribute to an increase in the use of the available resources in those sites with low activity levels.

For future steps in this type of analysis, the first point would be to perform a deeper analysis of the collected data, increasing the number of different measurement classifications or adding new subcategories to the ones already covered. As an example, it would be of interest to discriminate according to the schedule of the different schools or the context where they are immerse. Another interesting topic to study would be a user behavior analysis on an application level. Different applications such as the virtual library and educational platforms have already been

incorporated to Plan Ceibal. It is of high importance to know which applications are the most or least used ones. This type of data enables a deeper network usage analysis in order to be able not only to determine how much resources are used but also for which applications. This would reveal the kind of infrastructure necessary, regarding both capacity and quality of service. Finally, it would be interesting to incorporate data registered from the wireless side to the measurement campaigns. Air measurements are harder to obtain and it would be necessary to incorporate specific equipment. However, the information would be extremely useful in order to characterize and optimize the performance of the deployed infrastructure.

In the second half of Part I of this thesis, Chapter 3 introduces us to the basics of performance analysis of wireless networks based on the IEEE 802.11 standard. The main contribution presented deals with the design of a generic platform for monitoring and analyzing wireless networks. This wireless testbed is set in the RF protected environment of an anechoic room, allowing us to control the perturbation on the physical medium by generating noise. It also stands as a novelty the integration of pure physical signal measurement equipment, as the Lecroy oscilloscope, for very accurate measurements serving as ground truth. Based on the collected data, the second contribution of this analysis deals with exhibiting the importance of PHY parameters on network communication performance. The correlation between the physical environment and the communication performance is so strong that it is possible, by only monitoring the SNR and the RSS of the signal, to predict the performance level at the higher TCP/IP level. This result has been demonstrated using different kinds of models, in particular the SVR and k -NN models presented in this chapter.

The platform developed enables the realization of a much larger exploitation and more complex future studies. Indeed, for this preliminary stage, we just set simple scenarios with a single connection and simple noise model that can appear a bit far from realistic situations. These first simplistic scenarios were mandatory to validate the platform accuracy, and the monitoring and analysis tools, as well as for gaining the required skills required for this multi-thematic work, especially in the domain of the signal propagation and behavior. With this platform it is possible to generate large datasets with more complex and realistic scenarios, and for different kinds of wireless networks, not only 802.11-based, but also UMTS, LTE or any new ones to appear in the future. A further analysis on this new generated datasets will enable to understand how wireless networks behave, and then try to improve the way we use and manage them.

Part II

Dynamic Load-Balancing in Wireless Networks

Chapter 4

Routing and Forwarding in Wireless Mesh Networks

In the second part of this thesis we address the problem of routing and forwarding in Wireless Mesh Networks (WMNs). We consider a particular but very typical scenario: a planned WMN where all links do not interfere with each other. For example, this is the case of the deployment from Plan Ceibal introduced in Chapter 2, where WMNs are used to provide Internet access in suburban and rural schools. We also assume, based on the results from the same chapter, that it is possible to estimate the traffic demand at a certain time scale.

In the context of WMNs resources are intrinsically scarce, which has led to the proposal of dynamic routing in order to fully exploit the network capacity. We argue instead in favour of separating routing from forwarding (i.e. *à la* MPLS). Our proposal is a dynamic load-balancing scheme that forwards incoming packets along several pre-established paths in order to minimize a certain congestion function. We use a simple and versatile congestion function: the sum of the average queue length over all network nodes interfaces. We present a method to learn this function from measurements and several simulations to illustrate the framework, comparing our proposal with the IEEE 802.11s standard.

The rest of Part II of this thesis is structured as follows. After an introductory section, in Section 4.2 we describe some previous work and highlight some recent papers. In Section 4.3 we introduce the network model and most of the notation used. The chapter continues in Section 4.4 where we describe the procedure for learning the congestion function model from measurements. Then, the next chapter details the proposed method and discuss its implementation in a real network in Section 5.2 and 5.3. Finally, in Section 5.4 we present the simulation experiments and performance comparison, while conclusions are discussed in a separate chapter at the end of this part.

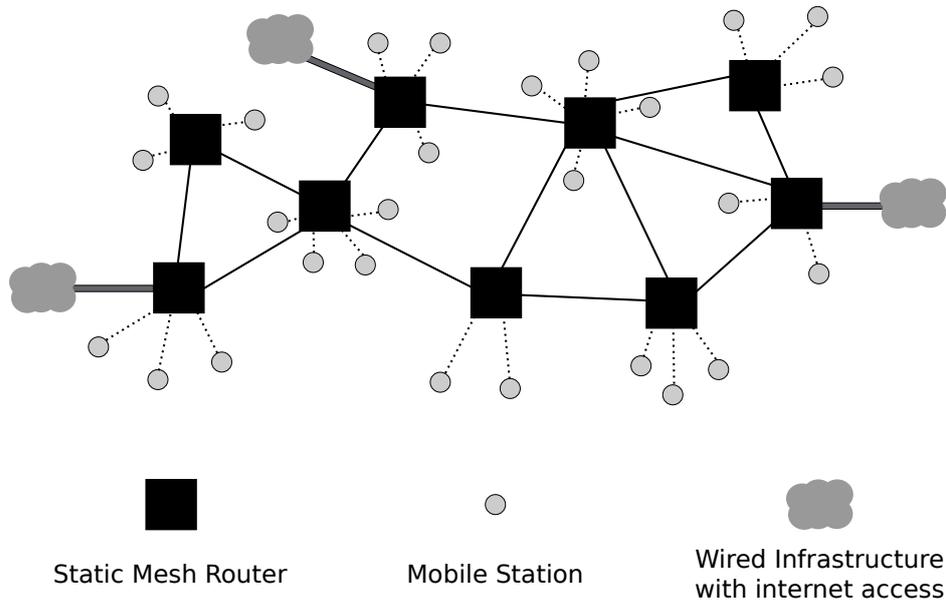


Figure 4.1: Wireless mesh network typical architecture.

4.1. Introduction

Wireless mesh networks (WMNs) [2, 40] are no longer just a promise for the future but a reality today, thanks mainly to the advantage offered in terms of cost compared to traditional wired access networks. In particular, outdoor community mesh networks [54] and rural deployments [66, 67] based on IEEE 802.11 have seen tremendous growth in the recent past. Lately, even service providers are beginning to use this technology, resulting in an increasing presence of carrier-class equipment in the market [57]. The typical architecture of a WMN (see Figure 4.1) includes one or more Internet gateways and several relay routers. Clearly, these intermediate routers increase the coverage of the access network without requiring more, and probably expensive, connections to the Internet. However, several problems arise that are specific of this kind of architectures.

The main challenge for this kind of networks, at the wireless mesh backbone level, is routing and forwarding. In the current IEEE 802.11s standard [14] (and in several other proposals [15]) each link has an associated metric value as cost. This cost is expected to change over time, and reflect current conditions (propagation conditions, interference, etc.), so as to maximize a certain criteria (e.g. throughput). To choose a path to its destination, each router executes a shortest path algorithm. This procedure is essentially the same than the one used in wired networks. The main difference is that, just like in the Internet until the early eighties, link costs are allowed to change at a timescale of some seconds [88]. The more static configuration that is used nowadays is due to the oscillations caused by these dynamic costs. It seems like history is repeating itself, since early experiments with WMNs have also reported routing oscillations [6, 89].

However, a completely static routing approach is not a suitable solution in this context. Static means non-optimized routing. In the wired case this is not such a big issue, since resources, specially in the core, are relatively inexpensive (in fact, most core networks are overprovisioned). On the contrary, in wireless networks resources are intrinsically scarce, and “upgrading” a link’s capacity is not always a possibility. Available resources must then be used at its maximum, and for this purpose a certain form of dynamism must be implemented in the network.

We present a novel approach which separates routing from forwarding, just like MPLS does in the wired context. That is to say, each ingress router has several possible paths towards the destinations, and these paths remain unchanged as long as no topological change takes place (e.g. a node failure). Please note that in the context of WMNs we may safely assume that nodes are fixed and do not change status nor position very often. Each new incoming flow will be forwarded along one of these paths, a decision that each ingress router will take depending on the current network condition. We shall call this procedure dynamic load-balancing. We propose one such scheme that forwards incoming packets along several pre-established paths in order to minimize a certain objective function. If correctly designed, load-balancing will bring improved performance over static routing, without the difficult to avoid oscillations of pure dynamic routing. For more arguments in favour of load-balancing see the discussion presented in [90], where Caesar *et al.* argue for a separation of timescale between *offline* computation of multiple paths and *online* spreading of load over these paths, or the analysis by Pham *et al.* [91] where single-path and multi-path routing protocols are compared in a wireless networks scenario, showing that the latter provides better performance.

We consider a particular but very typical scenario: a planned WMN, where all bidirectional point-to-point links do not interfere with each other. This assumption means either that all backhaul links use different channels or that links in the same channel are in different collision domains. There are many scenarios where this assumption holds, for example suburban or rural area networks and even campus networks, deployed with high directional antennas with proper RF design and channel assignment. This assumption also implies that the network topology is already defined, typically at infrastructure deployment phase. This means we cannot decide which backhaul links to establish but only how to use them, i.e. which traffic route through them. All these assumptions hold for the real world case scenario from Plan Ceibal, introduced in Chapter 2.

The question that remains is to what purpose should load-balancing serve and be worthwhile. That is to say, what function of the traffic distribution should be optimized (where “traffic distribution” refers to the portion of traffic sent along each path). In this work we argue that this function should be the sum over all nodes’ interfaces of the corresponding average queue length. As shall be discussed in Section 4.3, this is a very versatile and important performance indicator. The problem we address is then to find the traffic distribution that minimizes the sum over all interfaces of the average queue size. However, instead of relying in analytical expressions based on (arbitrary) models, we will strive at reflecting reality as much as possible, and design a measurement-based scheme. In this framework the

Chapter 4. Routing and Forwarding in Wireless Mesh Networks

relationship between the average queue length and the current traffic distribution will be learned from measurements, and the optimization shall be performed based on this learned function.

This kind of approach, using a network model developed from measurements of queue sizes and traffic loads, has already proved suitable for a wired scenario [92]. In this work, we extend the framework to the previously described wireless scenario. Furthermore, we also consider the dynamic gateway selection problem and we obtain a load-balancing solution using the proposed approach. Differently to the wired case, in the considered wireless scenario the average queue size at a given interface now depends not only on the incoming traffic, but also on the activity of the interface at the other end of the link. We model each link with only one average queue (the sum of both interfaces involved) which depends on the traffic in both link directions. A method to learn this bi-variable function is presented, whereas simulations illustrate the framework.

It is important to highlight that we are considering a WMN where links performance is stable and predictable, with a strong correlation between the error rate and the received signal strength. In the context of WMN, as stated in [93], interference (and not multipath fading) is the primary cause of unpredictable performance. In the scenario of interest there is no internal interference (as we assumed links do not interfere with each other), so we expect to achieve an accurate model, based on the results obtained in the previous chapter, with similar learning techniques. The difference in this case to what was done in Chapter 3, is that we will not try to infer directly the effective capacity of the wireless links, but we will model the packet queue behavior for each link, depending on the ongoing traffic.

4.2. Related Work

In the context of WMNs, several previous works presented new metrics for single path routing that take into account information from lower layers [15]. The need to increase the WMNs capacity led to the use of nodes with multiple radio interfaces which was analyzed in [22,23]. In this work we consider a planned WMN, where all links do not interfere with each other. Even in an unplanned scenario several algorithms have been proposed [18–20] which could be used to schedule the links so that they do not interfere with each other.

There are some recent related works that we would like to highlight. In [48] an optimization framework is presented to reach minimum average delay per packet in a single channel WMN. Starting from a Markov chain model for the medium access of a single node, they derived a closed form representation for the average system delay which is used as the objective function. The model takes into account the neighbors interference but several parameters of the Markov chain need to be calculated or defined which could difficult the implementation.

Another work that uses an analytical model in the context of single channel WMNs is [44]. In particular, the authors developed a queueing-based model which is used to estimate the network capacity and to identify network bottlenecks. Based on a load-aware routing metric they choose the corresponding path for

each new incoming flow, and then, based on the obtained model, a centralized entity performs admission control to guarantee network stability. They focused on per-flow performance and compare the results with the shortest-path first routing algorithm.

Concerning dynamic gateway selection, in [94] an heuristic algorithm was proposed to tackle the problem. A single channel WMN is considered between routers, but operating in a different channel than links between WMN nodes and mobile hosts. They assume that a routing protocol is executed in the WMN which establishes routes between every pair of nodes, including the gateways. They seek to minimize the maximum number of flows served by a gateway and minimize the cost of paths in order to avoid interference in the network. Contention regions are modeled as the maximal cliques¹ of the contention graph, which leads to a Mixed Integer Nonlinear Programming (MINLP) formulation of the problem. Their proposal solves gateway selection for Internet flows in a centralized manner using a greedy heuristic.

To the best of our knowledge, the only work that proposes a forwarding scheme for WMNs is the recent article [95], where the authors present an MPLS-based forwarding paradigm. However, two important differences with our proposal should be highlighted. Firstly, they allow traffic splitting at every node in the network while we only allow it at ingress routers. Secondly, and most importantly, they considered the *hose* traffic model (only knowledge about maximum traffic demands) which leads to a robust routing fashion to solve the problem. The optimization cost function of a routing solution is calculated as the average over all the feasible flows allocations, where the function used is a weighted average of the total utilizations over all the collision domains. We think that in the context of WMNs, it is more appropriate to consider a dynamic load-balancing solution rather than a robust routing scheme, because it is exactly in scenarios with highly dynamic traffic like WMNs where the former takes advantage over the latter. In addition, this scenario is the most typical one in the educational context presented in Chapter 2. For a deep comparison between both methods please refer to [96].

All in all, two major differences should be distinguished between our proposal and previous works. The first one is the introduction of a measurement-based model for 802.11 links, whereas most of the literature is based on (arbitrary) MAC layer models like the one presented in Bianchi’s seminal paper [9]. The second important difference is the timescale at which decisions are taken. Most of routing algorithms proposed for WMNs are based on a certain metric which changes at a timescale of seconds. Our framework operates with averages taken over tens of seconds and forwarding decision is taken with flow granularity. This fact enables decoupling the link model learning phase from the forwarding decision, and ensures better stability properties avoiding the route flapping problem.

¹A clique is a complete subgraph of at least 2 vertices.

4.3. Network Model and Problem Formulation

Firstly, let us remark that in the context of WMNs we may safely assume that nodes are fixed and do not change position very often. In addition, we will assume that power supply is available at every node, so we will completely ignore energy consumption. We will then concentrate on the performance as perceived by packets in terms of delay, dropping probability and throughput. Naturally, we will limit ourselves to the WMN, which means that throughput will refer to a quantity proportional to the inverse of the time that it takes any given packet to leave the network.

Before introducing the notation, let us highlight that throughout this part of the thesis we will assume that each node has a single FIFO queue attached to each of its (possibly several) interfaces. This means that all packets at each interface will receive the same treatment, independently of its destination, number of traversed hops, etc. This is not a very problematic assumption, since the only queue management that most wireless routers implement is some form of prioritization of certain particular and few packets (e.g. ARP packets).

Let $n = 1, \dots, N$ be the set of static wireless mesh routers (including gateways) which we shall call *nodes* and $l = 1, \dots, L$ the backbone bidirectional links in the network. Typically, high gain directional antennas are used for backhaul links with other nodes and sector panels or omnidirectional antennas are used to provide connectivity for mobile stations. Gateways nodes have also wired links to a fixed infrastructure network with Internet access. We will focus on the mesh core, so only backhaul links and aggregated traffic at mesh routers will be considered. Traffic generated at node n will refer to all traffic arriving at n from the mobile hosts attached to it. We will assume that this traffic uses different channels (e.g. 802.11b/g/n at 2.4 GHz) than the ones used within the mesh core (e.g. 802.11a/n at 5GHz). If n is a gateway, the generated traffic also includes that coming from the Internet to nodes in the WMN. As we mentioned before, we shall further assume that channels within the mesh core do not interfere with each other. Moreover, paths are assumed to be established *a priori* and how to choose them is out of the scope of the present work. In particular, we will use the k shortest paths.

Traffic generated at a node will have as final destination a set of nodes, which may contain for instance any other node in the WMN. This defines a set of possible origin-destination (OD) pairs, which we shall index by the integer $s = 1, \dots, S$. The amount of traffic corresponding to OD pair s will be noted by d_s and we further define the column vector $\mathbf{d} = [d_1 \dots d_S]^T$. We will assume that entries in \mathbf{d} are independent of each other. In particular, this means that the amount of traffic sent to the Internet through a particular gateway does not influence the amount of traffic that gateway generates.

Each pair will have a set of n_s fixed, established *a priori* paths, which we shall note as P_{si} for $i = 1, \dots, n_s$. The amount of traffic sent along path P_{si} shall be noted as $d_{P_{si}} = \alpha_{P_{si}} d_s$, where $\alpha_{P_{si}}$ is the traffic distribution coefficient for path P_{si} . We further define $\boldsymbol{\alpha} = [\alpha_{P_{11}} \dots \alpha_{P_{1n_1}} \alpha_{P_{21}} \dots \alpha_{P_{S1}} \dots \alpha_{P_{Sn_s}}]^T$ as the traffic distribution vector. The following two constraints should hold $\sum_{i=1}^{n_s} d_{P_{si}} = d_s \quad \forall s$

4.3. Network Model and Problem Formulation

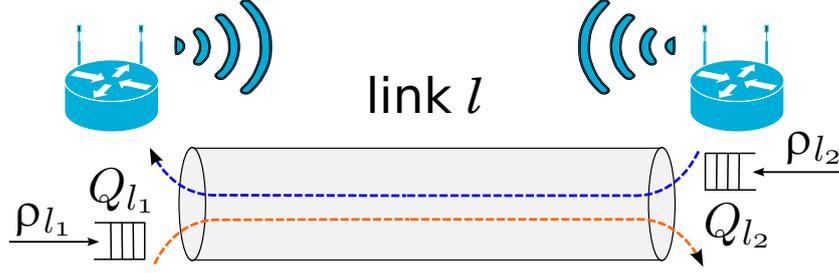


Figure 4.2: Wireless link queues and flows in both directions.

and $d_{P_{si}} \geq 0 \quad \forall s, i$, which implies $\sum_{i=1}^{n_s} \alpha_{P_{si}} = 1 \quad \forall s$ and $\alpha_{P_{si}} \geq 0 \quad \forall s, i$.

Within this context, for each link l we have two traffic loads, one for each direction of the communication, which we shall call ρ_{l_1} and ρ_{l_2} , taking any arbitrary convention (see Figure 4.2). Given a demand vector \mathbf{d} and a traffic distribution vector $\boldsymbol{\alpha}$, the total traffic load on link l in one direction (e.g. ρ_{l_1}) is given by the sum over all OD pairs of the traffic forwarded along those paths P_{si} which use the link in that direction. Let D_{l_1} be the average amount of time a packet spends at the queue of link l in the direction of load ρ_{l_1} . Naturally, this non-decreasing function depends on the traffic load ρ_{l_1} , which is the queue's input traffic intensity. However, and due to the half-duplex operation of the link and the 802.11 medium access control, D_{l_1} also depends on the load in the opposite direction (ρ_{l_2}).

Let us now discuss with more detail what this delay is composed of. Once a packet enters a node interface queue, it has to wait for several things to happen. Firstly, it has to reach the head of the line of the queue. What happens after then depends on whether the node is a gateway and the packet goes to the Internet, or not. In the former case, it has to wait for all its bits to be sent by the wired interface. In the latter case, it has to wait for the channel to be idle. Once this happens, the packet has to be correctly received by the destination node. This includes the transmission delay plus maybe some retransmissions. It is important to highlight then that queuing delay captures several aspects of the wireless link operation: congestion at the MAC layer, transmission errors at PHY layer and the chosen modulation rate.

Let D_P be the average end-to-end delay of path P . Note that, as mentioned above, the throughput of path P is proportional to the inverse of D_P . This fact in addition to what we discussed above suggests the use of the average end-to-end queuing delay in the network $D(\mathbf{d}, \boldsymbol{\alpha})$ as a total congestion measure:

$$D(\mathbf{d}, \boldsymbol{\alpha}) := \sum_{s=1}^S \sum_{i=1}^{n_s} d_{P_{si}} D_P = \sum_{s=1}^S \sum_{i=1}^{n_s} \alpha_{P_{si}} d_s D_P \quad (4.1)$$

Notice that this measure depends, on the one hand, of the vector \mathbf{d} , defined by the OD traffic demands, which cannot be set as desired because they are given by the network usage (e.g. the traffic demand shown in Chapter 2). On the other hand, the function also depends on the traffic distribution vector $\boldsymbol{\alpha}$, which we can

Chapter 4. Routing and Forwarding in Wireless Mesh Networks

control and will set so as to minimize the network congestion. Then, it is easy to prove that the sum over all the paths is equal to the sum over all the links, so we have:

$$D(\mathbf{d}, \boldsymbol{\alpha}) = \sum_{l=1}^L D_{l_1}(\rho_{l_1}, \rho_{l_2}) \rho_{l_1} + D_{l_2}(\rho_{l_2}, \rho_{l_1}) \rho_{l_2} \quad (4.2)$$

Let Q_{l_1} and Q_{l_2} be the mean amount of bytes on link l queues on each direction. Then, by Little's law we obtain the following result: $Q_{l_1} = D_{l_1} \times \rho_{l_1}$ and $Q_{l_2} = D_{l_2} \times \rho_{l_2}$. Finally $D(\mathbf{d}, \boldsymbol{\alpha})$ is given by:

$$D(\mathbf{d}, \boldsymbol{\alpha}) = \sum_{l=1}^L Q_{l_1}(\rho_{l_1}, \rho_{l_2}) + Q_{l_2}(\rho_{l_1}, \rho_{l_2}) \quad (4.3)$$

$$= \sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \quad (4.4)$$

where Q_l is the average sum over both link queues (i.e. $Q_{l_1} + Q_{l_2}$ in Figure 4.2). In Section 4.4 we will present a measurement-based scheme to characterize $Q_l(\rho_{l_1}, \rho_{l_2})$.

All in all, the dynamic load-balancing scheme should strive at solving the following problem:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & D(\mathbf{d}, \boldsymbol{\alpha}) = \sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \\ \text{s.t.} \quad & \sum_{i=1}^{n_s} \alpha_{P_{si}} = 1 \quad \forall s, \\ & \alpha_{P_{si}} \geq 0 \quad \forall s, i. \end{aligned} \quad (4.5)$$

Let us further justify our choice of the objective function. Equation 4.1 suggests that our objective function may be regarded as a weighted average end-to-end delay, where the weight of each path is how much traffic is being sent along it. This means that $\sum_l Q_l$ considers both delay and throughput at the same time. Concerning dropping probability, the last of the three performance indicators cited before, it should be clear that a larger value of it will result in a larger queue at the output air interface, resulting in a larger $\sum_l Q_l$. The conclusion of this discussion is that $\sum_l Q_l$ is a number that is affected by the three performance indicators, and as such reflects the three of them. We referred to this when we said before that $\sum_l Q_l$ is a versatile indicator.

4.4. Learning the Wireless Link Dynamics

In this section we present the procedure to choose the most appropriate function $Q_l(\rho_{l_1}, \rho_{l_2})$ for every 802.11 link in the network. We shall omit the subindex l for a matter of clarity, since the procedure is the same for every link. The function $Q(\rho_1, \rho_2)$ is not trivial as we are dealing with 802.11 wireless links which use CS-MA/CA as medium access control mechanism. Several works since [9] have tried

4.4. Learning the Wireless Link Dynamics

to find the relation between wireless link parameters and the corresponding TCP and UDP achievable throughput. We use a different approach, that has already proved suitable for wired links [92], which is learning the function from measurements. This way we avoid using an arbitrary model and reflect reality as much as possible. However, the learning procedure should be carried out with some care. For instance, differently to the wired case, the average queue length at a given link is now a bi-variable function, because it depends not only on the incoming traffic, but also on the traffic in the opposite direction.

Assume we have a set of N measurements $\{Q_1, Q_2, \dots, Q_N\}$ for the corresponding values $\{(\rho_{1_1}, \rho_{2_1}), (\rho_{1_2}, \rho_{2_2}), \dots, (\rho_{1_N}, \rho_{2_N})\}$ (also called training set). Assume that the response variable Q (the average queue length measurement) is related to (ρ_1, ρ_2) (the link average traffic loads measurements) by the following equation:

$$Q = f(\rho_1, \rho_2) + \epsilon \tag{4.6}$$

where ϵ is the measurement error and is modeled as a random variable such that $E\{\epsilon\} = 0$ and $\text{Var}\{\epsilon\} = \sigma < \infty$. The Weighted Least Squares (WLS) problem consists in finding the function f that minimizes the weighted sum of quadratic errors, assuming that f belongs to a given family of functions \mathcal{F} . The weights represent the relative importance of each measurement point with respect to the rest of the measurements in the training set.

We present a method that restricts the assumptions on the family of functions \mathcal{F} to the minimum. Regarding its shape, we have only two necessary assumptions:

1. $f(\rho_1, \rho_2)$ should be non-decreasing, since more load may never mean less queue length.
2. $f(\rho_1, \rho_2)$ should be convex in order to guarantee the existence and uniqueness of the optimum demand vector (later on we will discuss on this assumption).

We then consider \mathcal{F} as the family of continuous, monotonous increasing and convex functions. This WLS problem with such \mathcal{F} is called Convex Non-parametric Weighted Least Squares (CNWLS), a variation of the original unweighted Convex Non-parametric Least Squares (CNLS) [97]. The size of \mathcal{F} makes this problem very difficult to solve in such general form, which motivates to use instead a subfamily of \mathcal{F} , the piecewise linear functions included in \mathcal{F} . This leads us to a standard finite dimensional Quadratic Programming (QP) problem in order to solve the regression, for which mature methods to solve it exist (e.g. interior point algorithms) and several solver software are available (for instance, we used MOSEK [98]). This scheme is easily adaptable to update the function in real time through online learning as new measurements are gathered from the network. This fact could be useful to react properly to physical changes that may affect the link capacity (e.g. antenna misalignment or environmental changes).

4.4.1. Convexity Assumption

We will now discuss on the convexity assumption mentioned before. A necessary condition for the convexity of $Q(\rho_1, \rho_2)$ is that the feasible region of the link is

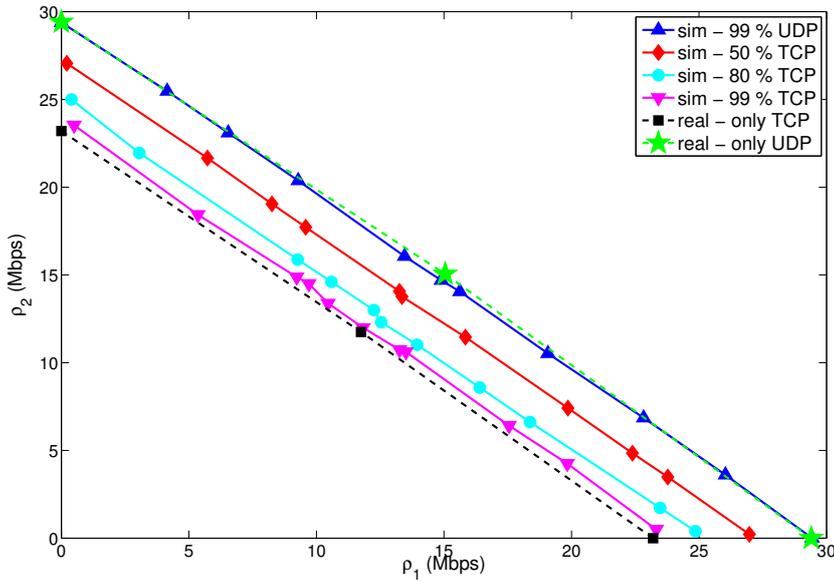


Figure 4.3: Feasible region analysis for a 802.11a link @54Mbps.

convex (i.e. the set of $\{(\rho_1, \rho_2)\}$ such that $Q(\rho_1, \rho_2) < \infty$). Several previous works studied the feasible region for 802.11 wireless multihop networks. This region is known to be not necessarily convex, which is demonstrated in [99] with models and simulations for different topologies. This fact is also analyzed in [100], where the log-convexity of this region is established, a fact that is taken as a basis for characterizing max-min fair rate allocations for 802.11 WMNs in [101]. However, the model presented in [102] approximates the feasible region by a convex polytope. The procedure is based on the computation of extreme points in order to get the polytope convex hull (boundary) and it is shown that most of the cases presented in [99] can be adequately captured by this model.

For the case we are considering in this work, a planned WMN, the analysis is much simpler because we have only two nodes that can interfere with each other (i.e. the endpoints of each link). This simplifies the feasibility region analysis to the study of the behavior of only one link as the traffic loads in both directions changes. For this purpose, first let us take a look at the well-known Bianchi model [9] to notice that the capacity for two nodes is 2.5% larger than for a single node. This fact indicates that two simultaneously transmitting nodes may support more traffic than only one, which means that feasible region of a 802.11 point to point link should be convex. We further studied the feasible region for a 802.11a link operating at 54 Mbps with simulations performed with the ns-3 simulator [103] and real data measurements. In Figure 4.3 we can see the results for different traffic compositions combining TCP and UDP flows. As we can see, the feasible region increases as the proportion of UDP traffic increase, with throughput ranging from 24 to almost 30 Mbps. It is clear from the results that for all cases it is suitable to use a convex model as an approximation, as used in [102].

4.4.2. Average Queue Regression Example

In order to illustrate the proposed procedure we will show an example with simulations performed with ns-3. We configured a wireless link operating in 802.11a, with a distance of 100m between nodes, while the propagation model used was fixed received signal strength (RSS = -65 dBm). This implies that the link is always operating at the same modulation rate (54Mbps in this case).

As we said before, we are considering a WMN where links performance is stable and predictable, with a strong correlation between the error rate and the received signal strength. Under this assumption, if we do not have much RSS variation for our network links, we will not have variation at all on each link modulation rate. This assumption is valid for a wide range of WMNs, not only in rural or suburban areas, but also in some urban scenarios with LOS links using directional antennas. As an example in Figure 4.4 we show the RSS for one week for two urban links from Plan Ceibal network. Both of them operate with line of sight and with an approximated distance of 200m between nodes. As we can see the RSS variation is not significant and enables a stable link operation at a fixed modulation rate, as the receiver sensitivity for 54Mbps is -71dBm. This fact is consistent with the data shown in [93] and with the measurements collected in Chapter 3.

Now, we present an example for one link to illustrate the procedure followed for every link in the network in the learning phase. In this example we generated a dataset of 484 measurements, 228 used for learning the function and the remaining 256 for testing the regression performance. To generate each flow with the desired traffic load ρ , we used a combination of random TCP and UDP flows (80 % and 20 % respectively). TCP flows were generated with exponential file sizes with mean 500 Kbytes. UDP flows were generated with a fixed rate of 100 kbps and exponential length with mean 30 seconds. The arrival rate distribution was also exponential for both cases, with mean according with the desired traffic loads (0.8ρ and 0.2ρ respectively). Each measurement corresponds to the average traffic load in both directions (ρ_1, ρ_2) and the average queue length Q , where averages are considered over 100 seconds.

In Figure 4.5 we present the resulting function after the regression in logarithmic scale (for the sake of clarity). Queue size is expressed in packets because both ns-3 simulator and typical wireless equipment use 802.11 packet-based queues [104]. The RMSE for training data was 4.3 packets, while the RMSE for test data was 5.5 packets. The relative RMSE for training data was 2.9 % with a maximum of 14.2 %, while for test data was 4 % with a maximum of 15.2 %. The RRMSE for test data is shown in Figure 4.6. This results show that the function approximation is suitable.

The presented example is only to show the procedure we followed for every link in the network in the learning phase. Once we have learned the function $Q_l(\rho_{l_1}, \rho_{l_2})$ for each link, we are in position to tackle the optimization problem defined in Section 4.3. The forwarding decision will come up from the optimum traffic distribution vector α which minimizes the total network congestion.

Chapter 4. Routing and Forwarding in Wireless Mesh Networks

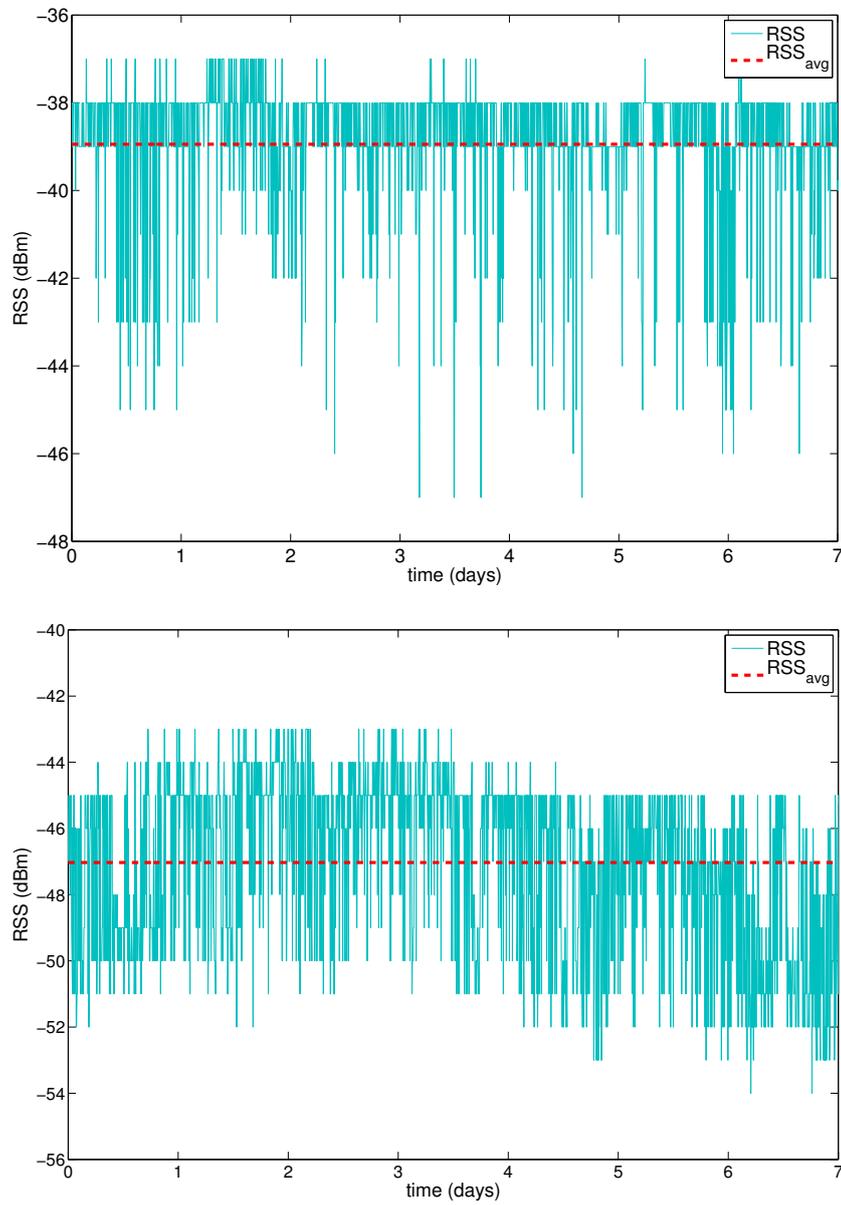


Figure 4.4: RSS measurements for two real 802.11 links.

4.4. Learning the Wireless Link Dynamics

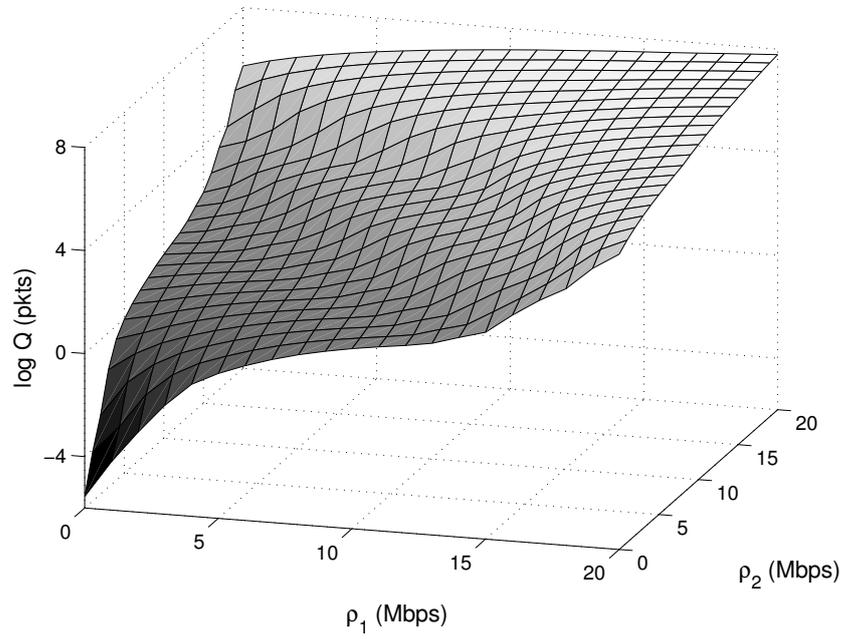


Figure 4.5: Learned function (log-scale) for the average queue size.

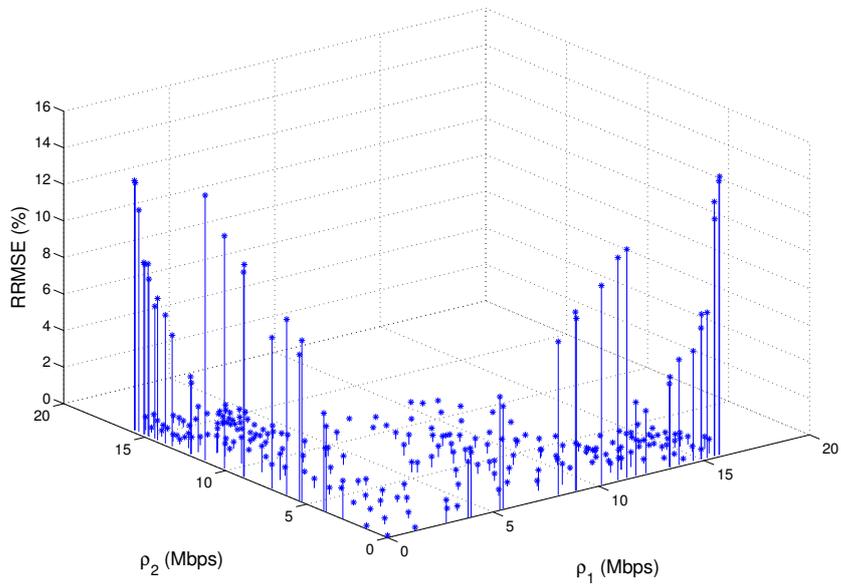


Figure 4.6: Relative RMSE for the example test data.

This page was intentionally left blank.

Chapter 5

Minimum Queue Length Load-Balancing

In this chapter we rely on the previously developed model, in order to arrive at a suitable solution of the optimization problem posed. The formulation achieved in the previous chapter can be seen as a particular case of convex optimization theory applied to a computer networks problem. One of the most famous frameworks of this kind is known as network utility maximization (NUM), developed by Kelly *et al.* more than two decades ago. Then, we discuss briefly about this point and the benefits this has to find an algorithm that leads to the solution of the optimization problem. Thereafter, we introduce the distributed algorithm that carries out the dynamic load-balancing scheme and we discuss some issues that arise in a real world implementation. Finally, several simulation experiments are presented to validate the framework and illustrate the advantages against other options.

5.1. Convex Optimization Problem Formulation

The process followed to pose a resource allocation issue as an optimization problem, which seems natural today, has its roots in the development of the Network Utility Maximization (NUM) framework. The NUM framework has its origin in the seminal paper of Kelly *et al.* [105] and dates from a time when a large research community was attracted by the congestion control problem in data networks. The formulation by Kelly *et al.* set the congestion control problem as a suitable convex optimization problem, where connections are represented as economic actors in a *bandwidth market*. The network congestion signals are interpreted as prices, and the market equilibrium becomes the resource allocation obtained through the decentralized mechanisms.

This formulation proved to be a valuable tool, connecting two previously disjoint areas, on the one hand the traditional layer analysis of network protocols and on the other the convex optimization theory. Since then, this approach has been extensively used to characterize real world network protocols and also to design new ones, that improve the network performance. While it was originally applied

to model the TCP congestion control mechanism, its use was later extended to many other problems. As an example we can cite the work of Lin *et al.* where it is used for cross-layer optimization in wireless networks [21]. Some other applications and extensions to the model can be found in [106, 107] and in the references therein.

The main idea of the NUM framework is to apply an economic network model to solve a resource allocation problem. The goal is to maximize a certain utility function U which depends on how much resources x are allocated. This maximization is subject to a certain capacity constraint, which limits the set of possible assignments. In our case, instead of seeking to maximize a utility function, what we have is the minimization of a cost function, which is completely equivalent (just consider a utility function equal to the opposite of the cost function $U(\mathbf{d}, \boldsymbol{\alpha}) = -D(\mathbf{d}, \boldsymbol{\alpha})$).

Convex optimization problems for routing and forwarding of the type considered here, were studied before in the context of wired networks (see Section 5.4 in Bertsekas and Gallager's book [108], and the references therein) . Also in this context, MPLS-based schemes similar to our approach were proposed by Elwalid *et al.* [109]. Let us recall the optimization problem raised:

$$\begin{aligned}
 & \underset{\boldsymbol{\alpha}}{\text{minimize}} && \sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \\
 & \text{subject to:} && \\
 & && \sum_{i=1}^{n_s} \alpha_{P_{si}} = 1 \quad \forall s, \\
 & && \alpha_{P_{si}} \geq 0 \quad \forall s, i.
 \end{aligned} \tag{5.1}$$

The problem constraints state that the sum of traffic distributions must equal the total traffic demand for each origin-destination pair and that a non-negative traffic portion must be allocated for each possible path. It is well known that a strictly convex function has a unique minimum over a closed and bounded convex set. Then, the problem satisfies the requirements as the cost function is strictly convex, and the constraint set is closed (since the aggregate traffic over all paths for each origin-destination pair is equal to a constant) and bounded (since the traffic demand for every origin-destination pair is finite). In addition, the constraint set for the minimization problem is convex which allows us to use the method of Lagrange multipliers and the Karush-Kuhn-Tucker (KKT) theorem [110]. This ensures that some $\boldsymbol{\alpha}$ satisfying the KKT conditions is the unique global minimum.

Considering the Lagrangian function we have:

$$L(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) + \sum_{i=1}^S \lambda_s \left(\sum_{i=1}^{n_s} \alpha_{P_{si}} - 1 \right) - \sum_{i=1}^S \sum_{i=1}^{n_s} \mu_{P_{si}} \alpha_{P_{si}} \tag{5.2}$$

The theorem ensures the existence of unique Lagrange multipliers $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^* \geq 0$ verifying the following KKT conditions, necessary for a local minimum $\boldsymbol{\alpha}^*$:

5.1. Convex Optimization Problem Formulation

- $\theta_{P_{si}} + \lambda_s^* - \mu_{P_{si}}^* = 0,$
- $\mu_{P_{si}}^* \alpha_{P_{si}}^* = 0,$

where $\theta_{P_{si}} = \frac{\partial}{\partial \alpha_{P_{si}}} \left(\sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \right) = \sum_{l \in P_{si}} \frac{\partial}{\partial \alpha_{P_{si}}} (Q_l(\rho_{l_1}, \rho_{l_2})).$

Note that the second condition means that the associated Lagrange multiplier $\mu_{P_{si}}$ is positive only for those paths that are not used at optimality ($\alpha_{P_{si}}^* = 0$), else it is equal to zero. Then, we shall define the cost for path P_{si} at the optimum α^* as:

$$\phi_{P_{si}}^* = \sum_{l \in P_{si}} \phi_l^* \doteq \sum_{l \in P_{si}} \left. \frac{\partial Q_l}{\partial \alpha_{P_{si}}} \right|_{\alpha^*} = \begin{cases} -\lambda_s^* & \text{if } \alpha_{P_{si}}^* > 0 \\ -\lambda_s^* + \mu_{P_{si}}^* & \text{if } \alpha_{P_{si}}^* = 0 \end{cases} \quad (5.3)$$

This necessary condition is true only if $Q_l(\rho_1, \rho_2)$ is continuously differentiable (i.e. its derivatives are continuous). As this function Q_l is convex, something that we imposed at the learning phase, the above condition becomes necessary and sufficient. Moreover, in such case the optimum is unique. All paths that are used at optimality have the same cost, which is actually the minimum cost among all paths of the corresponding origin-destination pair, and corresponds to the sum over all the links in the path of a certain link cost function ϕ_l , defined as the partial derivative of Q_l with respect to $\alpha_{P_{si}}$.

Considering the Lagrangian multipliers, a couple of facts can be highlighted:

- In general, for path P_{si} , the optimal traffic distribution coefficient $\alpha_{P_{si}}$ is only determined by the Lagrange multipliers on its route. This feature is extremely useful in designing decentralized algorithms to reach the optimal solution.
- The value of $\alpha_{P_{si}}$ is inversely proportional to the sum of the Lagrange multipliers on its route. In general, $\alpha_{P_{si}}$ is a decreasing function of the Lagrange multipliers. Thus, the Lagrange multiplier associated with a link can be thought of the price for using that link and the price of a path can be thought of as the sum of the prices of its links. If the price of a path increases, then the amount of traffic routed through that path decreases.

Because of the above, which is a typical property of this kind of models, the problem can be readily decomposed. Then, distributed algorithms can be developed, where each of the sources controls its local variable, such as the traffic distribution coefficients, based on local observables, such as link loads or path prices. By techniques such as Lyapunov function or the descent lemma, global or local asymptotic convergence towards the optimum can be proved for these distributed algorithms [106]. A key insight is that the effects of network protocols can be understood as the trajectories of a controlled dynamic system. In the next section we will introduce the design of the distributed load-balancing algorithm to reach the optimum for the particular problem proposed.

5.2. Distributed Algorithm Proposal

In order to drive the network to the desired operation point of minimum average congestion, we have to solve the optimization problem previously detailed:

$$\begin{aligned}
& \underset{\boldsymbol{\alpha}}{\text{minimize}} && \sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \\
& \text{subject to:} && \\
& && \sum_{i=1}^{n_s} \alpha_{P_{si}} = 1 \quad \forall s, \\
& && \alpha_{P_{si}} \geq 0 \quad \forall s, i.
\end{aligned} \tag{5.4}$$

For this purpose, we used a gradient descent method [110] to iteratively update the traffic distribution vector $\boldsymbol{\alpha}$ by setting the proper load balancing leading to the optimum. We can assure that there are no local minima because we are minimizing a sum of convex functions, which is also a convex function. To start the optimization algorithm we need an initialization step, so certain initial values have to be set to enable the network to begin the operation. Then, we consider a periodic update every ΔT seconds, given by:

$$\boldsymbol{\alpha}^{t+\Delta T} = \boldsymbol{\alpha}^t - \gamma \cdot \nabla \left(\sum_{l=1}^L Q_l(\rho_{l_1}, \rho_{l_2}) \right) \tag{5.5}$$

where γ is the gradient descent step size. Before updating $\boldsymbol{\alpha}$ we have a normalization step to guarantee the constraints on $\alpha_{P_{si}}$. With this procedure the demands are periodically adjusted, using the following equation for updating the traffic distribution coefficient which corresponds to the path P_{si} :

$$\hat{\alpha}_{P_{si}}^{t+\Delta T} = \left[\alpha_{P_{si}}^t - \gamma \sum_{l:l \in P_{si}} \frac{\partial Q_l}{\partial \rho_{l_{si}}}(\rho_{l_1}^t, \rho_{l_2}^t) \right]^+ \tag{5.6}$$

$$\alpha_{P_{si}}^{t+\Delta T} = \hat{\alpha}_{P_{si}}^{t+\Delta T} / \sum_{i=1}^{n_s} \hat{\alpha}_{P_{si}}^{t+\Delta T} \tag{5.7}$$

Notice that the partial derivatives in the second term of Equation 5.6 are with respect to $\rho_{l_{si}}$, which is the traffic load of link l in the direction that corresponds to path P_{si} . This fact implies that for updating the traffic distribution coefficients $\alpha_{P_{si}}$ we only need to know the learned functions for the links used by the path P_{si} , which means that edge routers only need information from the intermediate routers included in the pre-established paths they will use, enabling a decentralized implementation of the algorithm. All the notation used in this part of the thesis is summarized in Table 5.1.

The complete network operation is defined by the three processes: measurement-based learning of the objective function, update of traffic demands distribution via

Table 5.1: Index of key notations.

Variable	Description
$1, \dots, n, \dots, N$	Set of nodes (i.e., wireless mesh routers)
$1, \dots, l, \dots, L$	Set of bidirectional links
$1, \dots, s, \dots, S$	Set of OD pairs
d_s	Average traffic demand for OD pair s
n_s	Number of paths for OD pair s
P_{si}	i -th path for OD pair s
$d_{P_{si}}$	Average amount of traffic for path P_{si}
\mathbf{d}	Average traffic demands vector
$\alpha_{P_{si}}$	Traffic distribution coefficient for path P_{si}
$\boldsymbol{\alpha}$	Traffic distribution vector
ρ_{l_1}, ρ_{l_2}	Average traffic load on link l for each direction
D_{l_1}	Average delay at link l in the direction of load ρ_{l_1}
D_P	Average delay at path P
Q_{l_1}	Average queue size at link l in the direction of load ρ_{l_1}
Q_l	Sum of the average queues sizes at link l
$\alpha_{P_{si}}^t$	Traffic distribution coefficient for path P_{si} at time t
γ	Gradient descent step size

gradient descent optimization and packet forwarding on a per-flow basis. These processes operate at different timescales, which will be detailed next.

At the longest timescale we have the measurement-based learning of the average queue length function, which takes several hours of information to update the $Q_l(\rho_{l_1}, \rho_{l_2})$ for every link in the network following the procedure described in the previous section.

Then we have the update of traffic demands distribution in order to lead the network to the minimum queue length load-balancing (i.e. for each OD pair we use Eqs. 5.6 and 5.7). In this case each iteration is performed at a shorter timescale than model learning, but a much longer timescale than packet forwarding. The optimization takes into account average values, so we need an update period long enough to take good quality average measurements. On the other hand, this period should not be excessive in order to be able to respond quickly when traffic

Chapter 5. Minimum Queue Length Load-Balancing

conditions change abruptly. Typically a suitable period is some tens of seconds, which is the minimum time to get reasonable average measurements (e.g. we used 100 seconds).

The shortest timescale corresponds to the packet forwarding, which is performed with flow granularity. This means that every new traffic flow at an ingress router corresponding to OD pair s is associated with a certain path P_{si} with probability $\alpha_{P_{si}}$. Let us recall that we have certain pre-established paths defined by the network topology. This packet forwarding scheme is very similar to the one used in wired networks with MPLS. Several paths are defined at edge routers, where incoming traffic is labeled according to the corresponding path and then packet forwarding at relay routers is based on labels. This is why we say that our proposal of separating routing from forwarding is a solution *à la* MPLS.

With respect to the running time of each action, its precise value depends on the specific hardware at use (e.g. ingress router, relay nodes). However, it is clear that the more costly actions are the ones that operate at a longer timescale (i.e. function learning costs more than gradient descent and both of them more than forwarding). The different timescales involved in the proposed mechanism are resumed in Figure 5.1.

5.3. Discussion on Implementation Issues

The application of the proposed framework in a real-world network is relatively simple. First of all we need a routing protocol to establish the multiple routes for each OD pair defined by the wireless network topology. Once we have learned Q_l for every link l , each ingress router receives the values ρ_l from the links used by the OD flows with origin in that ingress router. A routing protocol that supports information distribution such as OSPF-TE may be used for this purpose. With that information, each ingress router is able to update the traffic portion that has to be routed through each path. This process is repeated indefinitely every some seconds.

With respect to the flow-based multipath forwarding implementation, the idea is to use an MPLS-based solution, similar to the wired case. Although an standard

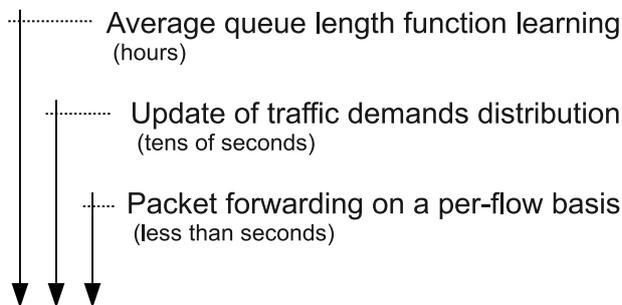


Figure 5.1: Processes involved in the proposed framework.

5.3. Discussion on Implementation Issues

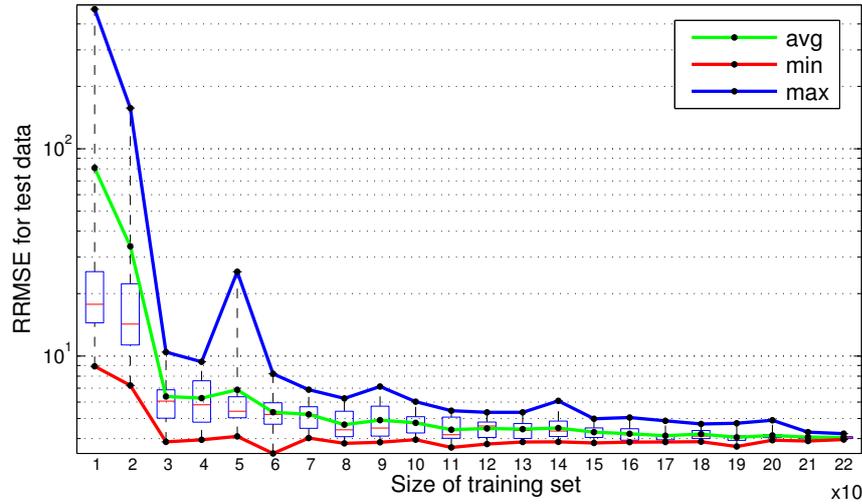


Figure 5.2: Training set size analysis. On each box, the central red mark is the median and the edges of the box are the 25th and 75th percentiles.

of MPLS over WMNs does not exist yet, several proposals were already presented. For example in [95] the proposal considers traffic splitting at every router and optimization over the average of all possible traffic matrices. Our proposal could be implemented reusing the same splitting-based scheme, but considering splitting only at ingress routers over all the different end-to-end paths and enabling dynamic load-balancing for the average load at each moment.

Regarding the learning phase we envisage several possibilities differing in the resulting architecture. One possibility is that a central entity gathers the measurements, performs the regression and communicates the obtained parameters to all ingress routers. This option has the advantage that the required new functionalities on routers are minimal. However, as all centralized architectures, it may not be suitable for some network scenarios, and handling the failure of this central entity could be very complicated. An alternative is that for each wireless link only the two directly involved routers perform the regression. They should keep the average queue size measurements for themselves, perform the regression and communicate the result to the ingress routers.

Another aspect that has different possibilities is what characterization (i.e. Q_l learned function) use at each moment and which measurements to keep for the training set. Measurements could be gathered every day, the regression performed, and its result could be used the next day or the same day the next week. In addition, it is clear that newer measurements should be given priority over older ones. A possible way to manage training data is to keep always the newer measurements and use weights in the regression to introduce temporal information (e.g. exponential decay). It may also be necessary to force keeping particular measurements to ensure a proper coverage density of the whole load value range.

Concerning the number of measurements needed for training, we now show how the considered learning algorithm (CNWLS) does not need a large number of measurements, as long as the training samples adequately covers the whole range of possible values. In Figure 5.2 we show the test error analysis for training sets with different sizes, using the same data as in the example discussed in section 4.4.2. In particular, for each size, we randomly sampled several training sets (we used 20) and computed the corresponding average RRMSE with the test data for the resulting learned function. As we can see, the RRMSE is always below 10% with only 60 training samples and falls below 5% with more than 150 samples.

Finally, rare events like node failures or changes in propagation conditions can be taken into account in our framework as follows. If interference on a particular link changes, this is captured when the learning of the function associated with that link is repeated. As we mentioned before, this learning process is periodically repeated. However, if several new measurements differ greatly from the learned model, one could decide to trigger a new learning process. Moreover, if a node fails, the ingress routers will not receive the corresponding link load information. If no such announcements are received for a certain period of time, this should lead to the decision of disabling all paths that use the faulty router.

5.4. Simulation Experiments

In order to validate the framework we tested the proposed minimum queue length load-balancing (MQLLB from now on) algorithm with simulations performed with ns-3. Most of the examples considered correspond to canonical topologies of WMNs [99] but also to typical configurations in real deployments (e.g. the wireless network from Plan Ceibal introduced in Chapter 2).

In this section we present four different examples considered for the simulation experiments. The first one is a three node topology used to describe the framework operation. In the second example we illustrate the gateway selection problem which can be solved within the same proposed framework. The third example corresponds to a four node topology where we deeply analyze the advantage of the proposed model under asymmetric traffic demands, comparing the performance with IEEE 802.11s. Finally, we present an example with a larger network, a 25 node uniform square grid, where we analyze convergence and scalability of the algorithm.

In all the examples the traffic considered is the same described before in Section 4.4 with a combination of TCP and UDP flows (80% and 20% respectively), both of them with exponential arrival rates. We also used exponential distributions for the file size (in case of TCP flows) and length (in case of UDP flows), with the same characteristics mentioned for the model learning example shown before. Wireless links were set to the standard 802.11a with a distance of 100m between nodes, while the propagation model used was fixed received signal strength (RSS = -65 dBm) which implies that links always operate at the same modulation rate (@54Mbps). The buffer size for each interface is 400 packets (ns-3 default) which is consistent with typical wireless equipment [104]. In every case, we used 235 measurements in the learning phase for each link, which is approximately 6.5

5.4. Simulation Experiments

hours of training data. Then, we implemented the MQLLB method which uses the described optimization framework to iteratively update α , taking the forwarding decision with a flow level granularity.

For performance comparison we considered as a benchmark the IEEE 802.11s routing scheme, which uses HWMP (Hybrid Wireless Mesh Protocol) [111] to compute paths. We think this benchmark is the most suitable one, as HWMP is the only algorithm included in an approved standard up to date and can be used by everyone to compare with. In addition, there are implementations available as the one included in the ns-3 simulator. Such protocol uses a routing metric called *airtime* metric which is designed to represent the channel resources needed for a frame to be transmitted over a wireless link and is calculated as follows:

$$airtime = \left(O_{ca} + O_p + \frac{B_t}{r} \right) \frac{1}{1 - e_{fr}} \quad (5.8)$$

where O_{ca} , O_p , and B_t are constants quantifying respectively the Channel Access Overhead, the Protocol Overhead, and the number of Bits in a probe frame. O_{ca} and O_p depend solely on the underlying modulation scheme, r is the transmission rate, and e_{fr} is the frame error rate. This routing metric is similar to ETX (Expected Transmission Count) and ETT (Expected Transmission Time) [15,22]. However, *airtime* further accounts for channel access and protocol overheads. An implementation of 802.11s is available in the ns-3 simulator.

For performance analysis and comparison we considered three metrics: average delay and jitter of UDP flows and average goodput of TCP flows, which corresponds to the amount of data per second carried by TCP flows discarding TCP ACKs. The analysis for each flow was done using the ns-3 flow monitor [112] which enables flow level statistical analysis of the simulation. We compared the results with the 802.11s performance for the different scenarios. We also considered static routing as a different alternative, using shortest path routing with hop count as metric.

5.4.1. Multipath Forwarding: 3-nodes Topology

The first example is presented to illustrate the framework and corresponds to the topology and flows shown in Figure 5.3. This topology has three links 1, 2 and 3, which implies we have also three functions Q_1 , Q_2 and Q_3 , each of them corresponding to the sum of the link queues in both directions. In this case we considered flows from node 1 to nodes 2 and 3 with traffic loads d_1 and d_2 respectively. When we apply the described framework to this particular topology and the considered traffic flows, we have the following function for the average end to end queueing delay in the network:

$$D(\mathbf{d}, \alpha) = Q_1(\rho_{11}, \rho_{12}) + Q_2(\rho_{21}, \rho_{22}) + Q_3(\rho_{31}, \rho_{32}) \quad (5.9)$$

For each OD pair we have two possible paths:

- $P_{11} = \{1, 2\}$ and $P_{12} = \{1, 3, 2\}$ for d_1 .

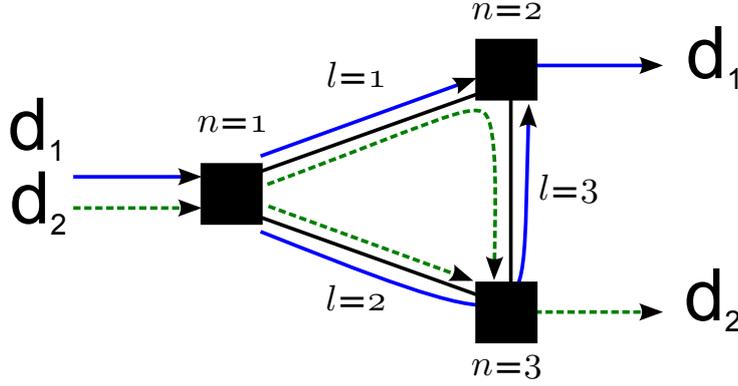


Figure 5.3: 3-nodes topology multipath forwarding example.

- $P_{21} = \{1, 2, 3\}$ and $P_{22} = \{1, 3\}$ for d_2 .

We will call $\alpha_{P_{11}}$ the portion of traffic d_1 that is routed through path P_{11} , which leaves $\alpha_{P_{12}} = 1 - \alpha_{P_{11}}$ through path P_{12} . We will call $\alpha_{P_{21}}$ the portion of traffic d_2 that is routed through path P_{21} , which leaves $\alpha_{P_{22}} = 1 - \alpha_{P_{21}}$ through path P_{22} . Functions Q_1 , Q_2 and Q_3 are learned from previous measurements following the procedure described in Section 4.4. Then, in order to find the optimum forwarding decision for a particular combination of the considered traffic flows, we have to find the optimum values of $\alpha_{P_{si}}$ which lead us to the minimum network congestion. The proposed framework applied to this particular case leads us to the following optimization problem:

$$\begin{aligned}
 & \underset{\alpha_{P_{11}}, \alpha_{P_{21}}, \alpha_{P_{12}}, \alpha_{P_{22}}}{\text{minimize}} && Q_1 + Q_2 + Q_3 \\
 & \text{subject to:} && \\
 & && \alpha_{P_{11}} + \alpha_{P_{12}} = 1, \\
 & && \alpha_{P_{21}} + \alpha_{P_{22}} = 1, \\
 & && \alpha_{P_{11}}, \alpha_{P_{21}}, \alpha_{P_{12}}, \alpha_{P_{22}} \geq 0.
 \end{aligned} \tag{5.10}$$

Then, in order to update $\alpha_{P_{11}}$ (for $\alpha_{P_{21}}$ is analogous) we have to use the following equations:

$$\hat{\alpha}_{P_{11}}^{t+\Delta T} = \left[\alpha_{P_{11}}^t - \gamma \left(\frac{\partial Q_1}{\partial \rho_{11}} - \frac{\partial Q_2}{\partial \rho_{22}} - \frac{\partial Q_3}{\partial \rho_{32}} \right) \right]^+ \tag{5.11}$$

$$\alpha_{P_{11}}^{t+\Delta T} = \min \left(\hat{\alpha}_{P_{11}}^{t+\Delta T}, 1 \right) \tag{5.12}$$

In order to choose the most appropriate function Q_l for each link we followed the measurement-based method described in Section 4.4. In the learning phase, to

5.4. Simulation Experiments

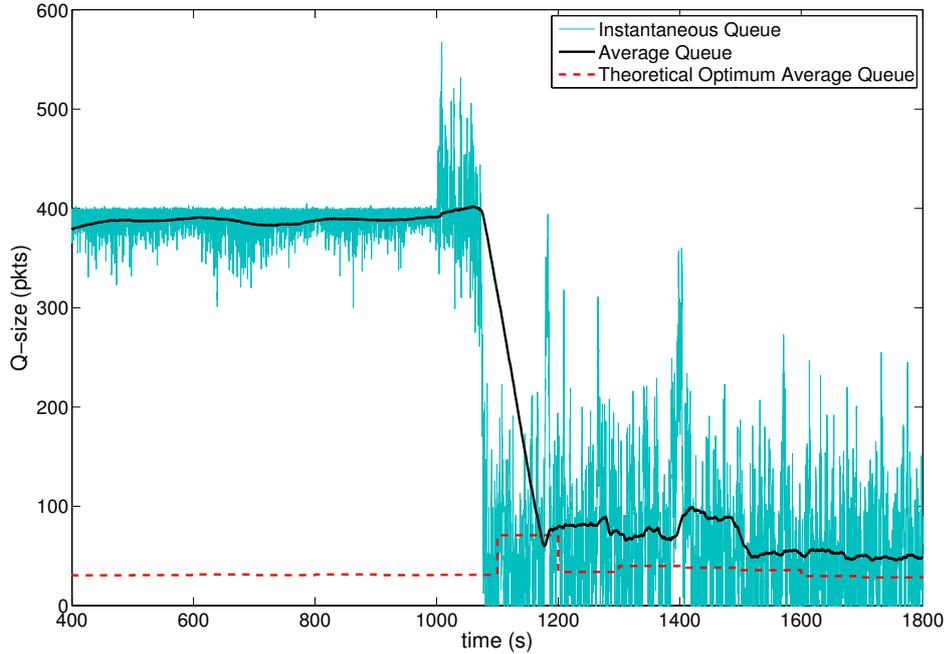


Figure 5.4: Total queue for the 3-nodes topology symmetric case.

Method	UDP flows Delay (ms)	UDP flows Jitter (ms)	TCP flows Goodput (Mbps)
MQLLB	14.6	6.7	15.2
802.11s	17.5	7.6	13.9
static routing	14.3	7.1	15.4

Table 5.2: Performance metrics for the 3-nodes topology symmetric case.

generate the training data we used simulations with different traffic distribution coefficients $\alpha_{P_{si}}$, uniformly covering all the possible values. Then, to calculate the partial derivatives of each link queue Q_l we used the learned functions in order to periodically update the $\alpha_{P_{si}}$.

Now, we will present the simulation results using the presented framework for two different traffic loads: symmetric and asymmetric cases. First we will show a symmetric example where traffic loads were $d_1 = d_2 = 13$ Mbps. In Figure 5.4 we can see the evolution during the simulation of the total queue size (expressed in packets), which corresponds to the sum of all interfaces queues in the network. We present the comparison of the instantaneous queue size and the 100-seconds average with the theoretical optimum queue length, which is calculated from the learned model and the traffic average measures. We show from time $t = 400$ s, when we have already reached steady state, starting with $\alpha_{P_{11}} = 1$ and $\alpha_{P_{21}} = 1$ (i.e. both flows forwarded through link 1), which causes saturation at link 1 and we start using MQLLB at time $t = 1000$ s. Concerning the performance metrics, the results

Chapter 5. Minimum Queue Length Load-Balancing

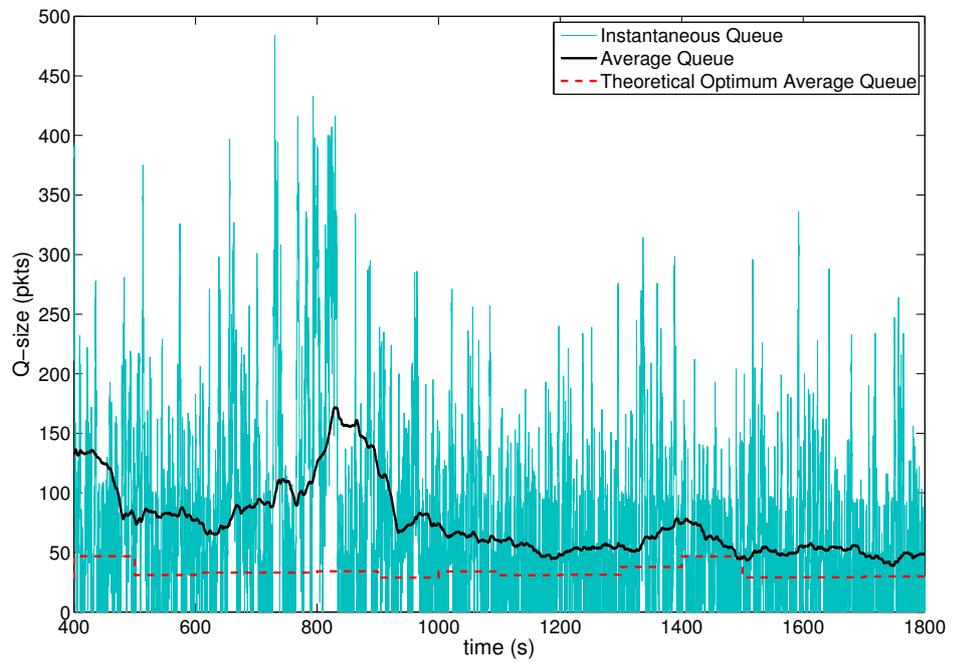
Method	UDP flows Delay (ms)	UDP flows Jitter (ms)	TCP flows Goodput (Mbps)
MQLLB	14.3	6.6	15.5
802.11s	43.1	8.9	9.6
static routing	35.1	8.6	11.2

Table 5.3: Performance metrics for the 3-nodes topology asymmetric case.

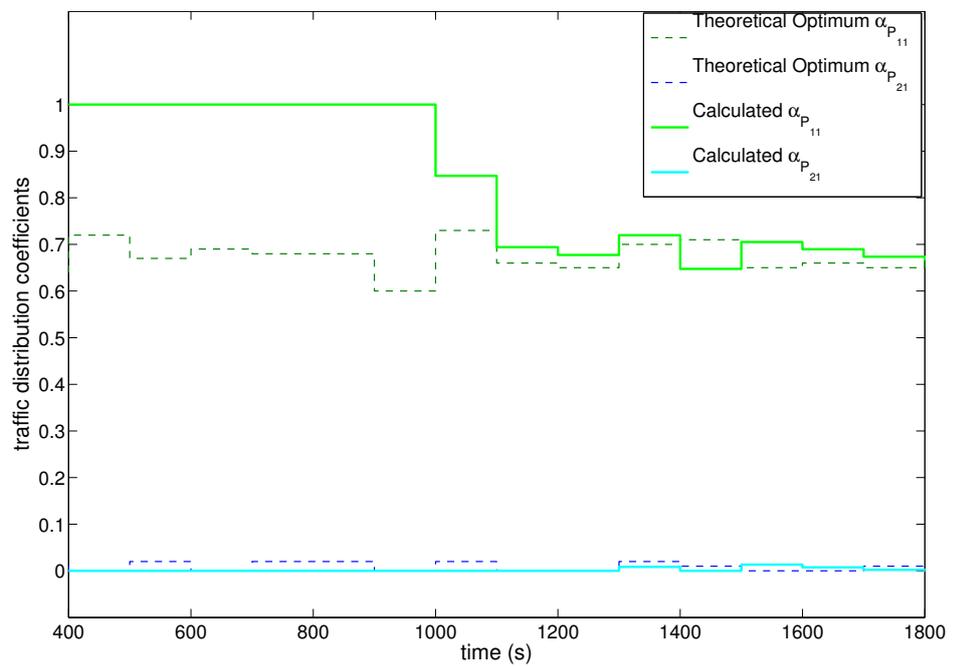
are summarized in Table 5.2. We can see that none of the metrics show significant differences between the three alternatives. It is clear that with symmetric traffic as in this case, static routing through shortest paths is a good alternative, as the results reflect. Notice that 802.11s presents slightly worse results, something which will be analyzed more deeply in the next simulations.

The other example with the three-node topology corresponds to an asymmetric case, where traffic loads were $d_1 = 20$ Mbps, $d_2 = 5$ Mbps. We started the simulation with $\alpha_{P_{11}} = 1$ and $\alpha_{P_{21}} = 0$ (i.e. only the one-hop path for each OD pair). In Figure 5.5(a) we can see the total queue size evolution from time $t = 400$ s. We started the operation of MQLLB at time $t = 1000$ s and as we can see the average queue size goes down which means the traffic load in the network is better balanced. Figure 5.5(b) shows the traffic distribution coefficients evolution. Notice that at time $t = 1100$ s, when the second update round happens, we already reached the optimum load-balancing. Looking at performance metrics shown in Table 5.3, we can see that the difference is clear in favour of MQLLB in this case where we have asymmetric traffic. As expected, for the asymmetric example we have an important improvement in the network performance due to the load-balancing mechanism.

5.4. Simulation Experiments



(a) Total queue size evolution.



(b) Traffic distribution coefficients evolution.

Figure 5.5: 3-nodes topology asymmetric case simulation.

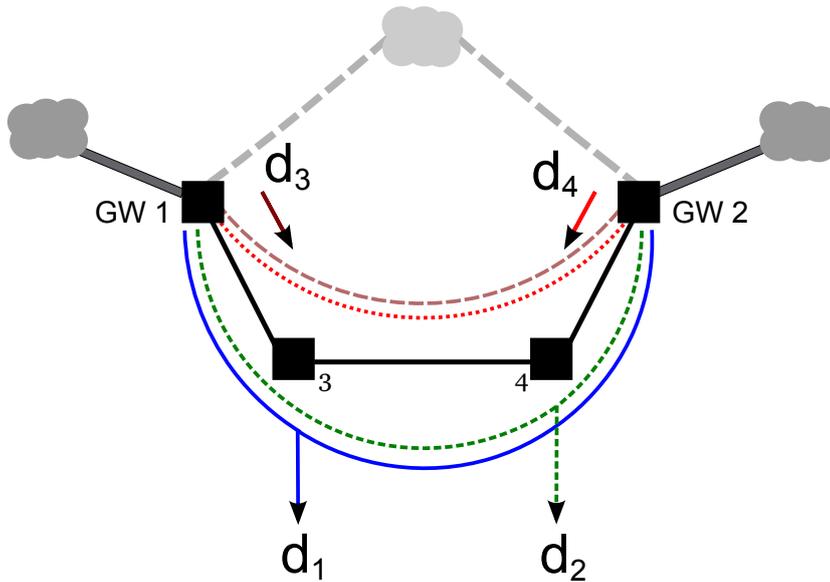


Figure 5.6: Gateway selection problem.

5.4.2. Gateway Selection Problem

In this subsection we will analyze an example corresponding to the gateway selection scenario shown in Figure 5.6. We will show that it is possible to solve this problem under the proposed framework, treated as an equivalent multipath forwarding one. In this topology we considered downlink flows to nodes 3 and 4, with demands d_1 and d_2 respectively, which can be distributed between the two gateways GW 1 and GW 2. Notice that both gateways could be considered as the same traffic origin (Internet). We can think this origin as a *super* node, connected to both gateways by links with infinite capacity (shown with dashed lines in Figure 5.6). Then, the gateway selection problem turns into a multipath forwarding problem, where we have to decide which portion of traffic demands d_1 and d_2 to forward through each of the possible paths from the *super* node (Internet), which is equivalent to decide which portion of traffic to route from each gateway.

In this example, we also considered inter-gateways flows from node 1 to node 2 and viceversa, with demands d_3 (from 1 to 2) and d_4 (from 2 to 1) respectively. These traffic flows may exist due to mobile hosts directly attached to one gateway that access resources allocated at servers in the other gateway. There is only one possible path for this flows, so there is no forwarding decision to take for those OD pairs. However, they affect the amount of traffic on each link, which leads the network to a different load condition than the one without inter-gateways flows. It is a desirable property of the algorithm that the existence of those inter-gateways flows does not affect the forwarding decision of the other flows.

We will analyze an asymmetric simulation example where traffic loads are $d_1 = 15$ Mbps, $d_2 = 5$ Mbps and $d_3 = d_4 = 3$ Mbps. In this case the network

5.4. Simulation Experiments

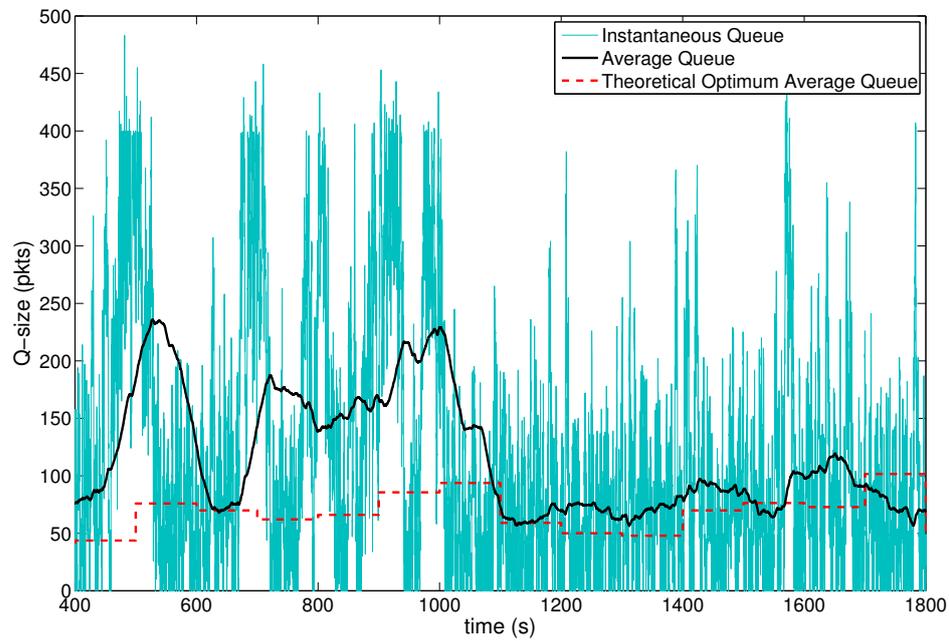
Method	UDP flows Delay (ms)	UDP flows Jitter (ms)	TCP flows Goodput (Mbps)
MQLLB	21.4	8.1	10.8
static routing	48.4	9.2	8.3

Table 5.4: Performance metrics for the gateway selection asymmetric case.

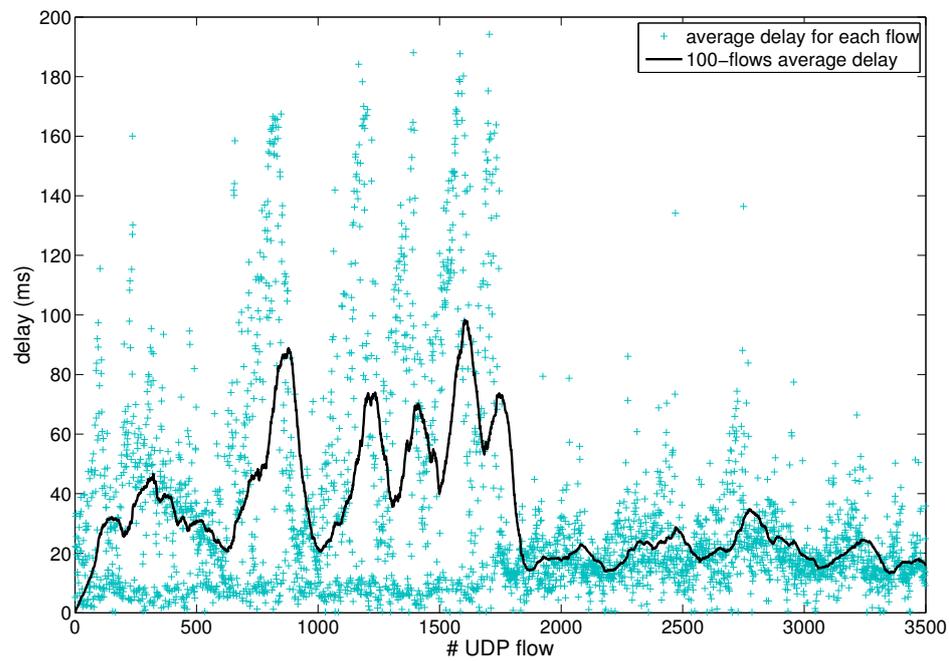
started operating with shortest path routing with hop count as routing metric (i.e. d_1 through GW 1 and d_2 through GW 2). The heavy traffic load from GW 1 to node 3 produces congestion in that link, which is visible in Figure 5.7(a) where the total average queue evolution is shown from $t = 400$ s, when we have already reached steady state. The operation of MQLLB starts at $t = 1000$ s and reached convergence at $t = 1200$ s. The final total average queue length as we reached convergence is 79 packets, which is almost 50 % smaller than before starting MQLLB where it was 154 packets (with peaks up to 235). In Figure 5.7(b) we show the average packet delay analysis for UDP flows. Please note that the x-axis does not correspond to time but to the flow index. It is clear that after MQLLB starts there is an important improvement with a smaller average delay. Performance metrics are summarized in Table 5.4, where we compare the results of MQLLB with static routing through the nearest gateway (802.11s was not considered in this gateway selection example). It is clear the advantage of using MQLLB in this case, particularly noticeable in the UDP flows delay with an improvement of more than 50 %.

For the gateway selection problem there is an important issue to solve for a real-world implementation. For the downlink case (traffic coming from the Internet) we cannot perform path selection at the ingress routers (i.e. the gateways) since we are distributing traffic between paths that do not share the same origin node. A simple alternative to solve this issue is to make gateway selection with client granularity. In this case, the routers which are directly connected to mobile hosts may decide the proper gateway for each client. In order to improve the performance of this approach these routers could monitor each client traffic demand. Thus, the optimization process could use a client granularity but including the client demand information, which allows a better traffic forwarding update at each step.

Chapter 5. Minimum Queue Length Load-Balancing



(a) Total queue size evolution.



(b) Average packet delay for UDP flows.

Figure 5.7: Gateway selection with asymmetric traffic loads.

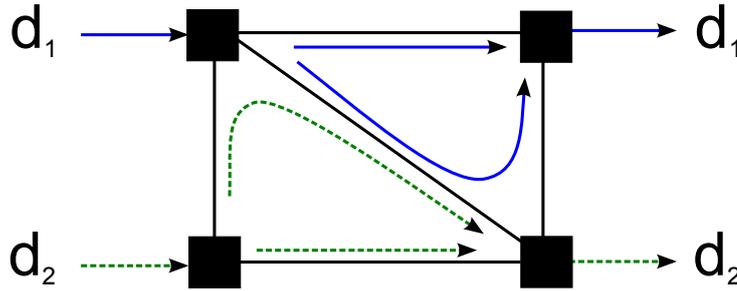


Figure 5.8: 4-node topology multipath forwarding example, situation 1.

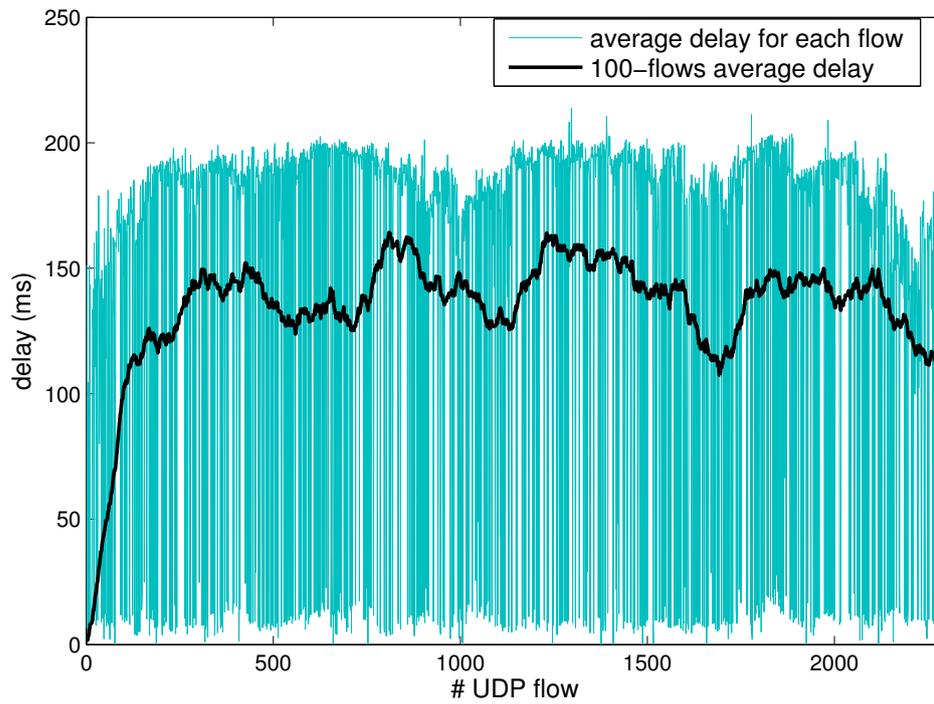
Method	UDP flows Delay (ms)	UDP flows Jitter (ms)	TCP flows Goodput (Mbps)
MQLLB	19.0	6.7	15.4
802.11s	141.3	8.0	4.8
static routing	104.0	8.4	6.8

Table 5.5: Performance metrics for 4-node topology example, situation 1.

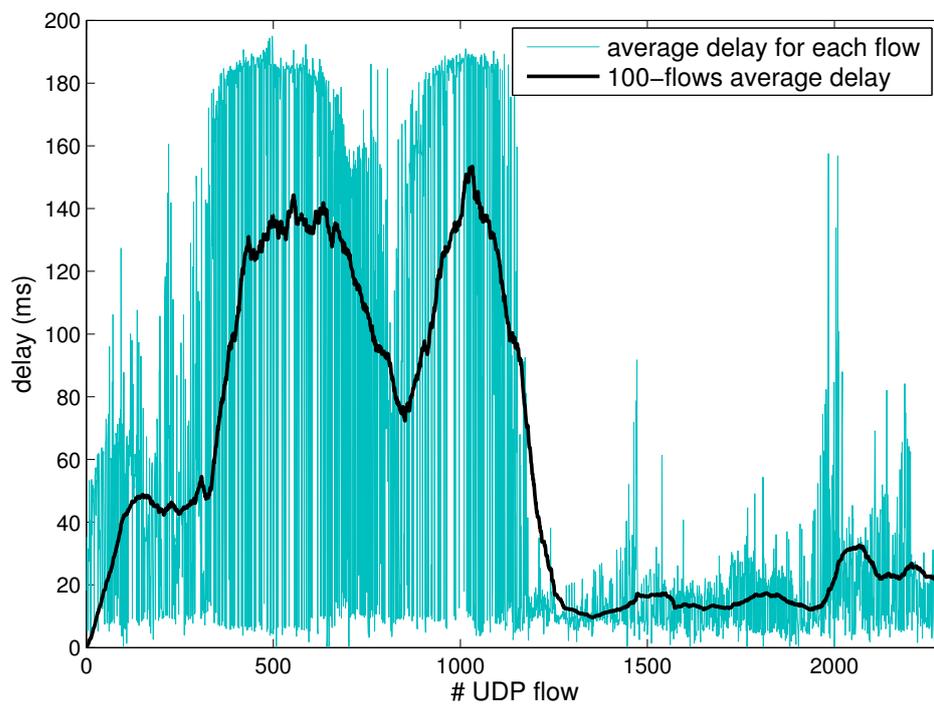
5.4.3. Multipath Forwarding: 4-nodes Topology

The next example corresponds to a four nodes topology with five 802.11 links and two OD pairs (see Figures 5.8 and 5.11). In this scenario we have three possible paths for each OD pair, each of them of distance 1, 2 and 3 links. We will consider only the two shortest paths for each one, so we have to decide for each OD pair, how much traffic to forward on each route. As we said before, the possible paths for each OD pair are defined by the network topology, but we can decide not to use any given path by configuration, because we want to simplify the network operation or just avoid the usage of a particular path. We will consider two different situations, both of them with asymmetric traffic demands, but the difference between them is how the paths share the different links.

First, we will analyze the situation shown in Figure 5.8, where both flows are from left to right, so links are shared by flows in the same direction. We simulated the scenario with $d_1 = 25$ Mbps and $d_2 = 10$ Mbps and compared the performance of MQLLB with 802.11s. Both simulations have a total duration of 2500s, in one case beginning with static routing using only the single-hop paths and MQLLB starting at time 500s and in the other case using 802.11s during all the simulation. The different performance metrics analyzed show a clear advantage of MQLLB over 802.11s and static routing. The results are summarized in Table 5.5, where we can see an improvement of more than 70% in the average delay for UDP flows and more than 100% in the average goodput for TCP flows. In Figures 5.9(a) and 5.9(b) we show the average packet delay evolution for UDP flows in time order during the first 1000s of the simulations. Similarly, in Figures 5.10(a) and 5.10(b) we show the average goodput evolution for TCP flows. In both cases it is clear the moment when MQLLB starts the operation (at 500s), which is reflected on



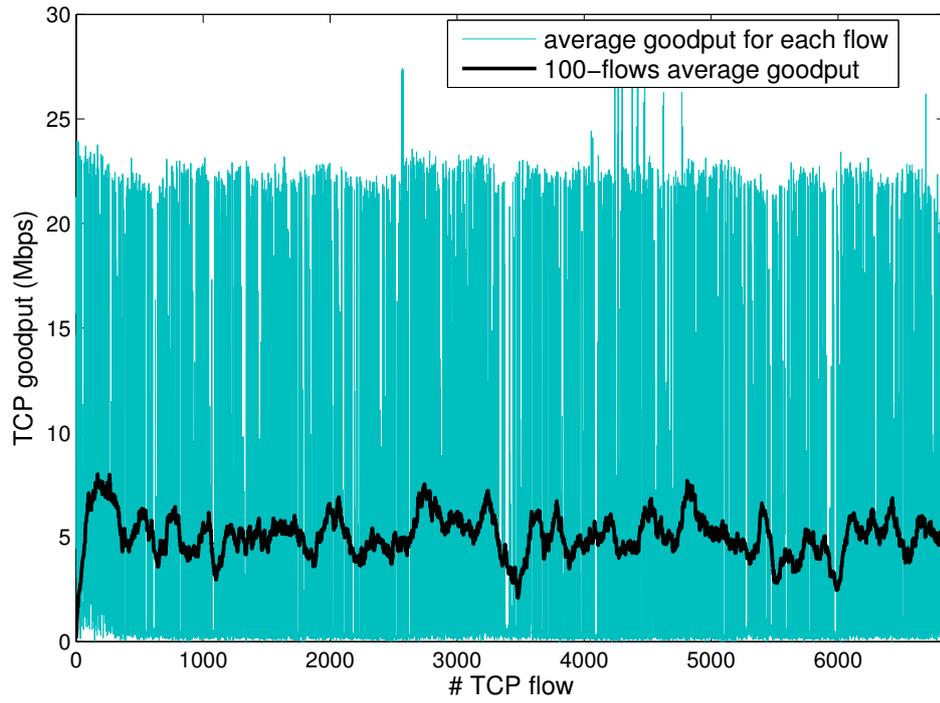
(a) Simulation with 802.11s.



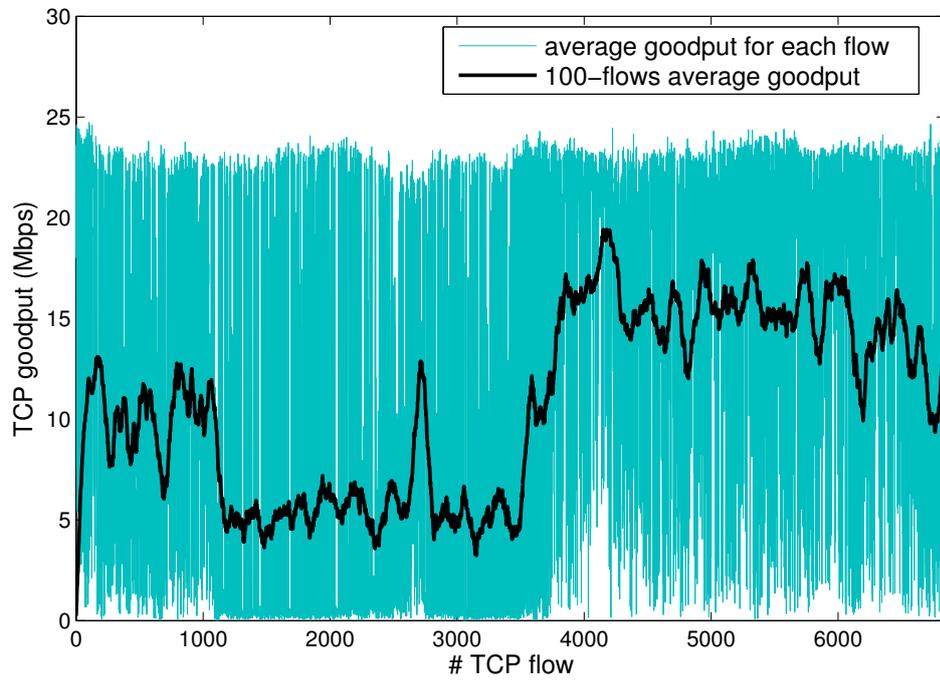
(b) Simulation with static routing and MQLLB.

Figure 5.9: UDP flows average delay analysis for 4-node topology example, situation 1.

5.4. Simulation Experiments



(a) Simulation with 802.11s.



(b) Simulation with static routing and MQLLB.

Figure 5.10: TCP flows average goodput analysis for 4-node topology example, situation 1.

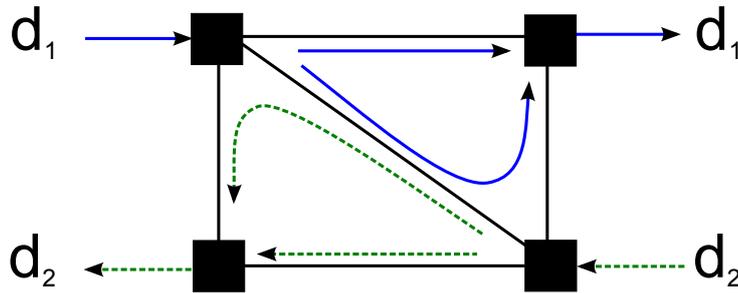


Figure 5.11: 4-node topology multipath forwarding example, situation 2.

the network performance with a smaller average delay for UDP flows and a larger average goodput for TCP flows.

The other considered situation is shown in Figure 5.11. The traffic loads are the same than before ($d_1 = 25$ Mbps and $d_2 = 10$ Mbps), but now d_1 is from left to right and d_2 from right to left, so links are shared by flows in the opposite direction. The results, which are summarized in Table 5.6, are quite similar to the previous situation, with significant improvements in all the analyzed performance metrics in favour of MQLLB. The purpose of this example is to show the ability of the proposed framework to cope with different link sharing situations, with traffic demands sharing the links both in the same direction or in opposite directions.

To explain the improvements of using an scheme like MQLLB instead of 802.11s, we must first note the advantage of considering multiple paths for each origin-destination pair, which allows a better adaptation to the particular traffic conditions. This fact is particularly clear when we analyze asymmetric traffic situations like the one of the examples. Second, we must consider the problems of using a metric that reflects the dynamics of each link at each moment as the *airtime* used by 802.11s. As studied in [6] routing oscillations may happen because of the dynamics of the different links metric. When more traffic is forwarded through a link, the metric is degraded, which causes that quickly we can find an unloaded link with a better metric. This fact causes that the node will change the selected path and it will start forwarding the traffic on the other link. The new selected link will suffer the same metric degradation that the other one had before, so the node will change the selected path again. This phenomenon is repeated indefinitely generating an oscillation of the chosen path. This phenomenon was also noticed in [7]

Method	UDP flows Delay (ms)	UDP flows Jitter (ms)	TCP flows Goodput (Mbps)
MQLLB	25.9	7.2	13.9
802.11s	141.6	8.4	4.7
static routing	101.9	8.2	6.8

Table 5.6: Performance metrics for 4-node topology example, situation 2.

5.4. Simulation Experiments

where it was called “ping-pong” effect, and the results reported in that work were similar with the ETX metric. This fact explains the bad performance of 802.11s, which is even worse than the one for static routing through one-hop paths in this examples. The proposed MQLLB uses average measurements to reflect the dynamics which allows a quick adaptation to traffic changes but ensuring an stable operation for steady state situations.

5.4.4. Gateway Selection: 25-nodes Topology

Finally, we present a gateway selection scenario in a 25-node topology to take a look into scalability and convergence of the proposed framework. The nodes are disposed in a 5 x 5 uniform square grid with side 500m and links are established between the closest nodes, all with a 100m distance. We call each node n_{ij} using matrix notation and we have two gateways corresponding to nodes n_{15} and n_{51} (top right and bottom left of the square respectively). We have a routing protocol (OSPF) which establishes routes between every pair of nodes, so, as we have two gateways, each node has two possibles routes to the Internet. We will use the proposed method to find the proper traffic distribution between gateways for each node, which in this example is called α_{ij} for the corresponding node ij .

The traffic considered in this example is all downlink (from the gateways to the other nodes) and it was generated with the same characteristics as in previous examples. We chose as a convention that $\alpha_{ij} = 1$ means that all the traffic for node ij comes from gateway n_{15} and if $\alpha_{ij} = 0$ all the traffic comes from gateway n_{51} . In the simulation, we started with $\alpha_{ij} = 0,5$ for all nodes, which corresponds to half of the traffic coming from each gateway for all of them. The load values used in the simulation were 5Mbps for nodes $\{n_{11}, n_{12}, n_{13}, n_{14}, n_{52}, n_{53}, n_{54}, n_{55}\}$ and 2.5Mbps for the rest of the nodes.

We enabled the operation of MQLLB at $t = 300s$. In Figure 5.12 we show the evolution of the traffic distribution α_{ij} for each node, while for the gateways we show the total traffic load that comes from each of them. We can see that all the nodes which are at the same distance from each gateway (nodes n_{ii} , at 4-hop distance to gateways) remained with $\alpha_{ii} = 0,5$ during all the simulation. On the other hand, nodes which are closer to gateway n_{15} changed to $\alpha_{ij} = 1$ while the ones closer to gateway n_{51} changed to $\alpha_{ij} = 0$. This means that nodes with one gateway closer than the other, change the traffic distribution in order to receive all the traffic from the closest gateway. Taking into account the convergence, we can see that nodes which are closer to gateways converge in less optimization steps than the others. For example, looking at gateway n_{51} we can see that nodes at one hop distance converge in one step, while nodes at two hop distance take two iterations to converge and finally nodes at three hop distance take three iterations to reach convergence. In Figure 5.13 we show the evolution of the total average queue, where we can appreciate its steep descent when MQLLB starts the operation.

Chapter 5. Minimum Queue Length Load-Balancing

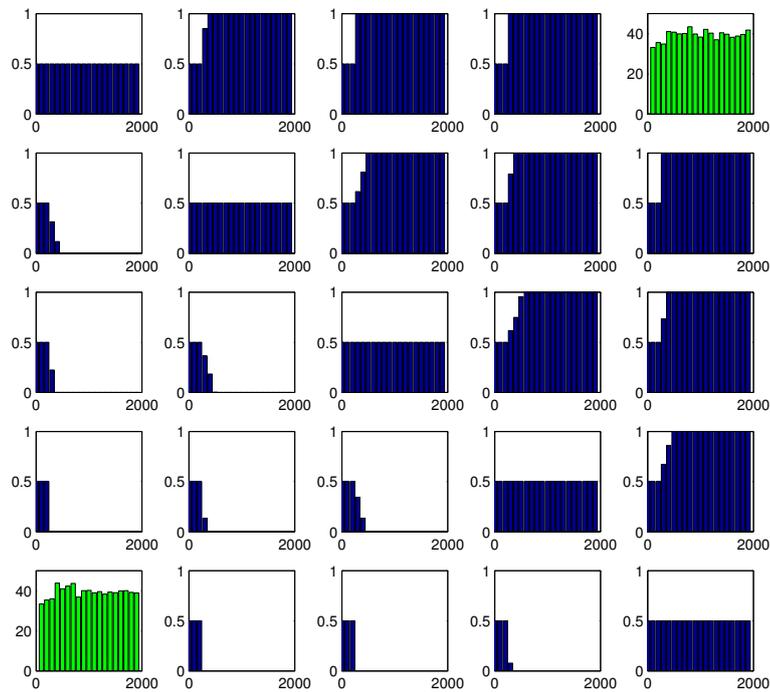


Figure 5.12: Traffic distribution and aggregate load at each GW as a function of time (subplot ij corresponds to node n_{ij}).

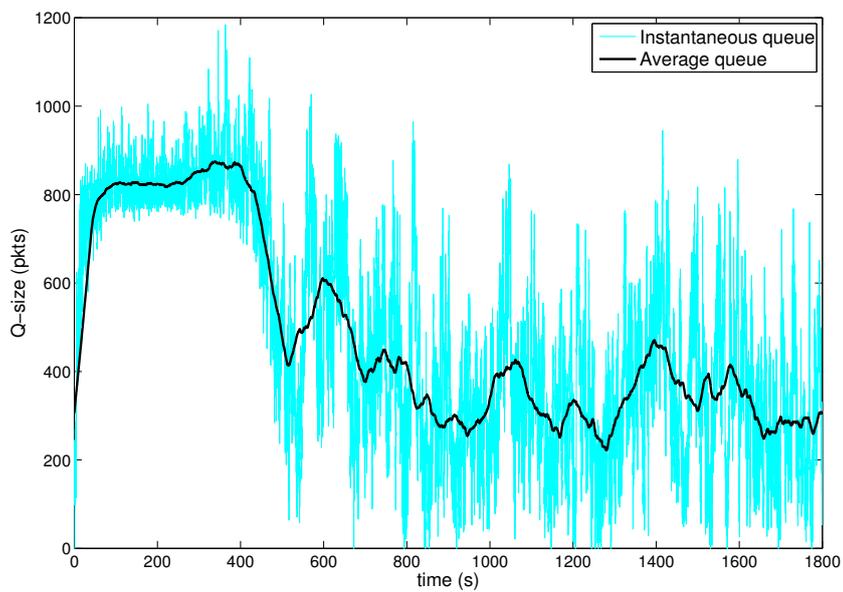


Figure 5.13: Total queue size evolution.

Conclusions of Part II

In this part of the thesis we addressed the problem of finding a suitable routing and forwarding scheme in a wireless mesh network. After developing a network model based on measurements, the forwarding is posed as an optimization problem. The achieved solution optimally distributes all end-to-end traffic over all possible paths for each pair origin-destination. With this dynamic multipath forwarding scheme, the algorithm enables load-balancing and conducts the network to operate at the minimum average congestion. The proposed framework also allows to solve the gateway selection problem in a wireless mesh network.

The problem formulation is based on learning from measurements the average queue length function for each wireless link in the network. This statistical model is constructed from actual network measurements, ensuring that the estimation is adjusted in the widest possible way to reality. Then, a classical optimization method is applied, in order to reach the minimum average queue length in the network. The proper evolution and convergence of the proposed method was verified by our packet-level simulations over several canonical topologies which served as a proof of concept.

We further analyzed the simulations taking several flow-level performance metrics as average delay and jitter for UDP traffic and average goodput for TCP traffic. With this metrics we studied the performance of the proposed MQLLB method compared with the IEEE 802.11s standard. The results show a clear advantage of MQLLB against a dynamic metric routing method like the one used by 802.11s. In all the simulations, independently of the topology size, we observed a quick adaptation of MQLLB to traffic changes and also an stable operation, avoiding the routing oscillations of 802.11s, already noticed before by [6, 7].

Among the points that could be studied in more detail in the future, we should perform the learning phase with real data, which includes the non-zero channel error rate, typical of a real-world wireless link. All the simulations presented in this work are done with synthetic traffic, so our analysis could be extended using real traffic data. It would also be very interesting to perform a statistical analysis of the behavior of the mean queue size with respect to the load. A possible objective would be to know how often does the regression function change over time (i.e. answer the question of whether the mean queue size function changes over time, and how often it does).

Another aspect that could be addressed in the future is the implementation of the proposed framework in a real-world network, which was briefly discussed in

this work. One possible way is to explore the adaptation of a recent MPLS-based routing scheme for WMNs [95] to our proposal. A testbed deployment would be useful for enhancing the algorithm and detecting real-world driven problems that need to be solved. An interesting point which could be more profoundly studied in the future is the optimization phase. This problem could be solved by several different methods and was not analyzed in the present work. Finally, the proposed framework was developed for a link disjoint WMN, so it could be extended to cope with scenarios that have not only point to point links but also point to multipoint links.

Part III

Spectrum Assignment in Cognitive Networks

Chapter 6

Spectrum Allocation in Cognitive Radio Networks

The third part of this thesis is dedicated to another resource allocation problem in wireless mesh networks, in this case the spectrum assignment. We address the analysis of optimum spectrum allocation mechanisms, based on the novel cognitive radio networks paradigm. A problem that is envisioned in the near future is the spectrum scarcity, which could be a serious threat to cope with the ever increasing demand. One of the possibilities that has emerged to solve this problem is to allow secondary assignments in licensed bands, by means of cognitive radio equipment.

The goal is to find a dynamic mechanism that conducts to the optimum spectrum allocation, given the physical medium conditions and the required traffic demand at each link. The novel robust method proposed meets these objectives, and also could be implemented with a distributed solution. The development is based on the assumptions validated in Chapter 2 and Chapter 3, that is to say that we assume that it is possible to estimate the traffic demand and predict the effective capacity of each link from the physical layer measurements.

The rest of Part III is structured as follows. In the next section we present the previous related work and highlight some recent papers. In Section 6.3 we introduce most of the notation used in this part of the thesis and the spectrum allocation problem model developed. The formulation results in a stochastic optimization problem, so Chapter 7 is entirely dedicated to review this topic. We concentrate on chance-constrained programming, with focus on a novel technique to find the equivalent deterministic optimization problem which leads to a robust solution. This part of the thesis closes with Chapter 8, where we introduce the proposed robust approach for spectrum allocation in cognitive radio multihop networks. The chapter continues with Section 8.2, where we describe the network architecture that enables the implementation of the proposed scheme in a decentralized way. Finally, in Section 8.3 we present the simulation experiments and performance comparison, while conclusions are discussed in a separate chapter at the end of this part.

6.1. Introduction

Over the last decade, we have seen an explosive growth in the deployment of wireless networks in unlicensed frequency bands, mainly driven by the great success of the IEEE 802.11 standard. During the same period of time, we have also witnessed the highest growth in the traffic load carried over wireless networks [63] and forecasts indicate that this growth will continue [113]. Moreover, the user density is also increasing, resulting in crowded scenarios where the technology is reaching its limits (e.g. classrooms, large conferences, shopping centers or sport events [59]).

Besides these most common scenarios, where we only have a wireless last hop, requirements also increase for the wireless transport networks we found today, also using 802.11-based technology in unlicensed bands. This is the case of the typical wireless mesh network (WMN) solution previously introduced in this thesis (see Figures 1.1 and 4.1), with the real world case scenario from Plan Ceibal described in Chapter 2, which provides Internet access for schools located in rural or suburban areas. In that case the problem is not about user density, as we only have point to point or point to multipoint links between a few nodes. Instead, we have higher throughput requirements, because we are talking about the network core. While standards are still evolving, achieving increasingly higher spectral efficiency, we may soon be faced with spectrum scarcity issues to properly cope with traffic demands. Regulators have taken note about this fact and some proposals already exist to extend the available spectrum [114].

Leaving aside traditional spectrum allocation, a new type of spectrum assignment has emerged some years ago: the so-called cognitive radio paradigm [8]. The main idea is to have two types of users; licensed or primary users (PUs from now on), which have the preferential right to use the band; and unlicensed or secondary users (SUs from now on), which can use the band only in the absence of the PUs. This type of spectrum allocation contributes to a more efficient use compared to traditional static assignments, as testified by some recent FCC rulings [52]. Although adoption is not yet massive, much industrial and academic efforts have been dedicated to this kind of technology. For instance, the IEEE 802.22 standard [50] was approved in 2011, which defines a Wireless Regional Area Network (WRAN) based on cognitive radio. Another industrial effort is the 802.11af amendment to enable the operation of the standard in TV bands, which has been recently published [51].

On the other hand, the development of cognitive radio equipment is still immature, particularly concerning sensing tasks to detect PUs, so the first solutions being deployed are based on databases queries to get the information about the available spectrum [115]. Some major providers such as Google are already authorized in the US to give such spectrum database service [116]. Everything suggests that in the short to medium term dynamic spectrum allocation will expand, and in a few years we will probably have several standards operating under this paradigm. This enables new possibilities for the development of radio communications equipment, which added to the advances in software defined radio (SDR) techno-

logies, may cause a significant change in the world of wireless communications we know so far.

While much research has been recently dedicated to cognitive radio networks and dynamic spectrum allocation, most of the works have mainly focused on the case where there are only licensed bands available [49]. In that case, unlicensed devices can only operate as SUs in the absence of PUs, greatly limiting their possibilities. We believe it is very complex to develop a useful solution in such scenario with high throughput requirements. Several issues arise working only with licensed bands, for example, a control channel should be available all the time in order to coordinate the communication, which might be not easy to ensure without any guaranteed frequency band to use. Moreover, it is possible to have circumstances under which the available spectrum is not sufficient to meet the throughput requirements, as the available capacity strongly depends on the PUs' dynamics. In this thesis we work in a mixed licensed and unlicensed scenario, which we believe is more appropriate to support high throughput requirements. This solution has not been deeply explored yet in the literature and we think it is the most suitable model for the equipment and regulations that we may have during the coming years.

Our focus is centered on the dynamic spectrum assignment in a WMN. That is to say, we will study possible methods to decide which frequency bands may be used by the network devices at any given time. It is worth to highlight that such an assignment means that the bands are *available* for the devices, and are not necessarily used. With this in mind, the natural question that arises is to what purpose this assignment should be performed [24]. In our particular context, examples include minimizing the number of licensed bands assigned [117] or maximizing the user's utility (as a function of the mean rate) [118] without exceeding a maximum interference threshold to other networks.

However, in the context of a cognitive WMN, we argue that the most natural objective would be to provide a lower bound to the resulting throughput in each link. As introduced in chapter 2, the traffic demand can be estimated from measurements of the live network, so we could infer the necessary capacity on each link to cope with the demand. Besides, the variations in traffic dynamics between spectrum assignment intervals could be solved with an algorithm as the one presented in Chapter 5, which adapts the routing and forwarding to the underlying links' capacity. The purpose of the spectrum allocation should be thus to ensure a certain effective capacity for each link, independently of the channel conditions and the PU's activity.

The other challenge that these systems pose is the timescale at which the assignment should be performed. One possibility is to re-assign (and thus re-optimize) every time a band is used or abandoned by PUs, or if significant changes in channel conditions occur. Although this event-driven solution will lead the system to operate with the optimal allocation all the time, it will typically result in a dramatically high signaling overhead. In this sense, we will assume, as many researchers, a periodic optimization every T time units, which leads us to a better performance tradeoff. However, T may include variations in PUs' activity. This

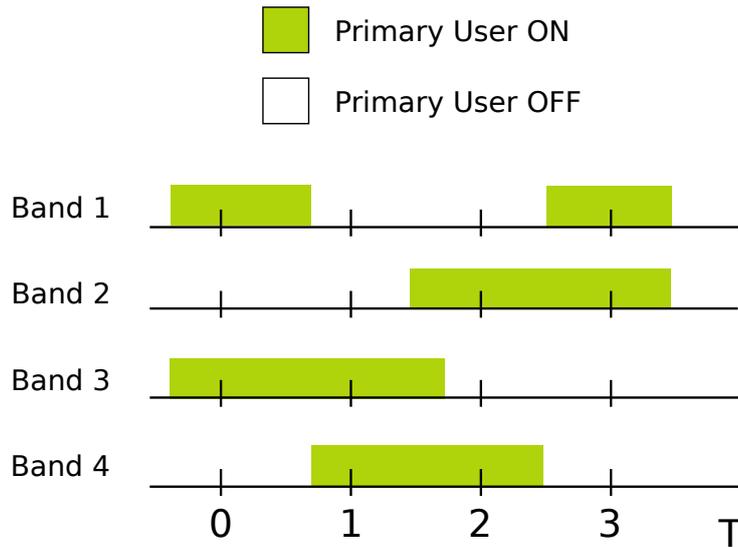


Figure 6.1: Spectrum occupancy example with four licensed bands.

fact implies that a licensed band assigned when the period starts might have to be abandoned, resulting in an effective capacity that is less than expected. In Figure 6.1 we present an example to clarify this situation. In it we have four licensed bands, with two of them available at the first spectrum assignment at time 0. During the interval between allocations, a PU starts using band 4, so it is no longer available. The problem occurs again in the second assignment, where bands 1 and 2 are available and the assignment is thus performed, but a PU occupies band 2 during the interval.

To address this issue, the most commonly used approach is to model the availability of licensed bands as random, and optimize the expected value of indicators such as interference or throughput, as discussed before. Although this means that in the long run the objective will be accomplished (e.g. the throughput will be maximized), at shorter timescales the resulting performance may be far from optimal. In contrast with previous works, we will present a frequency assignment scheme that provides the required throughput, which will hold with very high probability during the whole operating time. Naturally, such guarantee will require a certain degree of overprovisioning, but our simulations indicate that this is usually below 35 % of that required by an oracle that knows beforehand the PUs' activities. Moreover, the results show that simply considering an expected value approach leads us to a solution where the throughput requirement is not fulfilled more than 40 % of the time.

6.2. Related Work

More than a decade has already passed since the emergence of the cognitive radio networks (CRNs) paradigm, and a large amount of the research done in the

area during last years has been dedicated to spectrum assignment. An example of this is the number of papers that can be found as references in the broad survey by Tragos *et al.* [24]. However, as the authors state in the paper, there are still many issues and challenges to be solved, something which is also remarked in [119]. As we previously mentioned, most of the work so far is focused on a scenario with all licensed frequency bands, where cognitive nodes can only access to the spectrum as SUs while PUs are not present, discarding the use of unlicensed bands, available at any time. While this problem is still of interest for certain applications, such as delay-tolerant or sensor networks, it is not suitable for a transport mesh network, with high throughput and high availability requirements. In other cases, the spectrum allocation simply ignores the PUs, or just consider that SUs have a fixed set of available frequency channels, separated from the ones of the PUs.

This latter scenario reduces to the traditional spectrum allocation problem in a WMN, which has been the focus of several articles. In this problem, different variants arise, such as the number of radios per node, which can vary from a single radio per node [120,121], to the higher capacity multi radio case [122–124], which gives name to the multi-radio multi-channel (MR-MC) WMNs. Our work can be seen as an extension to this model, as we consider the same problem but under the paradigm of CRNs, which we believe should be the natural next step in the evolution of multihop wireless networks. Furthermore, we consider a novel robust approach, but in this case the uncertainty is not about the channel conditions [125,126], nor the traffic variations [127], but the PUs' activity. In particular, incorporating licensed bands generates a dynamic resources availability, so one of the requirements of the spectrum allocation is to be robust against such variations.

To the best of our knowledge, very few works have studied the resource allocation in a mixed licensed and unlicensed scenario. In [117] an opportunistic spectrum assignment is proposed in order to alleviate congestion in a WLAN environment. The problem is formulated as a binary linear program, where they seek to minimize the number of assigned bands without exceeding a maximum interference threshold. The proposal is limited to the allocation of a single frequency band for each access point, so channel aggregation is not considered, something already included in newer standards (802.11n and 802.11ac) and which is quite an important limitation in order to increase capacity when needed. A similar problem, but from the PUs' perspective, is studied in [118]. In that case the authors analyze the simultaneous use of both type of frequency bands by a mobile operator, in order to increase the capacity in a femtocell scenario.

We highlight the work in [128] where the authors studied a traffic engineering solution in the context of a multihop cognitive WMN. They considered the combined use of ISM bands and licensed bands in the absence of PUs, and also assumed nodes have cognitive sensing capabilities in order to exploit unused primary bands. The traffic engineering problem is formulated as a network utility maximization, which is solved with a stochastic primal-dual approach, without knowledge of the probability distribution of PUs' activity. The spectrum assignment is not treated directly as it is an underlying problem of the traffic engineering issue addressed in the paper, so they just assume the available spectrum for each link determines

its variable capacity. Our work is based on similar assumptions as the ones stated in [128], but we focus on the spectrum assignment problem. The main difference is that we consider a measurement-based approach where we estimate the probability distribution of PUs' activity, based on the nodes' cognitive sensing capabilities. In this work we thus take into account the PUs' activity, something which was not considered in many previous works, as stated in [24].

6.3. Network Model and Problem Formulation

In this thesis we study the spectrum allocation problem in a mixed licensed and unlicensed scenario. In the proposed scheme, devices operate always as unlicensed devices but in two types of frequency bands, licensed ones, where they are only allowed to operate when there is no presence of PUs, and unlicensed ones, where they can operate all the time. This offers greater flexibility to meet the requirements, given the scarcity of unlicensed spectrum. Furthermore, by having both type of bands, we simplify the protocol design complexity compared to solutions which only use licensed bands, as we can perform control communications through unlicensed bands, which are available all the time. To accomplish this goal we will impose that any possible assignment should include a minimum amount of unlicensed spectrum that guarantees a minimum capacity for control plane traffic (which we shall call w). This way we ensure the control plane connectivity between nodes, which makes possible the proper coordination for the use of the allocated frequency bands.

As in other previous works (e.g. [129]) we will assume that each node has a dedicated interface to enable cognitive sensing capabilities. By this mean, each node is able to keep a record of the PUs' activity on each licensed band. Besides, this interface is used to collect air measurement data, which are used to estimate the available capacity on each band, either licensed or unlicensed. In this part of the thesis we shall call effective capacity to the maximum achievable throughput in higher layers (i.e. network and transport layers), in the same way it was presented in chapter Chapter 3. This effective capacity depends on several factors such as channel conditions and other SUs' activity (devices from other networks that are not under our control), but it can be estimated passively through measurements [130] [131]. An example of how is it possible to infer the effective capacity from physical layer measurements was shown in Chapter 3.

We consider a solution where the assignment is performed every T time units and we will further assume that T is relatively small, so that an accurate estimation of each band's available capacity may be obtained using information from the previous interval. In this work we suppose that such estimation is exact, so as to focus only in the PUs' dynamics. We will also assume that devices can fully exploit the available spectrum (even disjoint available bands), using a PHY layer such as OFDM. We also assume there is a MAC layer mechanism in order to share the spectrum between nodes (e.g. 802.11 MAC layer).

6.3.1. Single Collision Domain

In this section we will focus on a single-domain spectrum assignment, that is to say, a network with a unique collision domain, corresponding to the case of just one point to point link between two nodes. In the next section we will present the model extension for a wireless mesh network (WMN) with multiple collision domains. Let $u = 1, \dots, U$, be the set of unlicensed frequency bands (i.e. no PUs, as in ISM bands). Let $b = 1, \dots, B_t$, be the set of licensed frequency bands (which are assigned to a PU) available at time t (i.e. PUs are not present).

We will note as $c_b(t)$ the effective capacity available on licensed frequency band b and $c_u(t)$ the effective capacity available on unlicensed frequency band u . This values should be estimated at each link, in a similar way to what we have done in Chapter 3, by means of the data collected at each node with the dedicated interface for sensing purposes. We define as spectrum assignment variables $\alpha_b(t)$ and $\alpha_u(t)$, which belong to $[0, 1]$, assuming partial band assignment is possible (e.g. via OFDMA or TDMA).

Now, we can define the total effective capacity assigned for the interval starting at T as:

$$C_{\text{eff}}(\boldsymbol{\alpha}^T) = \sum_{b=1}^B \alpha_b(T) c_b(T) h_b(T) + \sum_{u=1}^U \alpha_u(T) c_u(T) \quad (6.1)$$

where $h_b(T)$ is a real number in $[0, 1]$, according to how much time each licensed band was actually available during the interval. We will model h_b as a random variable, whose distribution will be learned from the previously observed dynamics. As we stated previously the objective is to provide a lower bound to the resulting throughput, so we will set this bound as a problem constraint, and we shall note it as d . This lower bound d is actually the minimum total capacity our system should have considering all nodes. We further define a cost function:

$$\mathcal{C}(\boldsymbol{\alpha}^t) = \mathcal{C}_{\text{lic}}(\alpha_1(t), \dots, \alpha_B(t)) + \mathcal{C}_{\text{unlic}}(\alpha_1(t), \dots, \alpha_U(t)) \quad (6.2)$$

The cost functions $\mathcal{C}_{\text{lic}}()$ and $\mathcal{C}_{\text{unlic}}()$ allow us to give different weights for each band, depending on the desired spectrum allocation goal. For example, it is possible to have different costs depending if the band corresponds to a higher or lower frequency, which may imply different transmission power requirements.

After all the stated assumptions, definitions and goals, we can now define an optimization problem which will lead us to the assignment algorithm for the single domain case. This problem should be solved periodically, so we will omit the time index from now on for a matter of clarity. That is to say, each time T we should

strive at solving the following problem:

$$\begin{aligned}
& \min_{\boldsymbol{\alpha}} \mathcal{C}(\boldsymbol{\alpha}), \\
& \text{s.t.} \quad \sum_{b=1}^B \alpha_b c_b h_b + \sum_{u=1}^U \alpha_u c_u \geq d, \\
& \quad \sum_{u=1}^U \alpha_u c_u \geq w, \\
& \quad \alpha_b \in [0, 1], b = 1, \dots, B, \\
& \quad \alpha_u \in [0, 1], u = 1, \dots, U.
\end{aligned} \tag{6.3}$$

In Chapter 8 we will present different approaches to deal with this problem. Our proposal, which we believe that fits best to the problem posed, is to use a probabilistic constraint, and find a suitable deterministic equivalent, using the technique presented in the next chapter. Thus it is possible to find a robust spectrum allocation, knowing only the mean and variance of the random data.

6.3.2. Model Extension for a Wireless Mesh Network

Now, we extend the previous model to the case of a wireless mesh network (WMN) with L links, where in the general case we may have multiple overlapping collision domains. We will consider for the spectrum allocation only the wireless links in the core of the WMN, assuming that the last hop with end clients is in other non interfering frequency bands. In this case, we can reuse the frequency bands in different links, but to avoid interference, we have to constrain the assignment on each collision domain. Thus, we want to ensure that if a certain frequency band u or b is assigned to a certain link l , then the same band cannot be assigned simultaneously by other links in the same collision domain.

In order to define the collision domains, we will consider interference between links and not between nodes. The reason to do this is that normally when we have communication between nodes, even when the data flows in only one direction, we still have information flowing in the opposite direction (e.g. acknowledgements). So, we will consider that two links interfere with each other if any node of one link is in the same collision domain than any node of the other link. Thus, we first need to know the conflict graph of the WMN, which is an undirected graph, where each vertex represents a wireless link and we have an edge between every pair of links that interfere with each other. Then, to list all the collision domains (noted with q , from 1 to Q), we have to look for all the maximal cliques of the conflict graph. Once we have all the Q collision domains in the WMN, we can properly define a binary matrix \mathbf{A} to reach the necessary additional constraint:

$$\mathbf{A} \cdot \boldsymbol{\alpha}^T \leq \mathbf{1}_{Q \times (B+U)} \tag{6.4}$$

using the matrix notation for the spectrum assignment variables defined in Table 6.1.

6.3. Network Model and Problem Formulation

$\alpha_U \in [0, 1]^{U \times L}$	unlicensed spectrum allocation variables
$\alpha_B \in [0, 1]^{B \times L}$	licensed spectrum allocation variables
$C_U \in \mathbb{R}^{+U \times L}$	capacity for unlicensed spectrum bands
$C_B \in \mathbb{R}^{+B \times L}$	capacity for licensed spectrum bands
$H_B \in [0, 1]^{B \times L}$	PU's activity on licensed bands
$d \in \mathbb{R}^{+L}$	capacity required for data plane traffic
$w \in \mathbb{R}^{+L}$	capacity required for control plane traffic
$A \in \{0, 1\}^{Q \times L}$	link interference matrix (conflict graph)
$\alpha = [\alpha_B; \alpha_U]$	all spectrum allocation variables
$C_{lic}(\alpha_B) \in \mathbb{R}^+$	cost function for licensed spectrum
$C_{unlic}(\alpha_U) \in \mathbb{R}^+$	cost function for unlicensed spectrum
$\mathbf{1}_M$	Column vector of ones of size M

Table 6.1: Matrix Notation

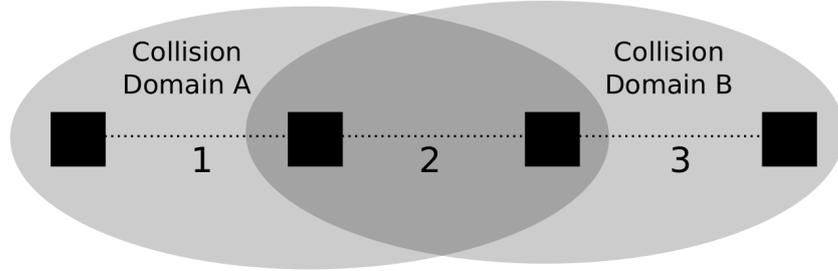


Figure 6.2: Wireless network with 3 links and 2 collision domains.

In Figure 6.2 we show an example topology with four nodes and three links, which will be later considered in the simulations presented in Chapter 8. The conflict graph of this network is given by the following adjacency matrix:

$$CG = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The nonzero entry $CG(1,2)$ indicates that the link number 1 and the link number 2 interfere with each other. The same happens between the link number 2 and the link number 3, given by the other nonzero entry in $CG(2,3)$. Notice that the matrix is symmetric which should be the typical case (if link A interferes link B , then is highly probable that link B also interferes link A). We have a total of two collision domains in the network, which correspond to the two maximal cliques of the conflict graph: $\{1,2\}$ and $\{2,3\}$. Then, the two resulting constraints for this example are:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_{B1}^\top & \alpha_{U1}^\top \\ \alpha_{B2}^\top & \alpha_{U2}^\top \\ \alpha_{B3}^\top & \alpha_{U3}^\top \end{pmatrix} \leq \begin{pmatrix} \mathbf{1}_{B+U} \\ \mathbf{1}_{B+U} \end{pmatrix}$$

Chapter 6. Spectrum Allocation in Cognitive Radio Networks

We are now able to define an optimization problem similar to the previous case of a single collision domain. We will omit again the time index for a matter of clarity. This way, the spectrum assignment in the WMN can be performed solving the following problem:

$$\begin{aligned}
 \min_{\boldsymbol{\alpha}} \quad & \mathcal{C}(\boldsymbol{\alpha}) = \mathcal{C}_{lic}(\boldsymbol{\alpha}_B) + \mathcal{C}_{unlic}(\boldsymbol{\alpha}_U), \\
 \text{s.t.} \quad & (\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{H}_B) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{d}, \\
 & (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{w}, \\
 & \mathbf{A} \cdot \boldsymbol{\alpha}^\top \leq \mathbf{1}_{Q \times (B+U)}, \\
 & \boldsymbol{\alpha}_B \in [0, 1]^{B \times L}, \\
 & \boldsymbol{\alpha}_U \in [0, 1]^{U \times L}.
 \end{aligned} \tag{6.5}$$

where \cdot stands for the common vector and matrix product operation and \odot stands for an element-wise matrix multiplication.

In the next chapter, after a brief review of chance-constrained programming, we will present a distributionally robust approach, which we argue best fits our needs to solve the stochastic optimization. Then, in Chapter 8 we will introduce the robust solution and a suitable distributed algorithm implementation.

Chapter 7

Chance Constrained Optimization

Many real-world engineering problems can be modeled as constrained optimization problems, where typically one or many of the constraints involve uncertain parameters. This chapter provides a quick review on chance-constrained optimization, and finally introduces to the method that we will use in the next chapter to develop the proposed robust spectrum allocation solution. The particular case of interest is a single-stage decision where the unknown data is random, with an uncertain probability distribution, but with known fixed values for the mean and covariance.

7.1. Introduction

Stochastic constrained optimization have found applications in many diverse contexts, such as network resource allocation problems (as our case of interest, for robust spectrum assignment), QoS management in multimedia networks, financial problems as portfolio optimization [132], electric power generation and optimal control of storage levels, and even diet and animal feed problems (included in Chapter 14 of [133]). In all cases, to find a suitable model, one should characterize the random parameters involved with a certain probability distribution. Two main approaches are generally available to address constrained optimization problems in presence of uncertainty: robust optimization and chance-constrained optimization. On the one hand, robust optimization is a deterministic paradigm where the solution that one looks for must simultaneously satisfy all the possible constraint sets corresponding to all the possible parameter realizations. For this reason, this paradigm is also known as max-min/min-max or worst-case approach. On the other hand, we have the chance-constrained optimization, where the deterministic equivalent corresponds to the computation of probabilistic constraints. In this case one seeks a solution that enforces the constraints up to a pre-defined amount of risk (ϵ) or security level ($1-\epsilon$).

Unfortunately, there is no standard way to select one option or the other, and both approaches may lead to a computationally intractable problem formulation. Thus, one may choose the most appropriate method for the problem at hand,

Chapter 7. Chance Constrained Optimization

which for our needs is chance-constrained optimization, as we will justify further in Chapter 8. In our case, the problem addressed is a single-stage decision problem, so we will consider in all cases a single stochastic optimization problem. A single-stage decision means that there is only one moment to make a decision from the available data, which differs for example from a two-stage decision, where a second stage with new information available enables to improve the previous decision. In the rest of this chapter we will present a generic stochastic problem and apply the classic chance-constrained programming approach introduced in [134]. We will then review different alternatives available to find a solution. Finally, we will focus the spotlight in a particular method which is robust to the underlying probability distribution.

7.2. Problem Formulation and Probabilistic Constraints

We present below a generic stochastic optimization problem:

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to:} && \\ & && Ax \leq b \end{aligned} \tag{7.1}$$

where $x \in \mathbb{R}^n$ is the vector containing the decision variables, $f(x)$ is a convex function, while the elements of matrix $A \in \mathbb{R}^m \times \mathbb{R}^n$ and vector $b \in \mathbb{R}^m$ are random variables with unknown probability distributions. The constraints of this stochastic program are:

$$a_i^\top x - b_i \leq 0, \quad i = 1, \dots, m \tag{7.2}$$

We consider an uncertainty in the data of stochastic nature, what is to say that the data vectors $d_i \doteq [a_i, -b_i]^\top$, for $i = 1, \dots, m$, are independent random vectors with dimension $n + 1$.

A classical approach to deal with the optimization problem under random uncertainty is to use the so-called chance-constrained approach [134]. In this method we introduce the risk levels $\epsilon_i \in (0, 1)$, for $i = 1, \dots, m$, associated with each constraint. Now, assuming a certain probability distribution for vectors d_i , we seek to enforce the constraints in probability, solving the following problem:

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to:} && \\ & && \text{Prob}\{a_i^\top x - b_i \leq 0\} \geq 1 - \epsilon_i, \quad i = 1, \dots, m \end{aligned} \tag{7.3}$$

While there is a vast literature on such kind of problems, we will only present a brief summary of the topic based on the work by Calafiore *et al.* [25]. For more details please refer to [25] and the references therein.

We do not intend here to survey all the research done in chance-constrained optimization, but only mention some of the fundamental keypoints. The first question that arises is under which hypotheses on the distribution of the random data

7.3. Random Data with Gaussian Distribution

d_i , the optimization problem (7.3) is a convex program. The first answer to this question, which is now a classical result, is for the case where d_i is Gaussian. In this case, the corresponding chance constraint imposes a conic quadratic constraint on x , so the problem is still convex. Similarly, it can be shown that, for the case where the values of a_i are fixed (just deterministic variables) and b_i has a log-concave probability density, then the corresponding chance constraint is also convex. A more recent result, which is an extension of the previous one, corresponds to the case when both a_i and b_i have joint log-concave and symmetric density, which was also proved to result in a convex problem.

Once we know that the problem is convex, the next step is to convert explicitly the probability constraint into a deterministic one, which depends on the particular probability distribution of the random data d_i . Again, this can be done straightforward in the case of a Gaussian distribution, while no such simple method is available in the literature for the case of other distributions. In the next section we will briefly look at the simpler Gaussian-distribution case, while in the following section we will present a novel solution introduced in [25] which tackles the case where the distribution of the random data d is unknown, but the first two moments (mean and covariance) are available.

7.3. Random Data with Gaussian Distribution

In order to illustrate the procedure to obtain the corresponding deterministic equivalent constraint, we first analyze the case where d is Gaussian. In the next section we introduce the distributionally robust approach that we will use later on this thesis. From now on, we will omit the subscript i , as the analysis is exactly the same for each constraint i . Setting $\tilde{x} \doteq [x^\top, 1]^\top$, the problem is to find a deterministic constraint, which should be equivalent to:

$$\text{Prob}\{d^\top \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad i = 1, \dots, m \quad (7.4)$$

If the random data d follows a Gaussian distribution, with mean \hat{d} and covariance matrix Γ , then the corresponding deterministic equivalent is given by:

$$\hat{d}^\top \tilde{x} + \Phi^{-1}(1 - \epsilon) \sqrt{\tilde{x}^\top \Gamma \tilde{x}} \leq 0, \quad i = 1, \dots, m \quad (7.5)$$

where $\Phi(\cdot)$ is the normal cumulative distribution function. The equivalence of both constraints means that the values of x satisfying one constraint or the other are exactly the same. This is a classical result which can be found as Theorem 10.4.1 in [133] and we reproduce the proof below.

Proof. We have that:

$$\begin{aligned} E(d^\top \tilde{x}) &= \hat{d}^\top \tilde{x} \\ \text{Var}(d^\top \tilde{x}) &= E[\tilde{x}^\top (d - \hat{d})(d - \hat{d})^\top \tilde{x}] \\ &= \tilde{x}^\top \Gamma \tilde{x} \end{aligned}$$

Chapter 7. Chance Constrained Optimization

If for some \tilde{x} we have $\tilde{x}^\top \Gamma \tilde{x} = 0$, then $d^\top \tilde{x} = \hat{d}^\top \tilde{x}$ with probability 1, and thus both constraints are equivalent. If, on the other hand, for some \tilde{x} we have $\tilde{x}^\top \Gamma \tilde{x} > 0$, then:

$$\frac{(d - \hat{d})^\top \tilde{x}}{\sqrt{\tilde{x}^\top \Gamma \tilde{x}}}$$

has a normal distribution, with mean 0 and variance 1. So, we can derive:

$$\begin{aligned} P(d^\top \tilde{x} \leq 0) &= P((\tilde{x}^\top \Gamma \tilde{x})^{-1/2} (d - \hat{d})^\top \tilde{x} \leq -(\tilde{x}^\top \Gamma \tilde{x})^{-1/2} \hat{d}^\top \tilde{x}) \\ &= \Phi(-(\tilde{x}^\top \Gamma \tilde{x})^{-1/2} \hat{d}^\top \tilde{x}) \end{aligned} \quad (7.6)$$

Thus, the probabilistic chance constraint is equivalent to:

$$\Phi\left(\frac{-\hat{d}^\top \tilde{x}}{\sqrt{\tilde{x}^\top \Gamma \tilde{x}}}\right) \geq 1 - \epsilon \quad (7.7)$$

and this is equivalent to the proposed deterministic constraint 7.5. \square

The corollary of this theorem (Corollary 10.4.2 in [133]) is that if the risk level satisfies $\epsilon \leq 1/2$, then the corresponding set of values x which satisfies the constraint is convex, and therefore so is the resulting optimization problem.

Proof. If $\epsilon \leq 1/2$, then $\Phi^{-1}(1 - \epsilon) \geq 0$ and then the result is directly derived from 7.5, as Γ is a covariance matrix, which implies it is a positive semi-definite matrix. \square

Thus, the resulting deterministic equivalent optimization problem is the following:

$$\begin{aligned} &\underset{x}{\text{minimize}} && f(x) \\ &\text{subject to:} && \\ &&& \hat{d}_i^\top \tilde{x} + \Phi^{-1}(1 - \epsilon_i) \sqrt{\tilde{x}^\top \Gamma_i \tilde{x}} \leq 0, \quad i = 1, \dots, m \end{aligned} \quad (7.8)$$

Note that in this case the resulting constraint is no longer linear but a second order cone constraint.

7.4. Distributionally Robust Approach

Now, we leave behind the Gaussian case, and we focus in a more general situation, which corresponds to the case where the distribution of the random data d is unknown, but the first two moments (mean and covariance) are known. To tackle this problem, a novel deterministic equivalent was proposed in [25], which leads us to what they have called a distributionally robust approach. In this case we look for a deterministic equivalent to the probabilistic constraint without knowledge of the particular distribution of d , and therefore should be enforced for all the family of distributions with given mean \hat{d} and covariance Γ .

We will omit the subscript i as before, for a matter of clarity. The problem considered by Calafiore *et al.* is to enforce the probabilistic constraint with respect

7.4. Distributionally Robust Approach

to the entire family \mathcal{D} of probability distributions on the random data d , with mean \hat{d} and covariance Γ . That is to say, we consider the problem of enforcing:

$$\inf_{d \sim \mathcal{D}} \text{Prob}\{d^\top x \leq 0\} \geq 1 - \epsilon \quad (7.9)$$

which is called the distributionally robust chance constraint, where the notation $d \sim \mathcal{D}$ means that we are considering the set of all the possible d that follow a probability distribution included in the family \mathcal{D} .

The equivalent presented in [25], for any $\epsilon \in (0, 1)$ is the following second order cone constraint:

$$\hat{d}^\top \tilde{x} + \kappa_\epsilon \sqrt{\tilde{x}^\top \Gamma \tilde{x}} \leq 0 \quad (7.10)$$

where $\kappa_\epsilon = \sqrt{(1 - \epsilon)/\epsilon}$. The result corresponds to Theorem 3.1 in [25], and the proof is reproduced below.

Proof. First, we consider the auxiliary random variable z , with $E\{z\} = 0$ and $\text{Var}\{z\} = I$, and we express d as:

$$d = \hat{d} + \Gamma_f z \quad (7.11)$$

where $\Gamma_f \in \mathbb{R}^{n+1, v}$ is a full-rank factor such that $\Gamma = \Gamma_f \Gamma_f^\top$, with $v \leq n + 1$ the rank of Γ .

We consider initially the case (a) when $\Gamma_f^\top \tilde{x} \neq 0$. Then, by means of a result from probabilistic inequalities (see the *one-sided Chebyshev inequality* in [135]) we have that:

$$\begin{aligned} \sup_{d \sim \mathcal{D}} \text{Prob}(d^\top \tilde{x} > 0) &= \sup_{z \sim (0, I)} \text{Prob}(z^\top \Gamma_f^\top \tilde{x} > -\hat{d}^\top \tilde{x}) \\ &= \frac{1}{1 + q^2} \end{aligned} \quad (7.12)$$

where:

$$q = \inf_{z^\top \Gamma_f^\top \tilde{x} > -\hat{d}^\top \tilde{x}} \|z\|^2 \quad (7.13)$$

In the same way than before, the notation $z \sim (0, I)$ means that we are considering the set of all the possible z that follow a probability distribution with mean 0 and covariance I .

We determine a closed-form expression for q^2 as follows. First, we notice that, if $\hat{d}^\top \tilde{x} > 0$, then we can just take $z = 0$ and obtain the infimum $q^2 = 0$. Assume then $\hat{d}^\top \tilde{x} \leq 0$. Then, the problem amounts to determining the squared distance from the origin to the hyperplane $\{z : z^\top \Gamma_f^\top \tilde{x} = -\hat{d}^\top \tilde{x}\}$, which results to be:

$$q^2 = (\hat{d}^\top \tilde{x})^2 / (\tilde{x}^\top \Gamma \tilde{x}) \quad (7.14)$$

Summarizing we have:

$$q^2 = \begin{cases} 0 & \text{if } \hat{d}^\top \tilde{x} > 0, \\ (\hat{d}^\top \tilde{x})^2 / (\tilde{x}^\top \Gamma \tilde{x}) & \text{if } \hat{d}^\top \tilde{x} \leq 0; \end{cases}$$

Chapter 7. Chance Constrained Optimization

hence, the probabilistic constraint 7.9 is satisfied if and only if:

$$\frac{1}{1+q^2} \leq \epsilon \quad (7.15)$$

i.e., if and only if:

$$\hat{d}^\top \tilde{x} \leq 0, \quad (\hat{d}^\top \tilde{x})^2 \geq \tilde{x}^\top \Gamma \tilde{x} (1-\epsilon)/\epsilon \quad (7.16)$$

or equivalently if and only if:

$$\kappa_\epsilon \sqrt{\tilde{x}^\top \Gamma \tilde{x}} \leq -\hat{d}^\top \tilde{x}, \quad \kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon}, \quad (7.17)$$

which proves that, in case (a), both constraints (7.9 and 7.10) are equivalent. On the other hand, in case (b), when $\Gamma_f^\top \tilde{x} = 0$, we simply have that:

$$\inf_{d \sim (\hat{d}, \Gamma)} \text{Prob}\{d^\top \tilde{x} \leq 0\} = 1, \quad \text{if } \hat{d}^\top \tilde{x} \leq 0 \quad (7.18)$$

and it is zero otherwise. Therefore, since $\sqrt{\tilde{x}^\top \Gamma \tilde{x}} = 0$, it follows that both constraints are still equivalent, which concludes the proof. \square

The most important thing of this result is that, through this robust deterministic equivalent, the problem raised in the previous chapter becomes convex. Thus, this result enables to find a solution to the problem posed, by means of standard convex optimization tools. In addition, we believe that this robust equivalent is an adequate approach for our problem, as it will be justified in the next chapter.

Some other references can be found in [25] using this result in other contexts, such as classification with kernel methods. The interested reader can also find referenced therein an alternative proof based on Lagrangian duality. Finally, in the same paper the result is extended to the particular case where not only the mean and covariance are known, but it is also known that the distribution of d is symmetric around the mean. That particular assumption is not true in general for our application, so we will end up here with the theoretical background in the area, but if the reader is interested it can go further on with [25] and the references therein.

Chapter 8

Robust Spectrum Allocation

In this chapter we return to the model developed in chapter 6 in order to present the proposed solution. A novel method is introduced for spectrum allocation in a cognitive radio multihop network, based on the distributionally robust approach presented in the previous chapter. As we will see, the proposed method can be implemented using a distributed algorithm, based on a primal-dual decomposition of the optimization problem posed. For this purpose, as mentioned in Chapter 6, we rely on the hypotheses validated in Chapter 2 and Chapter 3, assuming that it is possible to estimate the traffic demand and predict the effective capacity of each link from the physical layer measurements.

We believe that this approach is the most appropriate to solve the problem posed. On the one hand, we argue in favor of chance-constrained programming against robust optimization, because considering all the possible PUs' dynamics and optimizing for the worst-case will probably lead to a no solution problem in a spectrum scarcity context. That is to say, under robust optimization, none of the licensed bands in which there is some PU activity will be exploited, so this solution is clearly not suitable for our purposes. On the other hand, the introduction of a pre-defined risk ϵ allows to control the desired level of robustness of the spectrum allocation. We argue that this is a better approach than a classical expectation-based method, which only imposes to meet the requirements posed in average. Later on in this chapter we will illustrate the advantages of the proposed method through several simulation experiments.

8.1. Distributionally Robust Solution Proposed

In the last chapter we introduced a recent technique [25] to solve a stochastic optimization problem using chance-constrained programming. This method is focused on the particular case where the random data follows an unknown distribution, and we are only aware of the mean and covariance of the data. We apply this technique to our particular problem in order to reach a novel robust solution to the spectrum allocation.

8.1.1. Single Domain Spectrum Allocation

As we previously saw in Chapter 6, in order to address the spectrum allocation in the single-domain case, we have to solve each time T the following stochastic optimization problem:

$$\begin{aligned}
 & \min_{\alpha} \mathcal{C}(\alpha), \\
 & \text{s.t.} \quad \sum_{b=1}^B \alpha_b c_b h_b + \sum_{u=1}^U \alpha_u c_u \geq d, \\
 & \quad \sum_{u=1}^U \alpha_u c_u \geq w, \\
 & \quad \alpha_b \in [0, 1], b = 1, \dots, B, \\
 & \quad \alpha_u \in [0, 1], u = 1, \dots, U.
 \end{aligned} \tag{8.1}$$

The problem above is actually not well defined, as h_b is a random variable. To take into account this fact, the first and, as discussed in the introduction of chapter 6, most common approach, is to use the expected capacity, which leads us to the following equivalent deterministic constraint:

$$\bar{C}_{\text{eff}}(\alpha) = \sum_{b=1}^B \alpha_b c_b \mathbb{E}\{h_b\} + \sum_{u=1}^U \alpha_u c_u \geq d, \tag{8.2}$$

where $\mathbb{E}\{h_b\}$ can be estimated from the previously gathered PU's activity measurements. This way, we reach a convex optimization problem (assuming the defined cost functions are convex) which can be solved with standard optimization tools.

The alternative we propose, which we argue is better to address the problem at hand, is to change the expected effective capacity constraint for a probabilistic one:

$$\text{Prob} \left(\sum_{b=1}^B \alpha_b c_b h_b + \sum_{u=1}^U \alpha_u c_u \geq d \right) \geq 1 - \epsilon, \tag{8.3}$$

where ϵ is a fixed value (between 0 and 1), which leads us to a convex chance-constrained optimization problem [134] (convexity is assured in the general case only assuming h_b has a log-concave distribution and a symmetric density [136]). This approach is more difficult to solve in the general case and the deterministic equivalent constraint depends on the distribution of h_b . The solution we found suitable for this case, assuming the distribution of h_b is unknown, is to use the distributionally robust deterministic equivalent problem from [25], presented in the previous chapter. This solution is robust as it considers all the possible distributions of h_b with known mean and variance, which in our case can be estimated from the previous records of the PU's activity measurements.

Rewriting the constraint in terms of a and b ($a^T x \leq b$), according to the nomenclature and the theorem presented in the previous chapter, we have the deterministic equivalent in this case is given by:

$$\hat{a}^T x - b + \kappa_\epsilon \sqrt{[x^T, 1]^T \Sigma [x^T, 1]} \leq 0 \tag{8.4}$$

8.1. Distributionally Robust Solution Proposed

where $\hat{a} = \mathbb{E}\{a\}$, Σ is the covariance of vector $[a, -b]^\top$ and $\kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon}$. In our case, the corresponding values for a, b and x are:

$$\begin{aligned} a &= - \left[\{c_b h_b\}_{b=1, \dots, B}, \{c_u\}_{u=1, \dots, U} \right]^\top, \\ b &= -d, \\ x &= \left[\{\alpha_b\}_{b=1, \dots, B}, \{\alpha_u\}_{u=1, \dots, U} \right]^\top, \end{aligned} \tag{8.5}$$

and the covariance matrix Σ is given by:

$$\Sigma = \begin{pmatrix} c_1^2 \text{Var}\{h_1\} & 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 \\ 0 & c_2^2 \text{Var}\{h_2\} & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & & \vdots & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots & & \vdots \\ \vdots & \vdots & & & c_B^2 \text{Var}\{h_B\} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{8.6}$$

Note that all off-diagonal terms are zero because the cross terms correspond to the covariance between the activity in two different bands (h_i and h_j , with $i \neq j$), which we are assuming uncorrelated and therefore the covariance is zero (see below the justification of this assumption). Likewise, the terms corresponding to unlicensed bands and the corresponding term for the demand are zero, since in these cases the values are deterministic.

This way, we reach again a convex optimization problem, but now with a different deterministic equivalent constraint:

$$\bar{C}_{\text{eff}}(\boldsymbol{\alpha}) - \kappa_\epsilon \sqrt{\sum_{b=1}^B (\alpha_b c_b)^2 \text{Var}\{h_b\}} \geq d \tag{8.7}$$

where $\kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon}$. By this equivalence the constraint is no longer linear but a conic quadratic. Thus, the problem is still convex for convex cost functions, so it can also be solved by standard optimization tools.

Something that is worth to note is that we are basing our model in the hypothesis that the PUs' activity in one band is independent from the PUs' activity in other bands. For instance, this condition is fulfilled when the spectrum in licensed bands is assigned in amounts of at least the size of the channels used by PUs. For example, in the case of TVWS [51], the basic frequency band unit should correspond to at least the spectrum bandwidth of a single TV channel (6, 7 or 8 MHz, depending on the regulatory domain). This assumption allows us to model h_b as independent random variables on each licensed band and simplifies the resultant distributionally robust equivalent, as all the cross terms in the covariance matrix are zeros.

8.1.2. Extension for the Multiple Domain Case

Let us recall the extended model for the multiple domain case. In a generic WMN, the spectrum assignment should be performed solving the following problem:

$$\begin{aligned}
\min_{\boldsymbol{\alpha}} \quad & \mathcal{C}(\boldsymbol{\alpha}) = \mathcal{C}_{lic}(\boldsymbol{\alpha}_B) + \mathcal{C}_{unlic}(\boldsymbol{\alpha}_U), \\
\text{s.t.} \quad & (\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{H}_B) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{d}, \\
& (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{w}, \\
& \mathbf{A} \cdot \boldsymbol{\alpha}^\top \leq \mathbf{1}_{Q \times (B+U)}, \\
& \boldsymbol{\alpha}_B \in [0, 1]^{B \times L}, \\
& \boldsymbol{\alpha}_U \in [0, 1]^{U \times L}.
\end{aligned} \tag{8.8}$$

where \cdot stands for the common vector and matrix product operation and \odot stands for an element-wise matrix multiplication.

As in the previous case, we have to deal with the random variables \mathbf{H}_B . To do this we will use the same deterministic equivalents as before, on the one hand based on the expected value of \mathbf{H}_B and on the other hand considering the distributionally robust approach. For the first one, the deterministic equivalent constraint is:

$$(\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{E}\{\mathbf{H}_B\}) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{d} \tag{8.9}$$

where $\mathbf{E}\{\mathbf{H}_B\}$ is the element-wise expected value of \mathbf{H}_B .

Considering the robust approach and proceeding analogously to that shown for the single domain case, we arrive at the following deterministic equivalent constraint:

$$\begin{aligned}
& (\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{E}\{\mathbf{H}_B\}) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \\
& \quad - \kappa_\epsilon \left\| \boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \sqrt{\text{Var}\{\mathbf{H}_B\}} \right\| \cdot \mathbf{1}_B \geq \mathbf{d}
\end{aligned} \tag{8.10}$$

where $\sqrt{\text{Var}\{\mathbf{H}_B\}}$ corresponds to a matrix containing the standard deviation of each element in \mathbf{H}_B . We have again that all the cross terms between different bands or different links are zero, relying on the mild assumption that the covariance between the activity in different bands or different links is uncorrelated, and therefore the covariance is zero.

Both resulting problems are again convex if the cost functions are convex, and we have that, while in the first one the deterministic equivalent constraint is linear, in the second one it is a conic quadratic, just as in the single domain case. This ensures that both problems are convex for convex cost functions and can be solved with standard optimization tools. Next, we will develop a decentralized implementation of the algorithm, which is important in order to have a solution that scales properly as the size of the WMN grows.

8.2. Distributed Algorithm Architecture

In this section we show how to solve the optimization problem defined in the previous section in a distributed manner. Then, we propose a suitable architec-

ture for the algorithm implementation. Finally, we conclude the section with a discussion on some implementation issues.

8.2.1. Distributed Optimization

In order to find a distributed solution, we will use the dual decomposition of the described problem. This procedure is called *resource allocation via pricing* [137], because the Lagrange multipliers can be seen in a manner equivalent to the price of the resources. This approach is similar to that presented in Part II of this thesis, but in that case it was not necessary to use a dual decomposition to solve the problem. In this case the resources correspond to the frequency bands, which are then assigned to minimize the cost of the resulting allocation. The decomposition involves the relaxation of the coupling constraint, which in this case is the one imposed to avoid interference between links. Intuitively, it will be more expensive to allocate frequency bands for those links included in a higher number of collision domains. In turn, those collision domains with a larger number of links will have higher prices for the frequency bands (i.e. the greater the demand, the higher the prices).

For a matter of clarity, we will consider again the constraint with the random variable \mathbf{H}_B , which should be replaced in each case by the equivalent deterministic constraint corresponding to the previously presented approaches. In order to be able to solve the problem, we have to assume that the available spectrum is enough to cope with the demand, that is to say that the factible set is not empty. Then, the first step of the dual decomposition procedure is to form the Lagrangian by relaxing the coupling constraint. Thus, we shall consider the matrix $\boldsymbol{\lambda}$ of size $Q \times (B + U)$, with $\{\lambda_{qb}, \lambda_{qu}\} \in \mathbb{R}^+$, to get the following relaxed problem:

$$\begin{aligned}
\min_{\boldsymbol{\alpha}} \quad & \mathcal{C}(\boldsymbol{\alpha}) + \mathbf{1}_{B+U}^T \cdot (\boldsymbol{\lambda}^T \odot (\mathbf{A} \cdot \boldsymbol{\alpha}^T - \mathbf{1}_{Q \times (B+U)})) \cdot \mathbf{1}_{B+U} \\
\text{s.t.} \quad & (\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{H}_B) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{d}, \\
& (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{w}, \\
& \boldsymbol{\alpha}_B \in [0, 1]^{B \times L}, \\
& \boldsymbol{\alpha}_U \in [0, 1]^{U \times L}.
\end{aligned} \tag{8.11}$$

In the relaxed problem we add to the cost function a term which corresponds to the restriction (≤ 0) multiplied by the Lagrangian multipliers (≥ 0), so the resulting solution is a lower bound of the original problem optimum. Then, we have to maximize over $\boldsymbol{\lambda}$ in order to reach the optimum $\boldsymbol{\alpha}^*$ we are seeking, which results in this two-level optimization problem:

Chapter 8. Robust Spectrum Allocation

$$\begin{aligned}
\max_{\boldsymbol{\lambda}} \quad & \min_{\boldsymbol{\alpha}} \mathcal{C}(\boldsymbol{\alpha}) + \mathbf{1}_{B+U}^\top \cdot (\boldsymbol{\lambda}^\top \odot (\mathbf{A} \cdot \boldsymbol{\alpha}^\top - \mathbf{1}_{Q \times (B+U)})) \cdot \mathbf{1}_{B+U} \\
\text{s.t.} \quad & (\boldsymbol{\alpha}_B \odot \mathbf{C}_B \odot \mathbf{H}_B) \cdot \mathbf{1}_B + (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{d}, \\
& (\boldsymbol{\alpha}_U \odot \mathbf{C}_U) \cdot \mathbf{1}_U \geq \mathbf{w}, \\
& \boldsymbol{\alpha}_B \in [0, 1]^{B \times L}, \\
& \boldsymbol{\alpha}_U \in [0, 1]^{U \times L}. \\
\text{s.t.} \quad & \boldsymbol{\lambda} \geq 0
\end{aligned} \tag{8.12}$$

Through this relaxation, we can separate the optimization problem in two levels. We shall call $g(\boldsymbol{\lambda})$ the solution of the relaxed problem (8.11) for a given value of $\boldsymbol{\lambda}$. At a higher level, we have the master dual problem which corresponds to the update of the Lagrange multipliers $\boldsymbol{\lambda}$, variables of the dual problem:

$$\begin{aligned}
\max_{\boldsymbol{\lambda}} \quad & g(\boldsymbol{\lambda}) \\
\text{s.t.} \quad & \boldsymbol{\lambda} \geq 0
\end{aligned} \tag{8.13}$$

Then, at a lower level, and assuming we have a separable cost function, we can decompose the optimization in one sub-problem for each link l . We shall omit from the cost function the constant term in α , so the sub-problem for link l takes the form:

$$\begin{aligned}
\min_{\boldsymbol{\alpha}_l} \quad & \mathcal{C}_l(\boldsymbol{\alpha}_l) + \sum_{q \in Q_l} \lambda_q \alpha_l \\
\text{s.t.} \quad & \sum_{b=1}^B \alpha_{bl} c_{bl} h_{bl} + \sum_{u=1}^U \alpha_{ul} c_{ul} \geq d_l, \\
& \sum_{u=1}^U \alpha_{ul} c_{ul} \geq w_l, \\
& \alpha_{bl} \in [0, 1], b = 1, \dots, B, \\
& \alpha_{ul} \in [0, 1], u = 1, \dots, U.
\end{aligned} \tag{8.14}$$

where Q_l are the subset of the collision domains where the link l is included, λ_q is the row q of the Lagrange multipliers matrix $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}_l$ are the allocation variables for link l . It is worth to note that given the value of λ_q this problem can be solved locally by the link, as it has all the other necessary information. That is to say, both the estimation of the h_{bl} distribution parameters as well as the effective capacity values (c_{ul} and c_{bl}) are calculated locally, and they are directly used in the optimization, without need to forward them to any other node.

With this approach we actually solve the dual problem, so it will only work properly if we have strong duality, which holds if the original problem is convex and with strictly feasible solutions (which is commonly known as the Slater's condition, see Section 5.2.3 in [110]). If the function $g(\boldsymbol{\lambda})$ is differentiable, then the master dual problem can be solved with a gradient method (see Section 9.3 in [110]). Thus, the update of the Lagrange multipliers following this method is given by:

8.2. Distributed Algorithm Architecture

$$\lambda_{qb}^{t+1} = \left[\lambda_{qb}^t + \sigma \cdot \left(\frac{\partial g}{\partial \lambda_{qb}} \right)_t \right]^+ \quad (8.15)$$

where t is the iteration index, σ a positive suitable step-size (sufficiently small), and the projection $[\cdot]^+$ ensures the new value to be non-negative. Substituting by the corresponding gradient we reach the following:

$$\lambda_{qb}^{t+1} = \left[\lambda_{qb}^t + \sigma \cdot \left(\sum_{l \in q} \alpha_{bl}^* (\boldsymbol{\lambda}^t) - 1 \right)_t \right]^+ \quad (8.16)$$

which should be solved for each collision domain by some node in charge for all the domain. We shall call these nodes *domain referents* in the proposed hierarchy for the algorithm implementation, presented in the next section. Notice that $\alpha_{bl}^* (\boldsymbol{\lambda}^t)$ is the optimum of the sub-problem 8.14 for the previous value of the Lagrange multipliers $\boldsymbol{\lambda}^t$.

In summary, the relaxed problem $g(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} \geq 0$, can be decomposed as:

$$g(\boldsymbol{\lambda}) = \sum_l g_l(\boldsymbol{\lambda}) + \mathbf{1}_{B+U}^\top \cdot \boldsymbol{\lambda}^\top \cdot \mathbf{1}_Q \quad (8.17)$$

where $g_l(\boldsymbol{\lambda})$ is the subpart of the dual problem corresponding to link l . The dual decomposition results in each link l solving 8.14 for the given $\boldsymbol{\lambda}$, to obtain the optimum values $\alpha_{bl}^* (\boldsymbol{\lambda}^t)$ and $\alpha_{ul}^* (\boldsymbol{\lambda}^t)$, which are unique for strictly convex cost functions [110]. The gradient method ensures the dual variable $\boldsymbol{\lambda}^t$ will converge to the dual optimal $\boldsymbol{\lambda}^*$ as $t \rightarrow \infty$. Since the duality gap for the original problem is zero (as Slater's condition is satisfied) and the solution to the subproblems is unique, the primal variables $\alpha_{bl}^* (\boldsymbol{\lambda}^t)$ and $\alpha_{ul}^* (\boldsymbol{\lambda}^t)$ will also converge to the primal optimal variable $\boldsymbol{\alpha}^*$.

8.2.2. Proposed Algorithm Architecture

From the distributed optimization presented in the previous section we arrive at a decentralized implementation of the algorithm, according to the architecture described below. We say that it is a decentralized solution following the taxonomy described in [119] where it is stated that the allocation is performed by *more than one but not all of the nodes within the network*. In particular we work with a cluster-based solution where each cluster corresponds to a collision domain in the WMN. Each collision domain has a domain referent which is the head cluster in the proposed algorithm architecture.

In Figure 8.1 we can see the proposed hierarchy, where the lower level correspond to links, and the next level to the head clusters, which are the collision domain referents. It is worth to note that one link can belong to one or many collision domains as is shown in the example. In this case, the communication during the optimization should be with all the domain referents corresponding to all the collision domains where it belongs. This way it will receive all the updated prices

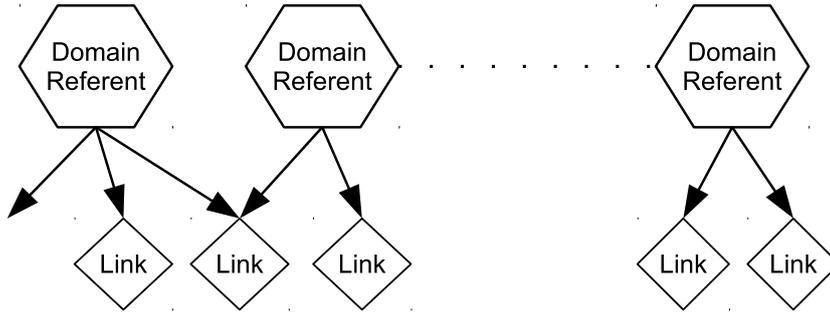


Figure 8.1: Proposed architecture for the decentralized implementation.

for the several collision domains in which it is included. Summarizing, the distributed optimization with the proposed architecture is solved with the following algorithm:

Dual decomposition algorithm for spectrum allocation

- Parameters: each link l estimates the local effective capacities c_{bl} and c_{ul} , and local PUs' activity statistics computing the mean and the variance of h_{bl} .
 - Inputs: each link l has its own capacity requirements for data and control traffic, given by d_l and w_l respectively.
 - Hierarchy: each collision domain has a predefined domain referent.
 - Initialization: at $t = 0$ set $\alpha = 0$ and $\lambda = 0$.
1. Each link locally solves the spectrum allocation by computing $\{\alpha_{bl}^*(\lambda^t), \alpha_{ul}^*(\lambda^t)\}$, the optimum of the corresponding lower level sub-problem 8.14, which is then communicated to each domain referent.
 2. Each domain referent receives from each link the previously computed allocation, updates the prices according to equation 8.15, and then, it broadcasts the new prices λ^{t+1} within the domain.
 3. IF $\max \left\{ \max_{\alpha_{bl}} |\alpha_{bl}^t - \alpha_{bl}^{t-1}|, \max_{\alpha_{ul}} |\alpha_{ul}^t - \alpha_{ul}^{t-1}| \right\}$ sufficiently small END, ELSE go to step 1.

8.2.3. Implementation Issues

For the purpose of an actual implementation of the proposed method there are some issues to solve in a real WMN. In this section we will discuss possible

8.2. Distributed Algorithm Architecture

solutions to these issues. First, we must resolve the conflict graph construction in order to find all the collision domains in the WMN. We envision several ways to do this, ranging from a planned solution at the deployment phase up to a distributed graph construction solution. Then, the next point which is related to the above, is to define who is the referent node in each collision domain. Finally, we will comment on the possibilities to implement the domain referent assignment, either by one or several nodes in the network, or even without being a physical solution but a distributed communication protocol. This is related to how the collision domain referents communicate with each other.

Starting with the conflict graph construction, on one hand, it is possible to pre-compute it during the network design stage. This graph can also be verified with measurements during the links' installation. This way, it is possible to know a priori all the interference conflicts. On the other hand, we can leverage on the sensing capabilities of the nodes in the network¹ to detect interferent links and communicate this information to a predefined central entity. With such information from every link centralized in a fusion center, this entity is able to construct the conflict graph.

The next step is to obtain the maximal cliques of the conflict graph, which correspond to the collision domains we are looking for. To solve this problem, which is commonly known as the maximal clique problem, we can use an efficient implementation of the well-known Bron-Kerbosch algorithm [138]. Once we have the complete list of collision domains, we have to proceed to select the referent for each one. To do this we can use as the first selection criterion those nodes that are in a higher number of collision domains, in order to simplify the system architecture, as we will have fewer referents. Then we can simply use an arbitrary criterion, e.g. the higher MAC address or just a pseudorandom selection, just to keep a unique referent per collision domain. Finally, when all the collision domains have the corresponding referent, we are able to carry out the periodic spectrum allocation, following the distributed algorithm described before.

Concerning the selection of a referent for each domain, one possibility is that a central entity (e.g a particular node or a set of redundant nodes) is in charge of selecting the corresponding node acting as referent for each collision domain. In any case, its role would only be important at the beginning of the network operation. Then, it would only be necessary to recompute if changes in the network topology occur, which depends strongly on whether the network nodes are fixed or mobile. As the main case of interest for us is with fixed nodes, it is unlikely that the central entity has much activity once the network is operative. At the other extreme, we can think of a completely decentralized solution, starting from a distributed mechanism for the conflict graph construction as presented in [139]. Then, after the conflict graph is known by every node in the network, each node can obtain the domain referents following the same procedure described before. This way, every node in the network will know who the referents are, including the referents themselves, without need to be informed by a central entity.

¹Recall we are assuming that each node has a dedicated interface for sensing purposes.

8.3. Simulation Experiments

In order to test the proposed framework we consider three simulation experiments. In the first case we evaluate the algorithm for a single point to point link, so it is the case with only one collision domain presented in Section 8.3.1. Then, we test the method for a simple network with four nodes and three links, now with two collision domains (topology shown in Figure 6.2). Finally, the last experiment corresponds to the topology of a real network which is part of the Plan Ceibal's rural Internet access deployment introduced in Chapter 2. In all the simulations the number of frequency bands used seeks to reflect a real-world situation, taking a quantity of unlicensed spectrum of similar order than the number of 5GHz U-NII bands², and on the other hand a considerable amount of licensed spectrum, which might correspond to TV or cellular frequency bands. A total of 50 frequency bands is considered for the first simulation, analyzing both the fixed case of 15 unlicensed and 35 licensed bands, and also varying the proportion of bands of each type. For the latter two experiments we sought to simulate a tighter situation with fixed amounts of each kind of spectrum, with 25 licensed bands and 15 unlicensed, totalizing 40 bands. The effective capacities for each band are all drawn from a uniform distribution at the beginning of each experiment, and remain the same during all the simulation.

For the experiments we set as goal to minimize the total assigned spectrum, which might be a suitable objective for the SUs. Thus, if all SUs operate with this objective, it ensures to have a friendly coexistence of multiple devices from different networks, sharing all the available spectrum. This leads us to use the following cost functions:

$$\mathcal{C}_{lic}(\boldsymbol{\alpha}_B) = \sum_{l=1}^L \sum_{b=1}^B \alpha_{bl} \quad \text{and} \quad \mathcal{C}_{unlic}(\boldsymbol{\alpha}_U) = \sum_{l=1}^L \sum_{u=1}^U \alpha_{ul} \quad (8.18)$$

Anyway, it is just an example to illustrate the algorithm operation and the proposed framework is more general, enabling to consider other targets of interest that would lead to different cost functions.

In order to model and simulate the PUs' activity in licensed bands, we consider a two-state On-Off discrete time markov chain (DTMC) spectrum occupancy model (see Figure 8.2), which has been proved to be suitable [140, 141]. The parameters involved in the model are the transition probabilities p_{on} and p_{off} , which will determine the average busy and non-busy time, $\pi_{on} = p_{on}/(p_{on} + p_{off})$ and $\pi_{off} = p_{off}/(p_{on} + p_{off})$, respectively. While it is not necessary for the implementation of the algorithms, as a measurement-based estimation is sufficient, it is possible with this model to obtain closed-form expressions of $E\{h_{bl}\}$ and $\text{Var}\{h_{bl}\}$ from the model parameters.

All bands are considered of equal spectrum bandwidth, each with a generic value BW. Each simulation is performed for a total time period of 1000 T, where

²From 6 to 24 non-overlapping WiFi channels in the US, depending on the channel bandwidth considered.

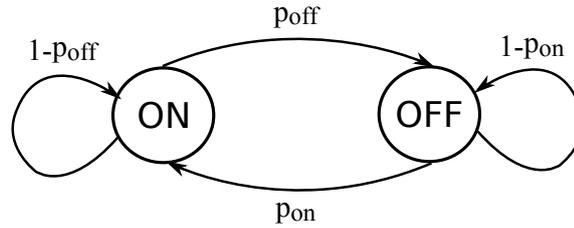


Figure 8.2: Two-state On-Off DTMC spectrum occupancy model.

FT: fortune teller (knows h_{bl} realizations in advance)

EXP: expectation based approach

CONS: conservative (only uses unlicensed spectrum)

ROB- ϵ : robust approach (ϵ - value taken by the parameter)

IND-EXP: individual decision for each link using EXP

IND-ROB- ϵ : individual decision for each link using ROB- ϵ

Table 8.1: Algorithms considered for performance comparison.

T (also a generic value) is the time interval between spectrum allocations. Finally, we use the DTMC spectrum occupancy model to simulate the PUs' activity, completing a total of 20 transitions during each interval, a fixed value used for all the experiments. The initial occupancy for each licensed band is drawn in all cases from the corresponding stationary distribution π_{on} , in order to start each simulation already at steady state.

As reference results we consider the solution to the proposed problem when the realizations of h_{bl} are known in advance. We shall call this method the fortune teller (FT). We also include as reference another simple approach to solve the problem, which we shall note as CONS (for conservative), and consists of assigning only unlicensed bands to meet the requirements. It is clear that this assignment is the safer one concerning the PUs, but it has the disadvantage of missing out on using all the available licensed bands. Furthermore, it cannot solve the problem when the unlicensed spectrum is not enough to reach the throughput lower bound.

To reference the proposed algorithms, we shall call EXP the expectation based approach with a mean value capacity constraint. On the other hand, we shall call ROB- ϵ the one that takes the robust deterministic equivalent constraint, where ϵ indicates the value taken by the parameter. Finally, for the cases with multiple links, we also consider the possibility that each link takes a decision individually. We will note those methods as IND-X, where X corresponds to the algorithm that each link uses to perform the spectrum allocation (e.g. EXP or ROB- ϵ). A summary of the aforementioned methods, which will be referenced throughout the simulations, is presented in Table 8.1

For performance comparison we analyze in all cases the spectrum allocated and the average effective capacity resulting from the assignment. We also study the short-term effectiveness (indicated as STE in the results) of the proposed methods,

Chapter 8. Robust Spectrum Allocation

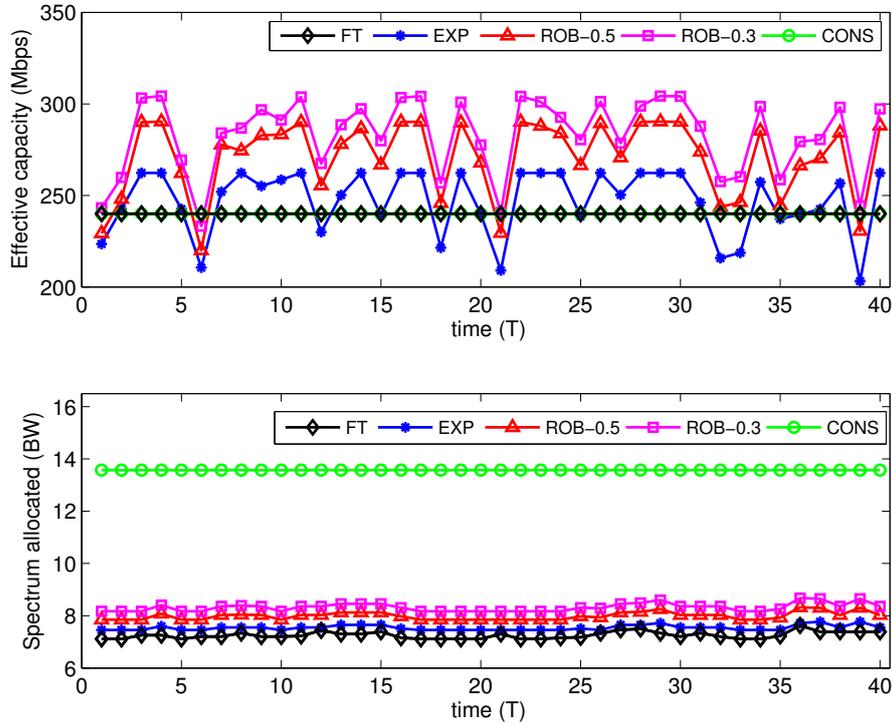


Figure 8.3: Experiment example with $d = 240$ Mbps, $p_{on} = 0,01$ and $\pi_{on} = 0,1$.

which is the percentage of time intervals where the effective capacity assigned is above or equal to the defined lower bound. Throughout the simulations we will see that although the expected value approach meets the requirements in average, and is the most efficient regarding spectrum usage, robust approaches perform much better at short scales, with a reasonable extra cost in terms of spectrum bandwidth allocated.

8.3.1. Single Domain Spectrum Allocation

The first example corresponds to the single domain case, which is the suitable model for a single point-to-point link. In this experiment we consider 15 unlicensed bands and 35 licensed ones, and we analyze the algorithm allocation for different values of p_{on} and π_{on} . Then, we vary the proportion of bands of each type, and we study the algorithm allocation for different values of d , now with fixed values of p_{on} and π_{on} . The capacities are taken from a uniform distribution between 5 Mbps and 25 Mbps for unlicensed bands, and values 50% higher for licensed bands. Typically unlicensed bands would be more crowded, so we try to reflect this fact in the selected capacity values for each band.

In Figure 8.3 we show an example simulation with parameters $p_{on} = 0,01$ and

8.3. Simulation Experiments

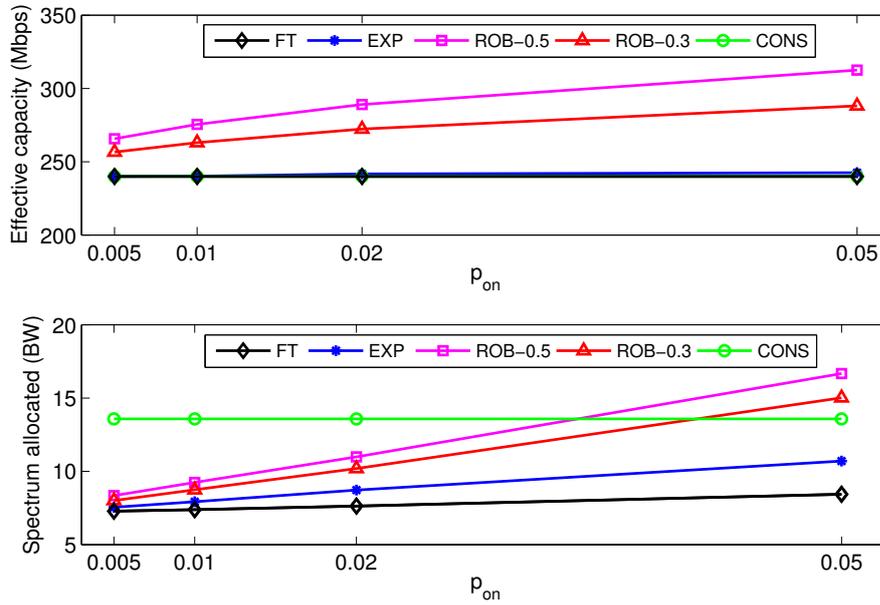


Figure 8.4: Effective capacity and spectrum allocated as a function of p_{on} when $\pi_{on} = 0,1$ and $d = 240$ Mbps (total unlicensed capacity = 248 Mbps).

$\pi_{on} = 0,1$, the same for all licensed bands. As we said before, each simulation lasts 1000 T, while in the example figure we only show 40 T for a matter of clarity. The throughput lower bound d is set at 240 Mbps, somewhat below the total unlicensed bands' capacity which is 248 Mbps. In all the simulated situations for this single domain case, only using unlicensed spectrum is enough to meet the requirements, which allows to get a solution with CONS. Notice that CONS and FT are superimposed in the capacity plot, as they both solve a deterministic optimization problem and reach the equality in the constraint, assigning exactly the required demand d . Furthermore, it can be seen that ROB-0.3 allocates more spectrum than ROB-0.5, since a smaller ϵ implies more robustness (and thus more spectrum required), and both assign more spectrum than EXP, which is the least robust one.

We first analyze the results for different values of p_{on} (see Figures 8.4 and 8.5). As we can see all the methods meet the throughput lower bound in average (they are all above or equal to $d = 240$ Mbps), something we ensure by placing it as a constraint in the problem formulation. Looking at the spectrum assignment, the stochastic approaches clearly outperform CONS, with better spectral efficiency and closer to the FT optimum solution as p_{on} goes to 0. It is clear that for lower values of p_{on} is when these methods make better sense, as it indicates higher possibilities of making profit from licensed bands. While robust approaches allocate more spectrum than the expected value solution, in exchange they get much better performance at short scale. The average success rate is between 90 % and 95 % for

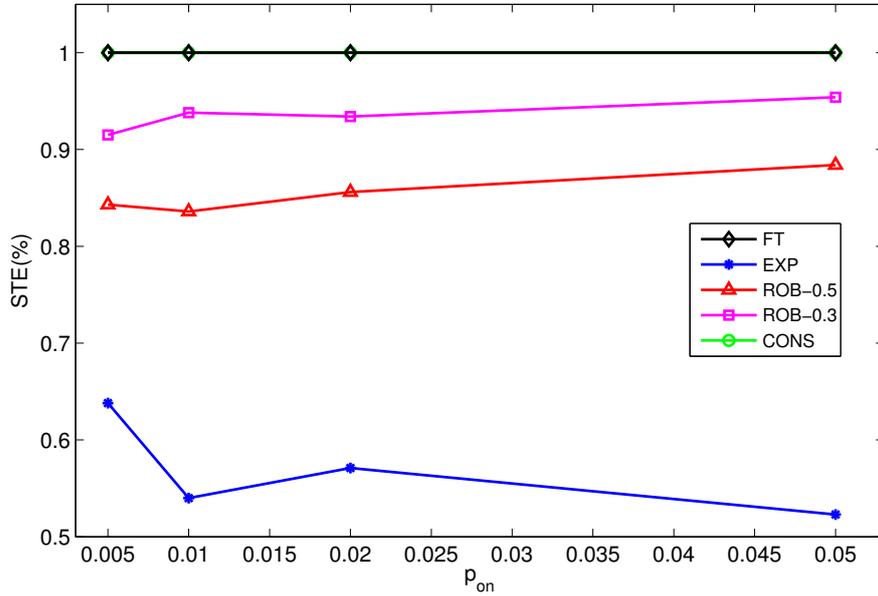


Figure 8.5: STE as a function of p_{on} when $\pi_{on} = 0,1$ and $d = 240$ Mbps (total unlicensed capacity = 248 Mbps).

ROB-0.3 and between 83 % and 90 % for ROB-0.5, while EXP is always below 65 %. The extra spectrum assigned in average by robust approaches implies some average capacity overallocation with respect to the stated throughput lower bound. However, this mild conservatism, allocating not much more spectrum than FT, is what enables a higher probability to meet the throughput lower bound also at the short scale.

Now, we analyze the performance for different busy times (see Figures 8.6 and 8.7). We have again a clear advantage of the stochastic methods against CONS, with less spectrum allocated to meet the same requirements. Furthermore, the advantage is higher for lower values of π_{on} , which are the most interesting situations to benefit from licensed spectrum. Robust approaches present again some average capacity overallocation, which is higher for lower values of π_{on} . In return, their short term effectiveness stands out again, with an average success rate of 93 % and 86 %, for ROB-0.3 and ROB-0.5 respectively, against a poor 57 % for EXP. This implies that, although the EXP solution meets the requirements in average, more than 40 % of the time the effective capacity assigned is below the stated throughput lower bound. Except for particular cases, where an expectation based solution might be sufficient, we argue instead that a robust approach will be more suitable in practice, with much higher short term performance at a reasonable cost in terms of spectrum.

Lastly, we set as fixed values $p_{on} = 0,01$ and $\pi_{on} = 0,1$, and we vary the number of unlicensed bands (from 10 to 20), keeping the same total number of bands

8.3. Simulation Experiments

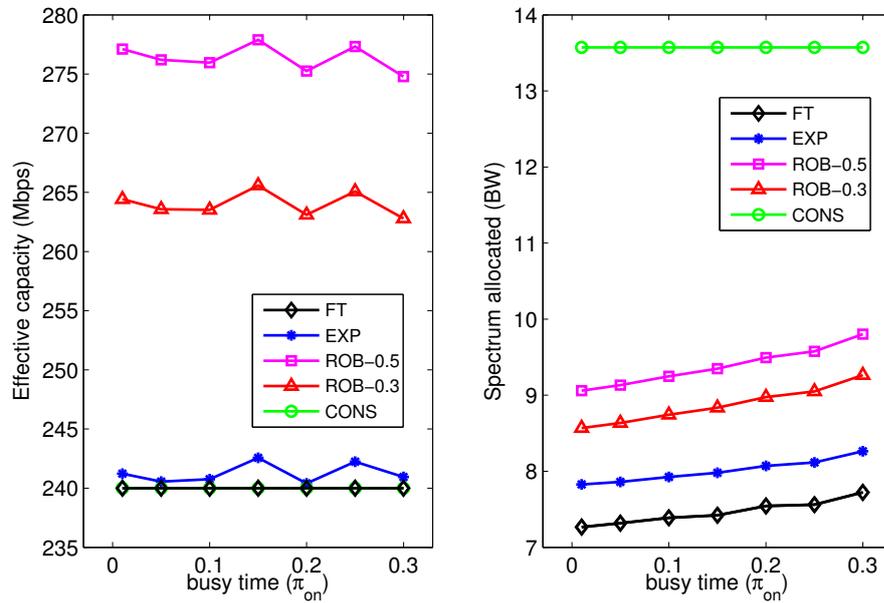


Figure 8.6: Effective capacity and spectrum allocated as a function of π_{on} when $p_{on} = 0,01$ and $d = 240$ Mbps (total unlicensed capacity = 248 Mbps).

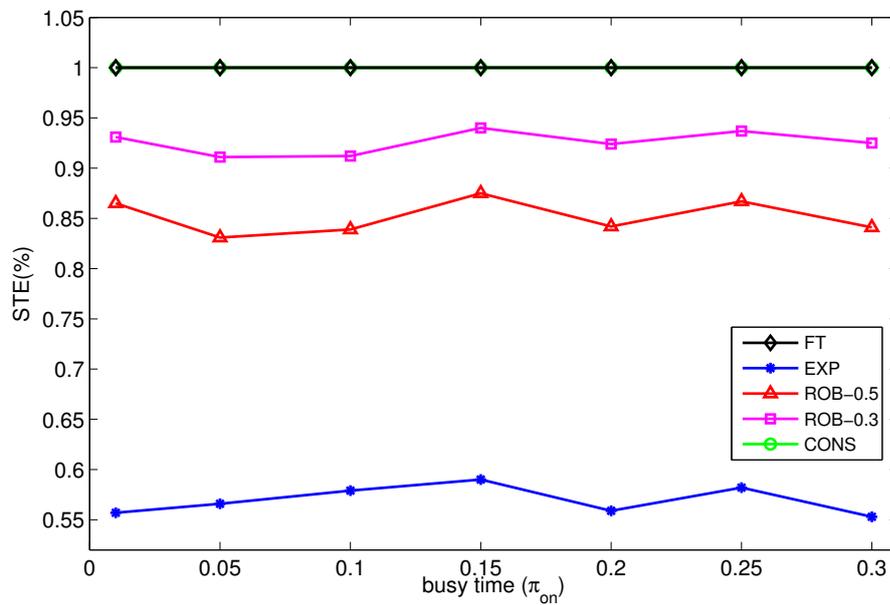


Figure 8.7: STE as a function of π_{on} when $p_{on} = 0,01$ and $d = 240$ Mbps (total unlicensed capacity = 248 Mbps).

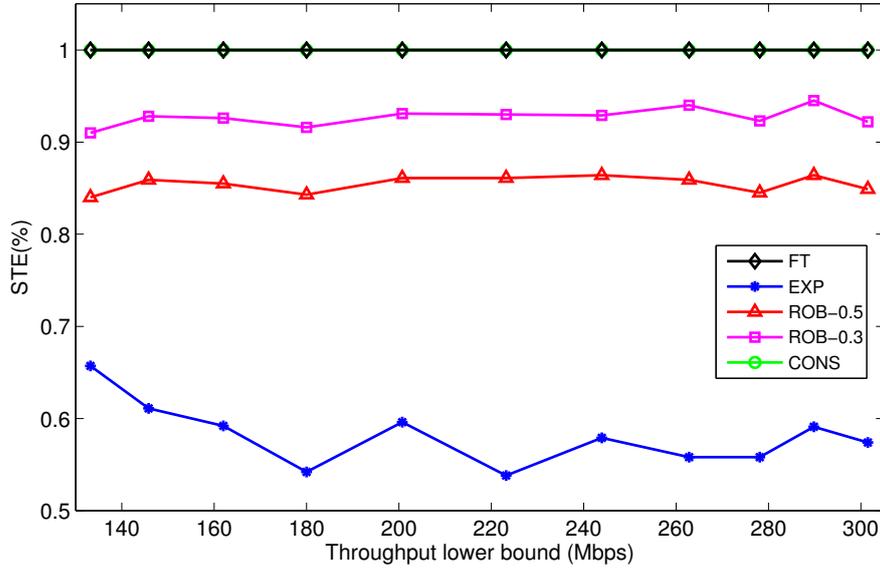
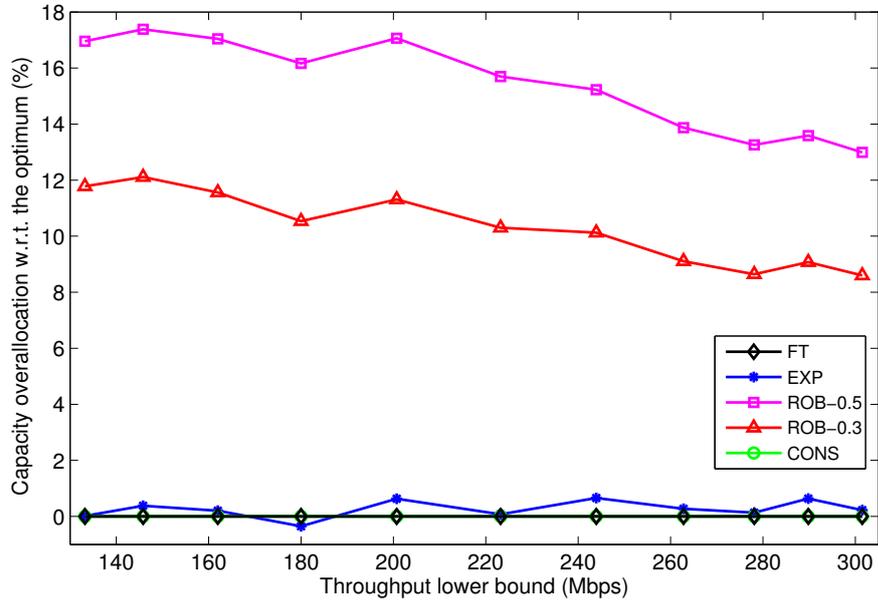


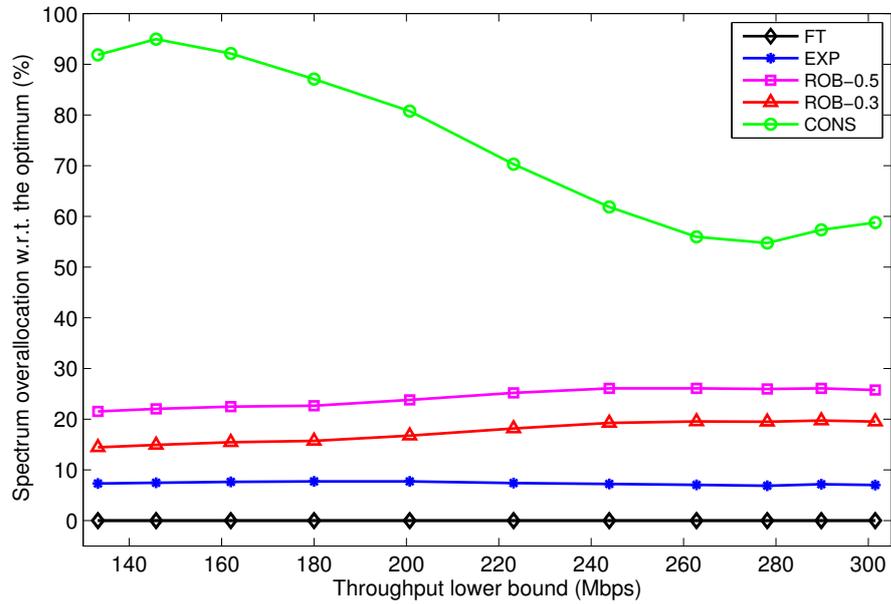
Figure 8.8: STE for $d = 0,9 \cdot \sum_u c_u$ ($p_{on} = 0,01$ and $\pi_{on} = 0,1$).

(50) for all cases. This way, the total unlicensed bands' capacity changes and we consider for each case a throughput lower bound equal to 90% of its value. In Figures 8.9(a) and 8.9(b) we can see the spectrum and effective capacity overallocation compared to the FT optimum. The stochastic methods clearly outperform CONS, with higher advantage for lower minimum throughput requirements, which is an expected result, as in this case it corresponds to a situation with more licensed bands. As the proportion of unlicensed spectrum gets higher, the benefit from using available licensed bands is lower, but it is still worth using it for reaching greater spectral efficiency. When we look at the short scale performance (see Figure 8.8) we can see again that the proposed robust approach clearly outperforms the expectation based solution. While ROB-0.3 achieves an average success of 92% and ROB-0.5 reaches 85%, EXP only gets a poor 58%. Furthermore, the price for that better performance is only between 6% to 15% more spectrum assigned than EXP, and between 10% to 25% more than the lower bound defined by the FT solution. It is neither too much if we look at capacity overallocation, with only between 8% to 18% more than the optimum FT solution.

8.3. Simulation Experiments



(a) Capacity overallocation with respect to the optimum.



(b) Capacity overallocation with respect to the optimum.

Figure 8.9: Simulation results for $d = 0,9 \cdot \sum_u c_u$ ($p_{on} = 0,01$ and $\pi_{on} = 0,1$).

8.3.2. Multi-Domain Case: 3-links Network Topology

The network considered for the second experiment is the example introduced in Chapter 6, shown in Figure 6.2. In this case the link number 1 and the link number 2 interfere with each other, and the same happens between the links number 2 and number 3. This gives us a total of two collision domains in the network, which correspond to the two maximal cliques of the conflict graph: the sets $\{1,2\}$ and $\{2,3\}$.

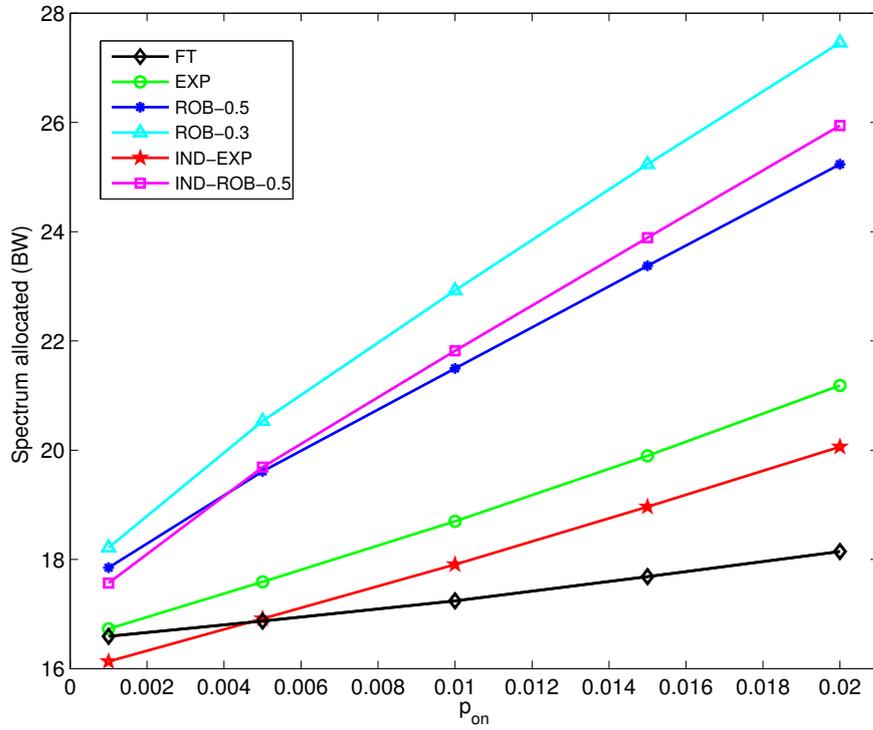
In this experiment we consider less spectrum than before, with 15 unlicensed bands and 25 licensed bands, totalizing 40 frequency bands. The capacities are all drawn from the same uniform distribution but with independent values for each link, and biased again with higher values for the licensed bands ($\sim 60\%$ more than unlicensed bands). In this case the throughput lower bound considered for each link is beyond the total capacity of unlicensed bands, so there is no possible solution using the conservative approach. We repeat the analysis from the previous experiment, varying the activity of the primary users through the values of p_{on} and π_{on} .

Now, an important thing to clarify is how to proceed with the evaluation of the STE performance indicator for a WMN. As we now have several links in the network, each of the them can independently reach or not the throughput lower bound, so we will consider two different STE values. On the one hand we have the average STE (A-STE), which is the average over all the links' STE individual values. On the other hand we have the global STE (G-STE), which is the percentage of time intervals where the effective capacity assigned is above or equal to the defined lower bound on all the links. The difference is that whereas in the former case all the situations when the capacity constraint holds in each link independently are counted as successful, in the latter we only are count as successful the cases when the constraint is accomplished in all the links simultaneously. That said, we can now comment on the results presented in Figures 8.10(a) and 8.10(b).

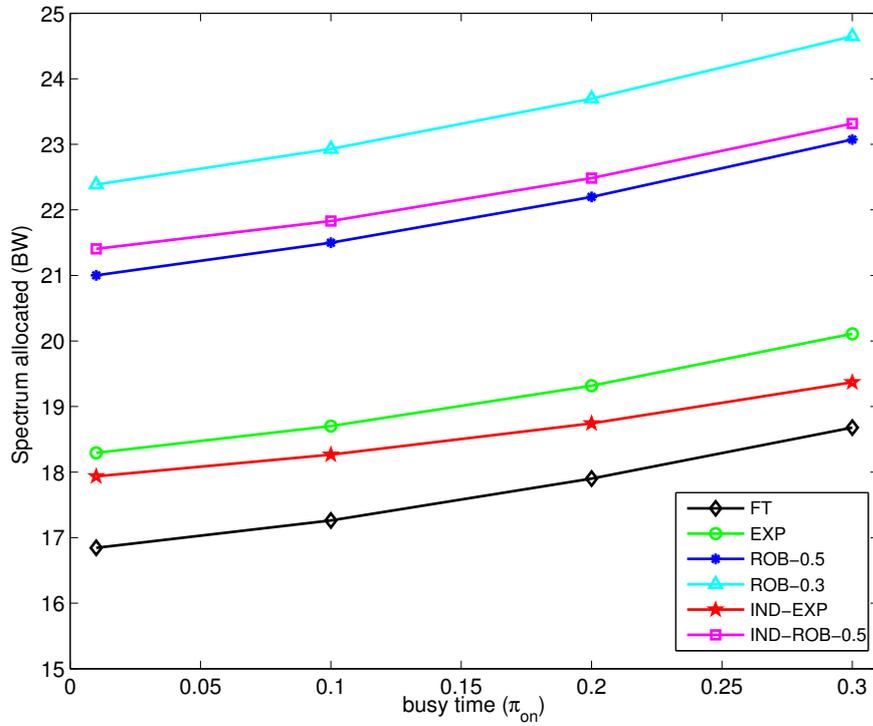
The first thing to notice is that we have again an increasing amount of spectrum allocated as p_{on} or π_{on} rises. If we look at the results for the individual methods, where each link makes a decision on its own, we can see that while the amount of spectrum allocated is similar than the other methods, they have both a null performance considering the STE. The reason for this fact is that no coordination between links is done, so two links in the same collision domain can assign the same frequency band ignoring the other, which results in less capacity than expected for each of them. Thus, with this kind of assignment we are always below the required capacities with a probability close to one.

Comparing the performance of the expected value approach and the proposed robust schemes, it becomes clear again looking at the results in Figures 8.11(a) and 8.11(b) the advantages of the latter. In one case, as p_{on} rises, we have a G-STE of 60% and 80% for ROB-0.5 and ROB-0.3 respectively, against a 20% for EXP. On the other hand, the extra spectrum required to reach this robustness is only 20-30% more than what EXP assigns, and 50-60% more than the optimum only reachable by a diviner. Similar results are obtained when we analyze the case where π_{on} varies, hence they are not included in the thesis. An

8.3. Simulation Experiments



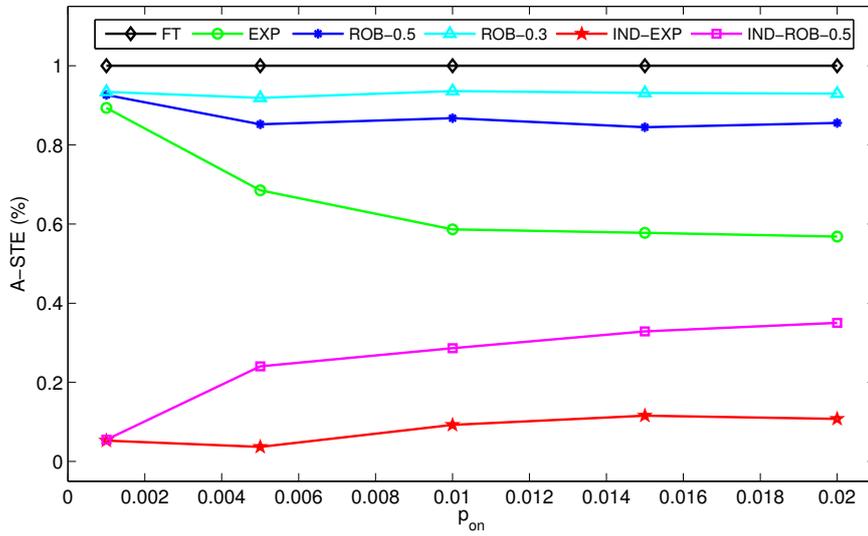
(a) Total spectrum allocated as a function of p_{on} .



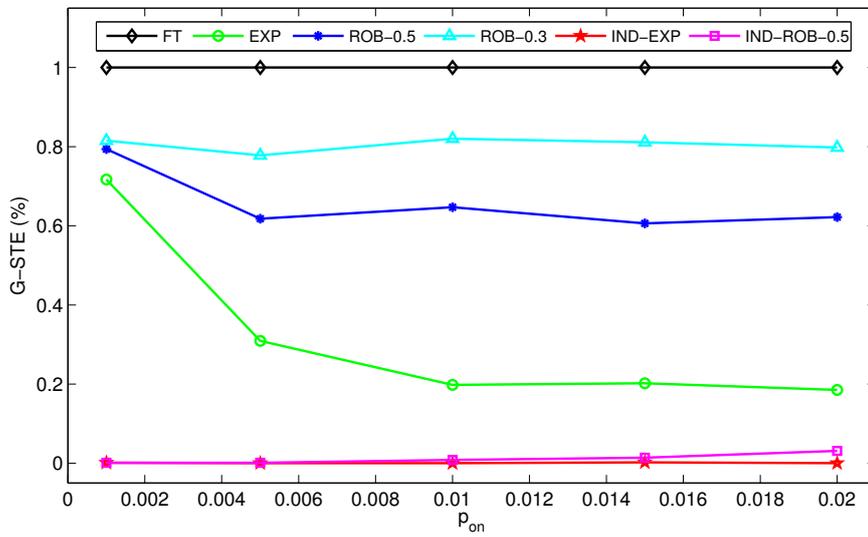
(b) Total spectrum allocated as a function of π_{on} .

Figure 8.10: Simulation results for the 3-links Network Topology.

Chapter 8. Robust Spectrum Allocation



(a) A-STE as a function of ρ_{on} .



(b) G-STE as a function of ρ_{on} .

Figure 8.11: Short term performance for the 3-links Network Topology.

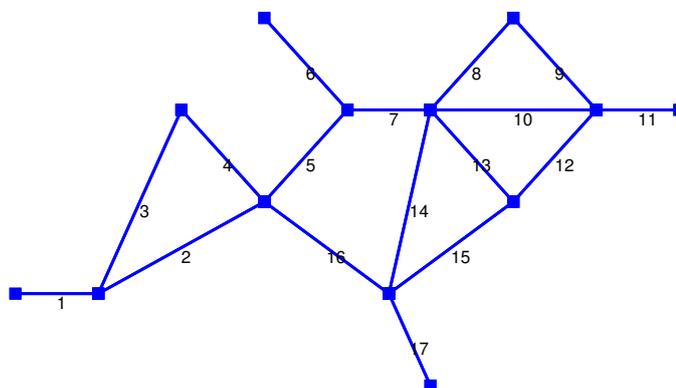


Figure 8.12: Real-world network example topology.

important property to note in the algorithms' comparison is the performance invariance with respect to the PUs activity parameters, which is noticed when we look at the flatness of the STE curves. This fact implies that when the PUs dynamics statistics are well estimated, the algorithm performance does not depend of it, which is a nice property of the spectrum allocation framework developed.

While the focus of our work is not centered in the particular optimization algorithm used, some comments regarding its convergence are in order. The number of iterations of each algorithm run depends on the accuracy desired for the assignment variables. As on each iteration the updated values have to be sent to each domain referent, this value will determine the control plane traffic load generated in the network. Based on our simulation experiments, typically 200 iterations are enough to reach an adequate precision in the allocation variables. We consider that this number of iterations is completely reasonable, taking into account that the algorithm is designed to run periodically at a medium to large timescale, and would not overload the network in a real-world implementation.

8.3.3. Real World Network Topology

The last simulation experiment corresponds to a real-world network, which is taken from a real deployment of rural Internet access for schools from Plan Ceibal (see Chapter 2). In this case the network is composed by 13 nodes and 17 links. In Figure 8.12 we can see the network topology, while Figure 8.13 illustrates the corresponding conflict graph. In this case we have a total of 16 maximal cliques in

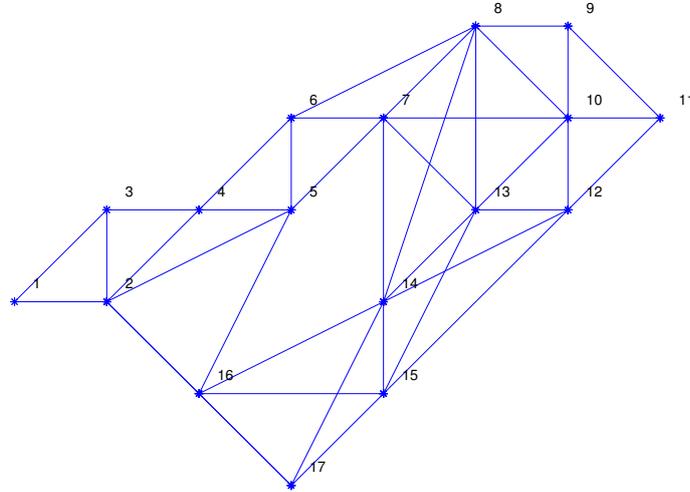


Figure 8.13: Real-world network example conflict graph.

the conflict graph, which are all the collision domains in the WMN.

For this experiment we consider the same spectrum than for previous case, with a total of 40 frequency bands, 15 unlicensed and 25 licensed. In this case the parameters of the PUs' activity are fixed, with $p_{on} = 0,01$ and $\pi_{on} = 0,1$. We use again a uniform distribution for the capacities, from which independent values were drawn for each link, and we also maintain the bias in favour of licensed bands, but now two different situations are considered. In the first case licensed bands have $\sim 60\%$ more capacity than unlicensed bands, while in the second case the bias is larger, reaching $\sim 160\%$ extra capacity. As in the previous case the total capacity of unlicensed bands is not enough to reach the throughput lower bound considered for each link, so there is no possible spectrum allocation using the conservative approach. In this experiment we also evaluate other values of ϵ , the robust algorithm parameter, in order to gain knowledge on how one should choose its value.

In Table 8.2 the results are summarized for both cases, the one with smaller and the one with larger bias against unlicensed bands capacities. As we can see, in both cases the performance increases as ϵ decreases, and the expected value approach is even worse than all of them. We also note that the fall of G-STE is much higher than that of A-STE, due to its exponential dependence on the number of links. In Figure 8.14 the comparison is illustrated for the case of a smaller bias. Based on these results one should choose the parameter ϵ according to the performance required in the specific network operation.

8.3. Simulation Experiments

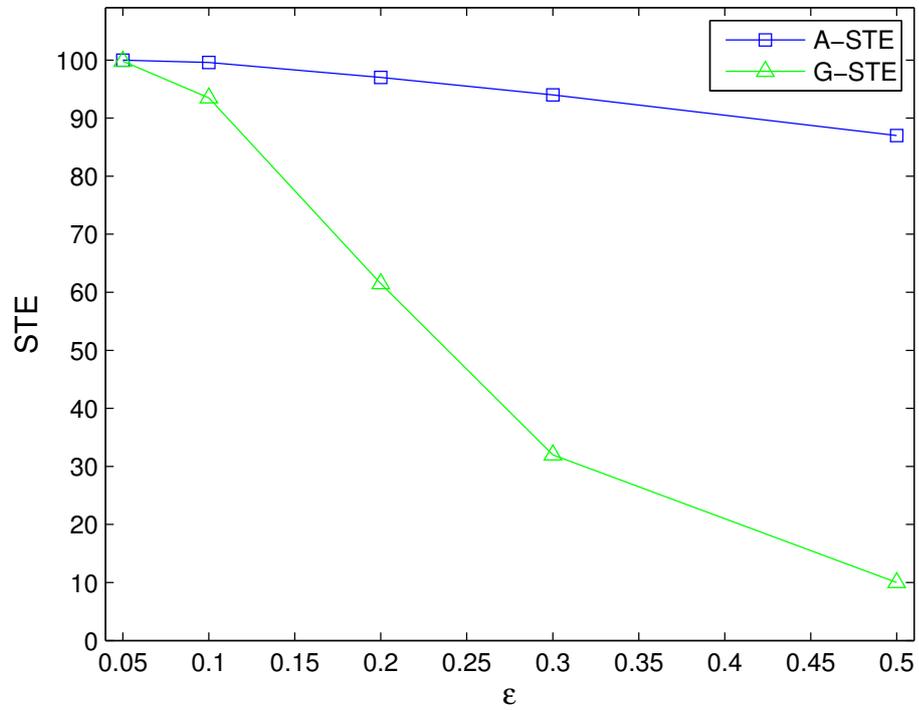


Figure 8.14: A-STE and G-STE for different values of the parameter ϵ .

Method	small bias		large bias	
	A-STE (%)	G-STE (%)	A-STE (%)	G-STE (%)
FT	100	100	100	100
ROB-0.05	99.988	99.8	99.96	99
ROB-0.1	99.6	93.5	99	92
ROB-0.2	97	61.5	96	50
ROB-0.3	94	32	92	26
ROB-0.5	87	10	85	6
EXP	74	1	64	0

Table 8.2: A-STE and G-STE comparison for the different methods.

Chapter 8. Robust Spectrum Allocation

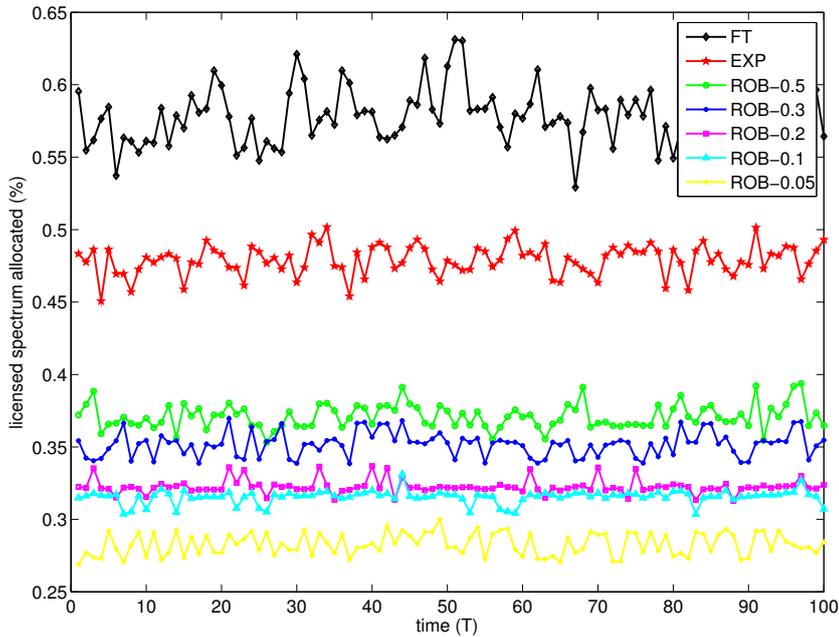


Figure 8.15: Proportion of licensed spectrum allocated for the different methods, when we have a smaller bias against unlicensed band capacities.

Finally, it is worth to note that for a larger bias, and taking into account the same value of ϵ , the performance is worse. This may seem counterintuitive at first, but it becomes clear when we look at the percentage of licensed and unlicensed spectrum assigned in each case. What happens is that when the bias is larger, it is more convenient to assign a greater proportion of licensed spectrum, but it also implies a higher risk, because these bands may become occupied. In Figures 8.15 and 8.16 we can see the evolution (100 first T of the simulation) of the licensed spectrum proportion assigned, which are clearly higher for the larger bias case. There is a compromise between the increased use of licensed spectrum, and the risk it takes to assign these bands. This is also clear when we look at the variation of the licensed spectrum proportion assigned with respect to the different values of ϵ . The more robust is the allocation, the lower the proportion of licensed spectrum assigned, and therefore higher the proportion of unlicensed spectrum. The latter cannot be *occupied* by a PU and hence the effective capacity is 100% available all the time. Remember that the effective capacity considered for each unlicensed band, already takes into account the time sharing with other interferent networks.

8.3. Simulation Experiments

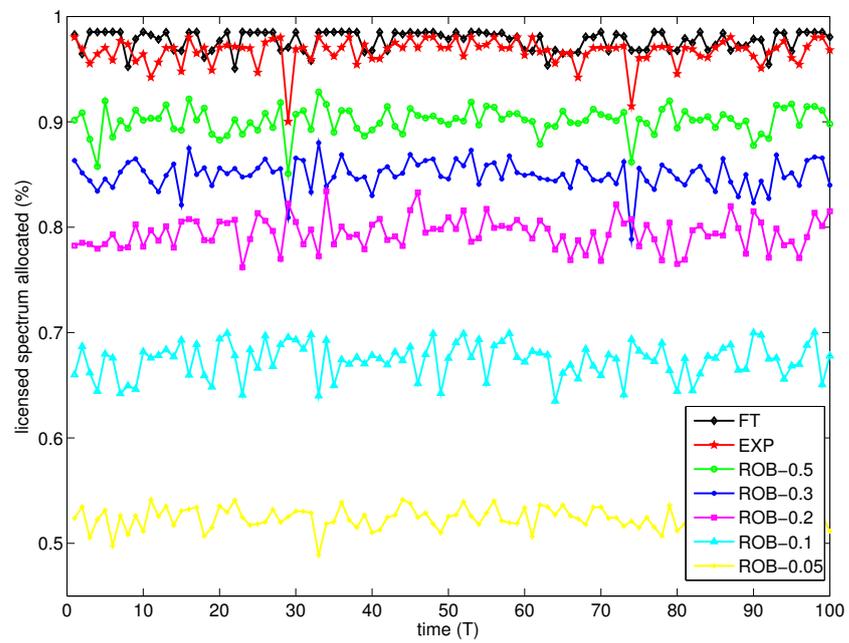


Figure 8.16: Proportion of licensed spectrum allocated for the different methods, when we have a larger bias against unlicensed band capacities.

This page was intentionally left blank.

Conclusions of Part III

The spectrum allocation in a cognitive wireless mesh network (WMN) was studied, considering a mixed scenario with both type of frequency bands, licensed and non licensed. The problem was analyzed from the perspective of SUs, which might use licensed bands whenever available and unlicensed bands all the time. We developed a general stochastic formulation considering a periodically scheduled assignment. We proposed a novel robust approach to solve the problem and analyzed the advantage against an expectation based solution, comparing their performance by extensive simulations. We also presented a decentralized implementation of the proposed framework, allowing the algorithm to scale properly. We believe the proposed solution is suitable for WMN Internet access solutions, as the one aforementioned from Plan Ceibal's schools, in order to meet the capacity requirements they will face in the coming years.

The results show that the proposed solution presents much better performance than the expectation based approach, with not much additional spectrum allocated. The robust approach guarantees the required throughput in each link with very high probability, while with a mean value solution the requirements are not accomplished more than 40 % of the time. The algorithm performance was analyzed in depth, for different situations of the PUs' activity, and the simulation experiments indicate that the extra spectrum required to guarantee the algorithm robustness, is usually below 35 % of that required by an oracle that knows beforehand the PUs' activities. Finally, we simulated the proposed scheme in a real-world network, analyzing the operation for different values of the algorithm parameter ϵ and looking how it should be chosen to meet certain previously specified requirements.

In this part of the thesis we considered a spectrum allocation framework, with fixed *a priori* requirements for each link in the WMN. In our future work, we would like to extend this framework, considering a cross-layer approach, which integrates this framework, with automatic selection of dynamic requirements on each link, based on real-time network demand measurements. It would be also interesting to compare the periodically scheduled allocation proposed with an event-driven solution. We could analyze which is the threshold in the PUs' dynamics when signaling overhead of the latter becomes tolerable to get a more efficient spectrum allocation. We believe that each approach may have advantages and disadvantages, so it would be worth to do a thorough comparison of the two alternatives. Another point that can be explored in depth is the choice of different cost functions for specific requirements. We present a general framework, which is then simulated for

a particular case. It would be interesting to look for other possible cost functions, according to different practical requirements, to analyze how the same framework could be applied and what are the results we can obtain.

A couple of additional points that could also be analyzed in the future, one more theoretical, the other more practical, are on the one hand the robust equivalence and on the other hand the optimization algorithm. While we considered an approach which is robust with respect to the distribution of the PUs' activity, it could be possible to find better equivalences if we know more about it. Maybe in a particular case, with an adequate statistical model of the PUs' dynamics, it is possible to take advantage of this information for a better solution. About the optimization itself, this work was not focused on it and we just used a gradient descent algorithm. Hence, it is possible to look for efficient alternatives, particularly adapted to the proposed scheme.

Finally, the next stage in our line of research would be to implement the algorithm in a real field deployment, for example using WiFi bands as unlicensed spectrum and TV bands as licensed spectrum (e.g. TVWS technologies).

Chapter 9

General Conclusions and Perspectives

Throughout this thesis different topics were studied associated with the analysis, design and optimization of wireless access networks with multiple hops. The issues addressed are important challenges that should be solved for the next generation wireless access networks. In particular, the development and implementation of standards under the novel cognitive radio networks paradigm, despite opening a wide range of new possibilities, it also poses many problems still unsolved. The proposed methods tackle some of these problems and we believe are a promising approach to meet the requirements we will need from the future infrastructure.

We have seen that the volume of traffic we have to be able to manage continues to grow and this ever-increasing demand is inevitable, so the future access networks will have to deal with it. As we saw in the introduction of this thesis, one of the most important changes we are witnessing are in the area of education, which has become an important vertical for the wireless networks industry. The educational revolution is underway and there is no going back, with several one-to-one programs already running around the planet. This massive adoption of ICT initiatives need the deployment of an infrastructure that is able to support it, which poses major technological challenges, particularly for those countries with fewer resources. Something similar happens in other areas where the ICT initiatives have great impact, as the health care system or the development of digital citizenship.

Another aspect which is common thread along this thesis is the incorporation of statistical learning tools in the design and optimization of wireless networks. We highlight the great utility of this kind of techniques, particularly today that we are able to collect tons of data from measurements in real time, with the focus on finding practical solutions adapted for real-world scenarios. Some of the use cases in this thesis are the performance estimation of an 802.11-based wireless network from physical layer measurements or the model inference for the queue of a 802.11 point to point link. These techniques enable to dynamically optimize the allocation of the available resources.

In this thesis the focus was placed on access technologies for rural areas, where wireless mesh networks have already shown to be a viable solution, as demonstrated by several existing deployments. The problem is how to optimize these

networks to meet the increasing demands and be able to cope with the future requirements. In this regard, two different aspects were addressed in our work. On one hand, we worked on the development of traffic engineering tools to maximize the exploitation of network resources, optimizing the packet routing considering the origin-destination flows. On the other hand, based on the novel cognitive radio paradigm, we developed a framework to extend the available resources, taking advantage of the idle licensed spectrum.

Concerning the traffic engineering in a wireless mesh network, a routing and forwarding scheme was developed, based on a queue model for each link, inferred from measurements collected from the network. The statistical model learned from measurements guarantees that the resultant model is the one that best fits to the real link queue behavior. A suitable convex optimization problem was formulated, which enables a decentralized algorithm to reach the solution. This method optimally distributes the end-to-end traffic load for each pair origin-destination over all possible paths, and also allows to solve the gateway selection problem. The advantages of the proposed technique in comparison with other classical dynamic metric routing methods (like the one used by 802.11s) was illustrated by several packet-level simulations. In all cases, we observed a quick adaptation of the method to traffic changes and also an stable operation, avoiding the routing oscillations dynamic metrics can cause.

With respect to the expansion of network resources, a new framework for spectrum allocation in wireless multihop networks was defined, based on the novel cognitive radio paradigm. Taking advantage of the free spaces of licensed bands, but also maintaining the use of unlicensed bands, a suitable stochastic optimization problem was formulated to model the resource allocation issue. A novel solution was introduced, which is robust to the primary users activity. Then, we presented a dual decomposition of the problem, which enables a decentralized implementation of the proposed method. Several simulation experiments were presented, where we highlighted the advantages of the proposed spectrum assignment in comparison with other classical approaches, such as an expectation-based solution. The algorithm performance was analyzed in depth for different PUs' activity level and the results obtained indicate that the amount of spectrum required to guarantee the algorithm robustness is, at most, only 35% more than the one required by an oracle which knows beforehand the PUs' dynamics.

An interesting analysis for future research, is how to combine the two proposed mechanisms from Part II and Part III, to operate simultaneously in the network. That is to say, we should study the interaction between both algorithms, the timescales at which operation is performed and verify if the proposed distributed implementations still conducts to the optimal allocations.

It is clear that there is still a long way to go to see massive deployments of infrastructure based on the novel cognitive radio paradigm. Several challenges still need to be solved, new standards must be approved and the wireless industry will have to develop the necessary equipment. Anyway, the seed seems to be germinating, and various efforts are ongoing in parallel today. A clear example is that lots of research is being devoted to the subject, major technological advances in soft-

were defined radio happen daily, and also many field tests are already underway, such as TVWS trials in several places around the world [142,143].

Many years ago, the emergence of packet switched networks led to the prediction of a convergent future, where all services (data, voice, video) would be transmitted through a single network. It took quite time to become a reality, but that reality we can say today is finally happening. This new paradigm left behind the old networks based on circuit switching, which allowed much less flexibility to manage resources, but certainly guaranteed the end-to-end requirements. Despite the flexibility introduced by packet switched networks, even though it was successful from the beginning for data services, many efforts were needed to implement resource allocation mechanisms to achieve the necessary end-to-end requirements.

Today, with the new paradigm of cognitive networks, we evidence a kind of analogy with the previously described situation. Clearly, this new approach enables much more flexibility for the resource management task, adding the spectrum as a new allocation variable. However, it is also clear that, to handle this new degree of freedom, the necessary mechanisms should be developed in order to maintain the QoS compliance, particularly for applications with the most demanding requirements. We believe that this thesis is a contribution in this direction, and probably, even though the road may be tough, we envision that finally cognitive networks will become a reality, as well as convergent networks are today.

This page was intentionally left blank.

Bibliography

- [1] IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), “IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” Tech. Rep., 2007.
- [2] I. F. Akyildiz, X. Wang, and W. Wang, “Wireless mesh networks: a survey,” *Comput. Netw. ISDN Syst.*, vol. 47, pp. 445–487, March 2005.
- [3] A. Aziz, M. Durvy, O. Dousse, and P. Thiran, “Models of 802.11 multi-hop networks: Theoretical insights and experimental validation,” in *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*, Jan. 2011, pp. 1–10.
- [4] R. Bruno, M. Conti, and A. Pinizzotto, “Routing internet traffic in heterogeneous mesh networks: Analysis and algorithms,” *Perform. Eval.*, vol. 68, pp. 841–858, September 2011.
- [5] Plan Ceibal, “Plan Ceibal: One Laptop per Child implementation in Uruguay.” 2007-2015. [Online]. Available: <http://www.ceibal.org.uy/>
- [6] R. G. Garroppo, S. Giordano, and L. Tavanti, “A joint experimental and simulation study of the IEEE 802.11s HWMP protocol and airtime link metric,” *Int. J. Commun. Syst.*, vol. 25, no. 2, pp. 92–110, Feb. 2012.
- [7] R. M. Abid, T. Benbrahim, and S. Biaz, “IEEE 802.11s wireless mesh networks for last-mile internet access: An open-source real-world indoor testbed implementation,” *Wireless Sensor Network*, vol. 2, no. 10, pp. 725–738, 2010.
- [8] J. Mitola, “Cognitive Radio — An Integrated Agent Architecture for Software Defined Radio,” PhD thesis, Royal Institute of Technology (KTH), Kista, Sweden, May 2000.
- [9] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination function,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
- [10] K. Medepalli and F. A. Tobagi, “Towards performance modeling of IEEE 802.11 based wireless networks: A unified framework and its applications,” in *IEEE INFOCOM*, April 2006, pp. 1–12.

Bibliography

- [11] N. Choi, J. Ryu, Y. Seok, T. Kwon, and Y. Choi, "Optimizing Aggregate Throughput of Upstream TCP Flows over IEEE 802.11 Wireless LANs," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, Sept. 2007, pp. 1–5.
- [12] R. Bruno, M. Conti, and E. Gregori, "Throughput Analysis and Measurements in IEEE 802.11 WLANs with TCP and UDP Traffic Flows," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 2, pp. 171–186, Feb. 2008.
- [13] S. Shioda and M. Komatsu, "Queueing-Model-Based Analysis for IEEE 802.11 Wireless LANs with Non-Saturated Nodes," in *Radio Communications, Alessandro Bazzi (Ed.)*, 2010.
- [14] G. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke, "IEEE 802.11s: The WLAN mesh standard," *IEEE Wireless Commun.*, vol. 17, no. 1, pp. 104–111, Feb. 2010.
- [15] R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks," in *ACM SIGCOMM*, 2004, pp. 133–144.
- [16] Y. Yang, J. Wang, and R. Kravets, "Designing routing metrics for mesh networks," in *WiMesh*, 2005.
- [17] P. Hsiao, A. Hwang, H. T. Kung, and D. Vlah, "Load-balancing routing for wireless access networks," in *IEEE INFOCOM*, 2001, pp. 986–995.
- [18] V. Mhatre, H. Lundgren, F. Baccelli, and C. Diot, "Joint mac-aware routing and load balancing in mesh networks," in *ACM CoNEXT*, 2007, pp. 19:1–19:12.
- [19] Y. Bejerano, S.-J. Han, and A. Kumar, "Efficient load-balancing routing for wireless mesh networks," *Comput. Netw.*, vol. 51, pp. 2450–2466, July 2007.
- [20] E. Alotaibi, V. Ramamurthi, M. Batayneh, and B. Mukherjee, "Interference-aware routing for multi-hop wireless mesh networks," *Comput. Commun.*, vol. 33, pp. 1961–1971, October 2010.
- [21] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on selected areas in communications*, vol. 24, pp. 1452–1463, 2006.
- [22] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *ACM MobiCom*. ACM Press, 2004, pp. 114–128.
- [23] A. Raniwala and T. cker Chiueh, "Architecture and algorithms for an ieee 802.11-based multi-channel wireless mesh network," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 3, march 2005, pp. 2223 – 2234 vol. 3.

- [24] E. Tragos, S. Zeadally, A. Fragkiadakis, and V. Siris, "Spectrum assignment in cognitive radio networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1108–1135, 2013.
- [25] G. Calafiore and L. Ghaoui, "On distributionally robust chance-constrained linear programs," *Journal of Optimization Theory and Applications*, vol. 130, no. 1, pp. 1–22, 2006.
- [26] International Chamber of Commerce, Commission on the Digital Economy, "ICTs and the Internet Can Strengthen Economic Growth and Recovery," July 2012.
- [27] International Telecommunication Union, "Impact of Broadband on the Economy: Research to Date and Policy Issues," April 2012.
- [28] C. Scott, "Does broadband Internet access actually spur economic growth?" *Information and Communications Technology For The Developing World Project, Berkeley University*, December 2012.
- [29] Inter-American Development Bank, *Development Connections: Unveiling The Impact of New Information Technologies*, April 2012.
- [30] International Development Research Centre, *ICT Pathways to Poverty Reduction: Empirical evidence from East and Southern Africa*, 2014.
- [31] "Conectar Igualdad," 2010. [Online]. Available: <http://www.conectarigualdad.gob.ar/>
- [32] "ConnectED initiative," 2013. [Online]. Available: <http://www.whitehouse.gov/issues/education/k-12/connected>
- [33] International Telecommunication Union, Broadband Commission, "The state of broadband 2014: broadband for all." September 2014.
- [34] F. C. Commission, "2015 Broadband Progress Report and Notice of Inquiry on Immediate Action to Accelerate Deployment," February 2015.
- [35] R. Flickenger, C. E. Aichele, C. Fonda, J. Forster, I. Howard, T. Krag, and M. Zennaro, "Wireless networking in the developing world," 2013. [Online]. Available: <http://wndw.net/>
- [36] R. Flickenger, S. Okay, E. Pietrosemoli, M. Zennaro, and C. Fonda, "Very Long Distance Wi-fi Networks," in *Proceedings of the Second ACM SIGCOMM Workshop on Networked Systems for Developing Regions*, ser. NSDR '08. New York, NY, USA: ACM, 2008, pp. 1–6.
- [37] R. K. Patra, "A multi-tier network architecture for long distance rural wireless networks in developing regions," Ph.D. dissertation, EECS Department, University of California, Berkeley, Jun 2009. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-92.html>

Bibliography

- [38] J. Simó, P. Belzarena, and P. Ludeña, *Las redes de telecomunicación basadas en WiLD (IEEE 802.11 modificado para largas distancias)*. CYTED, 2012, pp. 105–134. [Online]. Available: <http://iie.fing.edu.uy/publicaciones/2012/SBL12>
- [39] A. Rendón, P. Ludeña, and A. Martínez, *Tecnologías de la información y las comunicaciones para zonas rurales*. CYTED, 2012.
- [40] S. Misra, S. C. Misra, and I. Woungang, *Guide to Wireless Mesh Networks*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [41] I. S. 802.11-2012, “IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” 2012.
- [42] J. Gálvez, P. Ruiz, and A. Skarmeta, “A Feedback-based Adaptive Online Algorithm for Multi-Gateway Load-Balancing in Wireless Mesh Networks,” in *IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*, June 2010, pp. 1–9.
- [43] F. Li, Y. Fang, F. Hu, and X. Liu, “Load-aware multicast routing metrics in multi-radio multi-channel wireless mesh networks,” *Comput. Netw.*, vol. 55, no. 9, pp. 2150 – 2167, 2011.
- [44] E. Ancillotti, R. Bruno, M. Conti, and A. Pinizzotto, “Load-aware routing in mesh networks: Models, algorithms and experimentation,” *Comput Commun.*, vol. 34, no. 8, pp. 948 – 961, 2011.
- [45] V. Gambiroza, B. Sadeghi, and E. W. Knightly, “End-to-end performance and fairness in multihop wireless backhaul networks,” in *ACM MobiCom*, 2004, pp. 287–301.
- [46] J.-F. Lee and W.-H. Kuo, “Review: Fairness provisioning in multi-hop wireless backhaul networks: Challenges and solutions,” *Comput. Commun.*, vol. 33, pp. 1767–1772, September 2010.
- [47] G. Narlikar, G. Wilfong, and L. Zhang, “Designing multihop wireless backhaul networks with delay guarantees,” *Wirel. Netw.*, vol. 16, pp. 237–254, January 2010.
- [48] J. Zhou and K. Mitchell, “A scalable delay based analytical framework for CSMA/CA wireless mesh networks,” *Comput. Netw.*, vol. 54, pp. 304–318, February 2010.
- [49] B. Wang and K. Liu, “Advances in cognitive radio networks: A survey,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, February 2011.

- [50] IEEE Std 802.22-2011 Standard for Wireless Regional Area Networks (RAN) — Part 22, “Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Policies and procedures for operation in the TV Bands,” July 2011.
- [51] “IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Television White Spaces (TVWS) Operation,” *IEEE Std 802.11af-2013*, pp. 1–198, February 2014.
- [52] FCC Second Memorandum Opinion and Order 10-174, “Unlicensed Operation in the TV Broadcast Bands, Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band,” September 2010.
- [53] B. Ishibashi and R. Boutaba, “The symbiosis of cognitive radio and wireless mesh networks,” in *Guide to Wireless Mesh Networks*, ser. Computer Communications and Networks, S. Misra, S. Misra, and I. Woungang, Eds. Springer London, 2009, pp. 471–496.
- [54] Personal Telco Project, “Wireless Communities,” 2014. [Online]. Available: <https://personaltelco.net/wiki/WirelessCommunities>
- [55] Cisco Systems, “Evolution of municipal wireless networks,” whitepaper, 2006.
- [56] Plan Ceibal, “Equidad, Tecnología y Educación para el desarrollo humano.” 2012.
- [57] Infonetics Research, “Carrier WiFi equipment 2014,” [Online; accessed 10-April-2015]. [Online]. Available: <http://www.infonetics.com/pr/2014/1H14-Carrier-WiFi-Equipment-Market-Highlights.asp>
- [58] RnR Market Research, “Outdoor Wi-Fi Market: Global Advancements, Business Models, Worldwide Market Forecasts and Analysis (2013 – 2018),” 2013.
- [59] Extreme Networks, “Super Bowl XLVIII Stats Infographic,” [Online; accessed 10-April-2015]. [Online]. Available: <http://www.extremenetworks.com/super-bowl-stats/>
- [60] D. Schwab and R. Bunt, “Characterising the use of a campus wireless network,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, 2004, pp. 862–870 vol.2.
- [61] D. Kotz and K. Essien, “Analysis of a campus-wide wireless network,” *Wirel. Netw.*, vol. 11, no. 1-2, pp. 115–133, Enero 2005.

Bibliography

- [62] T. Henderson, D. Kotz, and I. Abyzov, “The changing usage of a mature campus-wide wireless network,” *Comput. Netw.*, vol. 52, no. 14, pp. 2690–2712, Octubre 2008.
- [63] E. Halepovic, C. Williamson, and M. Ghaderi, “Wireless data traffic: a decade of change,” *IEEE Network*, vol. 23, no. 2, pp. 20–26, 2009.
- [64] OLPC, “One Laptop Per Child.” <http://one.laptop.org/>. [Online]. Available: <http://one.laptop.org/>
- [65] Presidencia de la República Oriental del Uruguay, “Decreto EC579. 18 de abril de 2007.” [Online]. Available: http://archivo.presidencia.gub.uy/_Web/decretos/2007/04/EC579_18%2004%202007_00001.PDF
- [66] F. J. Simó Reigadas, P. Osuna García, D. Espinoza, L. Camacho, and R. Quispe, “Application of IEEE 802.11 technology for health isolated rural environments,” in *WCIT 2006*, Santiago de Chile, August 2006.
- [67] B. Raman and K. Chebrolu, “Experiences in using wifi for rural internet in india,” *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 104–110, Jan. 2007.
- [68] F. Larroca and F. Rodríguez, “An Overview of WLAN Performance, Some Important Case-Scenarios and Their Associated Models,” *Wireless Personal Communications*, vol. 79, no. 1, pp. 131–184, 2014.
- [69] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” *Artificial Intelligence in Medicine*, vol. 23, pp. 89–109, 2001.
- [70] T. Fawcett and F. Provost, “Combining data mining and machine learning for effective user profiling.” AAAI Press, 1996, pp. 8–13.
- [71] T. T. Nguyen and G. Armitage, “A survey of techniques for internet traffic classification using machine learning,” *Commun. Surveys Tuts.*, vol. 10, no. 4, pp. 56–76, Oct. 2008.
- [72] T. Claveirole and M. D. de Amorim, “Wipal and wscout, two hands-on tools for wireless packet traces manipulation and visualization,” in *ACM Mobicom Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization*, 2008.
- [73] A. Lecointre, D. Dragomirescu, and R. Plana, “New methodology to design advanced mb-iruwb communication system,” *IEE Electronics Letters*, vol. 11, 2008.
- [74] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. Singh, “Overview of the orbit radio grid testbed for evaluation of next-generation wireless network protocols,” in *Wireless Communications and Networking Conference, 2005 IEEE*, vol. 3, 2005, pp. 1664–1669 Vol. 3.

- [75] V. Vapnik, S. E. Golowich, and A. Smola, “Support vector method for function approximation, regression estimation, and signal processing,” in *Advances in Neural Information Processing Systems 9*, vol. 9, 1997.
- [76] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, “Top 10 algorithms in data mining,” *Knowl. Inf. Syst.*, vol. 14, no. 1, Dec. 2007.
- [77] N. Sapankevych and R. Sankar, “Time series prediction using support vector machines: A survey,” *Computational Intelligence Magazine, IEEE*, vol. 4, no. 2, pp. 24–38, 2009.
- [78] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, Aug. 2004.
- [79] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992. [Online]. Available: <http://www.jstor.org/stable/2685209>
- [80] C. G. Atkeson, A. W. Moore, and S. Schaal, “Locally weighted learning,” *Artif. Intell. Rev.*, vol. 11, no. 1-5, pp. 11–73, Feb. 1997.
- [81] A. Aamodt and E. Plaza, “Case-based reasoning: Foundational issues, methodological variations, and system approaches,” *AI Commun.*, vol. 7, no. 1, pp. 39–59, Mar. 1994.
- [82] Avila GW2348-4 Network Processor, “A full featured quad Mini-PCI Intel XScale IXP425/533MHz processor for enterprise and residential network applications.” <http://www.gateworks.com/product/item/avila-gw2348-4-network-processor>, 2015. [Online]. Available: <http://www.gateworks.com/product/item/avila-gw2348-4-network-processor>
- [83] “OpenWrt: Linux distribution for embedded devices,” [Online; accessed 10-May-2015]. [Online]. Available: <http://openwrt.org/>
- [84] “Atheros ath5k driver.” [Online]. Available: <https://wireless.wiki.kernel.org/en/users/drivers/ath5k>
- [85] “Iperf: The TCP/UDP Bandwidth Measurement Tool,” [Online; accessed 10-May-2015]. [Online]. Available: <http://iperf.fr/>
- [86] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [87] Grupo Radar, “Perfil del Internauta Uruguayo,” 10a edición, Setiembre 2013.

Bibliography

- [88] A. Khanna and J. Zinky, “The revised ARPANET routing metric,” *SIGCOMM Comput. Commun. Rev.*, vol. 19, pp. 45–56, August 1989.
- [89] K. Ramachandran, I. Sheriff, E. Belding, and K. Almeroth, “Routing stability in static wireless mesh networks,” in *PAM*, 2007, pp. 73–83.
- [90] M. Caesar, M. Casado, T. Koponen, J. Rexford, and S. Shenker, “Dynamic route recomputation considered harmful,” *SIGCOMM Comput. Commun. Rev.*, vol. 40, pp. 66–71, April 2010.
- [91] P. P. Pham and S. Perreau, “Increasing the network performance using multi-path routing mechanism with load balance,” *Ad Hoc Networks*, vol. 2, no. 4, pp. 433 – 459, 2004.
- [92] F. Larroca and J.-L. Rougier, “Minimum delay load-balancing via nonparametric regression and no-regret algorithms,” *Computer Networks*, vol. 56, no. 4, pp. 1152 – 1166, 2012.
- [93] B. Raman, K. Chebrolu, D. Gokhale, and S. Sen, “On the feasibility of the link abstraction in wireless mesh networks,” *Networking, IEEE/ACM Transactions on*, vol. 17, no. 2, pp. 528 –541, April 2009.
- [94] J. J. Gálvez, P. M. Ruiz, and A. F. Gómez-Skarmeta, “Responsive on-line gateway load-balancing for wireless mesh networks,” *Ad Hoc Networks*, vol. 10, no. 1, pp. 46–61, 2012.
- [95] S. Avallone and G. Di Stasi, “A new MPLS-based forwarding paradigm for multi-radio wireless mesh networks,” *Wireless Communications, IEEE Transactions on*, vol. 12, no. 8, pp. 3968–3979, 2013.
- [96] P. Casas, F. Larroca, J.-L. Rougier, and S. Vaton, “Taming traffic dynamics: Analysis and improvements,” *Comput. Commun.*, vol. 35, no. 5, pp. 565–578, Mar. 2012.
- [97] T. Kuosmanen, “Representation theorem for convex nonparametric least squares,” *Econometrics Journal*, vol. 11, no. 2, pp. 308–325, 2008.
- [98] “The MOSEK optimization software.” [Online]. Available: <http://www.mosek.com/>
- [99] A. Jindal and K. Psounis, “The achievable rate region of 802.11-scheduled multihop networks,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1118–1131, Aug. 2009.
- [100] D. J. Leith, V. G. Subramanian, and K. R. Duffy, “Log-convexity of rate region in 802.11e w lans,” *Comm. Letters.*, vol. 14, no. 1, pp. 57–59, Jan. 2010.
- [101] D. J. Leith, Q. Cao, and V. G. Subramanian, “Max-min fairness in 802.11 mesh networks,” *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 756–769, 2012.

- [102] T. Salonidis, G. Sotiropoulos, R. Guerin, and R. Govindan, “Online optimization of 802.11 mesh networks,” in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, ser. CO-NEXT '09, 2009, pp. 61–72.
- [103] “The ns-3 Network Simulator.” [Online]. Available: <http://www.nsnam.org/releases/>
- [104] F. Li, M. Li, R. Lu, H. Wu, M. Claypool, and R. E. Kinicki, “Measuring queue capacities of IEEE 802.11 wireless access points,” in *BROADNETS*, 2007, pp. 846–853.
- [105] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, “Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability,” *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.
- [106] M. Chiang, S. Low, A. Calderbank, and J. Doyle, “Layering as optimization decomposition: A mathematical theory of network architectures,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan 2007.
- [107] Y. Yi and M. Chiang, “Stochastic network utility maximization,” *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 421–442, June 2008.
- [108] D. Bertsekas and R. Gallager, *Data Networks (2Nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1992.
- [109] A. Elwalid, C. Jin, S. Low, and I. Widjaja, “MATE: MPLS adaptive traffic engineering,” in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, 2001, pp. 1300–1309 vol.3.
- [110] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [111] Y. Zhang, J. Luo, and H. Hu, *Wireless Mesh Networking: Architectures, Protocols and Standards*, ser. Wireless Networks and Mobile Communications. CRC Press, 2006.
- [112] G. Carneiro, P. Fortuna, and M. Ricardo, “Flowmonitor: a network monitoring framework for the network simulator 3 (ns-3),” in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, ser. VALUETOOLS '09, 2009, pp. 1:1–1:10.
- [113] Cisco White Paper, “Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018,” February 2014.
- [114] FCC Notice of Proposed Rulemaking 13-22, “Revision of Part 15 of the Commission’s Rules to Permit Unlicensed National Information Infrastructure (U-NII) Devices in the 5 GHz Band,” February 2013.

Bibliography

- [115] Craig Mathias, “Rethinking Spectrum Scarcity: Database-Driven Cognitive Radio,” *A Farpoint Group White Paper*, September 2010.
- [116] FCC Public Notice DA-1472, “Office of Engineering and Technology Announces the Approval of Google, Inc.’s TV Bands Database System for Operation,” June 2013.
- [117] F. Novillo, M. Churchman, R. Ferrus, and R. Agusti, “A channel allocation algorithm for OSA-enabled IEEE 802.11 WLANs,” in *International Symposium on Wireless Communication Systems*, September 2009.
- [118] F. Liu, E. Erkip, M. C. Beluri, R. Yang, and E. Bala, “Dual-band femto-cell traffic balancing over licensed and unlicensed bands,” in *International Conference on Communications*, 2012.
- [119] E. Ahmed, A. Gani, S. Abolfazli, L. Yao, and S. Khan, “Channel assignment algorithms in cognitive radio networks: Taxonomy, open issues, and challenges,” *Communications Surveys & Tutorials, IEEE*, vol. PP, no. 99, pp. 1–1, 2014.
- [120] J. Crichigno, M.-Y. Wu, and W. Shu, “Protocols and architectures for channel assignment in wireless mesh networks,” *Ad Hoc Networks*, vol. 6, no. 7, pp. 1051 – 1077, 2008.
- [121] R. Soua and P. Minet, “Multichannel assignment protocols in wireless sensor networks: A comprehensive survey,” *Pervasive and Mobile Computing*, no. 0, pp.–, 2014.
- [122] A. Prodan and V. Mirchandani, “Channel assignment techniques for 802.11-based multiradio wireless mesh networks,” in *Guide to Wireless Mesh Networks*, ser. Computer Communications and Networks, S. Misra, S. Misra, and I. Woungang, Eds. Springer London, 2009, pp. 119–146.
- [123] S. Avallone and I. F. Akyildiz, “A channel assignment algorithm for multi-radio wireless mesh networks,” *Computer Communications*, vol. 31, no. 7, pp. 1343 – 1353, 2008, special Issue: Resource Management and routing in Wireless Mesh Networks.
- [124] W. Si, S. Selvakennedy, and A. Y. Zomaya, “An overview of channel assignment methods for multi-radio multi-channel wireless mesh networks,” *Journal of Parallel and Distributed Computing*, vol. 70, no. 5, pp. 505 – 524, 2010.
- [125] T. K. Thuc, S. Guruacharya, and D. Niyato, “Robust bandwidth allocation in wireless mesh network,” in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 1704–1709.
- [126] X. Shao, C. Hua, and A. Huang, “Robust resource allocation for multi-hop wireless mesh networks with end-to-end traffic specifications,” *Ad Hoc Networks*, vol. 13, Part A, no. 0, pp. 123 – 133, 2014.

- [127] J. Wellons and Y. Xue, “The robust joint solution for channel assignment and routing for wireless mesh networks with time partitioning,” *Ad Hoc Networks*, vol. 13, Part A, no. 0, pp. 210 – 221, 2014.
- [128] Y. Song, C. Zhang, and Y. Fang, “Stochastic traffic engineering in multihop cognitive wireless mesh networks,” *IEEE Transactions on Mobile Computing*, vol. 9, no. 3, pp. 305–316, March 2010.
- [129] M. Timmers, S. Pollin, A. Dejonghe, L. V. der Perre, and F. Catthoor, “A distributed multichannel mac protocol for multihop cognitive radio networks,” *IEEE Trans. on Vehicular Technology*, vol. 59, no. 1, pp. 446–459, 2010.
- [130] A. Kashyap, S. Ganguly, and S. R. Das, “A measurement-based approach to modeling link capacity in 802.11-based wireless networks,” in *International Conference on Mobile Computing and Networking*, 2007.
- [131] J. Guerin, S. Glass, P. Hu, W. L. Tan, and M. Portmann, “Time-based and low-cost bandwidth estimation for IEEE 802.11 links,” in *International Wireless Communications and Mobile Computing Conference*, August 2012.
- [132] S. Uryasev and P. M. Pardalos, *Stochastic Optimization: Algorithms and Applications*. Springer US, 2001.
- [133] A. Prékopa, *Stochastic Programming*. Springer Netherlands, 1995.
- [134] A. Charnes and W. W. Cooper, “Chance-constrained programming,” *Management Science*, vol. 6, no. 1, pp. 73–79, 1959.
- [135] D. Bertsimas and I. Popescu, “Optimal inequalities in probability theory: A convex optimization approach,” *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 780–804, 2005.
- [136] C. Lagoa, X. Li, and M. Sznaier, “Application of probabilistically constrained linear programs to risk-adjusted controller design,” in *American Control Conference, 2001. Proceedings of the 2001*, vol. 2, 2001, pp. 738–743 vol.2.
- [137] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1439–1451, 2006.
- [138] C. Bron and J. Kerbosch, “Algorithm 457: Finding all cliques of an undirected graph,” *Commun. ACM*, vol. 16, no. 9, pp. 575–577, Sep. 1973.
- [139] A. Plummer, T. Wu, and S. Biswas, “A cognitive spectrum assignment protocol using distributed conflict graph construction,” in *Military Communications Conference, 2007. MILCOM 2007. IEEE*, Oct 2007, pp. 1–7.

Bibliography

- [140] M. Lopez-Benitez and F. Casadevall, “Discrete-time spectrum occupancy model based on markov chain and duty cycle models,” in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, May 2011, pp. 90–99.
- [141] Y. Saleem and M. H. Rehmani, “Primary radio user activity models for cognitive radio networks: A survey,” *Journal of Network and Computer Applications*, vol. 43, no. 0, pp. 1 – 16, 2014.
- [142] J. Mack and J. Cartmell, “Field trial results for a Wi-Fi based spectrum sharing technology in TVWS,” in *Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island*, May 2014, pp. 1–6.
- [143] A. Lysko, M. Masonta, M. Mofolo, L. Mfupe, L. Montsi, D. Johnson, F. Mekuria, D. Ngwenya, N. Ntlatlapa, A. Hart, C. Harding, and A. Lee, “First large TV white spaces trial in South Africa: A brief overview,” in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2014 6th International Congress on*, Oct 2014, pp. 407–414.

Esta es la última página.
Compilado el domingo 2 agosto, 2015.
<http://iie.fing.edu.uy/~gcapde/tesis.pdf>