MODELLING THE IEEE 802.11 PROTOCOL IN WIRELESS MULTI-HOP NETWORKS

THÈSE N^{O} 3950 (2007)

PRÉSENTÉE LE 23 NOVEMBRE 2007

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

Laboratoire pour les communications informatiques et leurs applications 3

SECTION DES SYSTÈMES DE COMMUNICATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mathilde DURVY

ingénieure en systèmes de communication diplômée EPF
et de nationalité française

acceptée sur proposition du jury:

Prof. M. Hasler, président du jury
Prof. P. Thiran, directeur de thèse
Prof. F. Baccelli, rapporteur
Prof. F. Kelly, rapporteur
Prof. J.-Y. Le Boudec, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL
2007

Abstract

IEEE 802.11 is probably the most widely used, medium access control protocol in current wireless networks. In the Wireless LAN (i.e., single-hop) setting, its performance is by now quite well understood. However, in the multi-hop setting where relay nodes are used to achieve end-to-end communication, there is, to date, no widely accepted model. Consequently, when confronted with experimental results, people often find it hard to interpret them.

The goals of this thesis are (i) to model protocols "à la 802.11" in the context of multi-hop ad hoc networks, (ii) to derive theoretical limits for their performance, (iii) to contrast the performance of the current IEEE 802.11 protocol with these limits and (iv) to identify all the factors that prevent IEEE 802.11 from reaching these limits.

Most of this thesis is dedicated to achieving the two first goals. We begin by proposing an idealized version of IEEE 802.11. We model this idealized protocol using a continuous Markov chain. We then use the properties and the stationary distribution of this Markov chain to derive the performance of the idealized 802.11 protocol.

We first look at its spatial reuse or, in other words, at its ability to schedule a large number of concurrent successful transmissions. We show that the idealized 802.11 protocol organizes the transmissions in space in such a way that it leads to an optimal spatial reuse when its access intensity is large. This is encouraging, as it shows that a protocol using only local interactions can find a global optimum in a completely decentralize way.

We then consider the short and long-term fairness properties of the idealized 802.11 protocol. We observe a clear trade-off between its spatial reuse and its fairness. At low access intensities, its fairness is high but its spatial reuse is low; whereas at high access intensities, the reverse is true. As a result, the access intensity of the protocol can be used to adapt its performance to fit the requirements of the applications running on top of it. The fairness performance of 802.11 also highly depends on the underlying network topology – 802.11 only amplifies the existing topological inequalities. In regular lattice topologies these inequalities arise only at the border where the nodes have fewer neighbors than the nodes inside the network. We demonstrate that, in large line networks and for all finite

access-intensities, this border effect does not propagate inside the network, as a result 802.11 is fair. In contrast, we demonstrate that in large grid topologies a phase transition occurs. Under a certain access intensity, the border effect fades away; whereas above a certain access intensity, it propagates throughout the network, and the protocol is severely unfair.

Finally, after extending our model to consider different node sensing and capture capabilities, we compare the performance of the ns-2 implementation of IEEE 802.11 and of the idealized protocol. We observe a large gap between the theoretical and practical performance. We identify the three problems that are responsible for this gap. We then propose a remedy to address each of these problems, and show that a 'cured' IEEE 802.11 can achieve the level of performance of the idealized 802.11 protocol.

Keywords

Medium access control, multi-hop ad hoc networks, modeling, IEEE 802.11, spatial reuse, fairness, starvation phenomenon, phase transition.

Résumé

IEEE 802.11 est probablement le protocole de gestion d'accès au canal le plus utilisé dans les réseaux sans-fil d'aujourd'hui. Ses performances dans les réseaux locaux sont maintenant bien connues. En revanche, dans les réseaux à sauts multiples où une communication doit utiliser des nœuds relais pour arriver à sa destination finale, aucun modèle n'est pour l'instant reconnu. Par conséquent, il reste difficile d'interpréter les résultats expérimentaux obtenus pour ce type de réseaux.

Les objectifs de cette thèse sont (i) la modélisation de protocoles "à la 802.11" dans les réseaux ad hoc à sauts multiples, (ii) l'obtention de limites théoriques pour leurs performances, (iii) la comparaison entre ces limites et les performances actuelles du protocole IEEE 802.11 et (iv) l'identification des facteurs qui empêchent IEEE 802.11 d'opérer au niveau des limites théoriques.

Cette thèse est consacrée en grande partie aux deux premiers objectifs. La première étape consiste à modéliser une version idéalisée du protocole 802.11 en ayant recours à une chaîne de Markov à temps continu. Les caractéristiques et la distribution stationnaire de cette chaîne de Markov sont ensuite utilisées pour obtenir les performances du protocole 802.11.

Dans un premier temps, nous évaluons la densité spatiale des transmissions sous le protocole 802.11 idéalisé ou, en d'autre termes, sa capacité à programmer en parallèle un grand nombre de transmissions réussies. Nous montrons que ce protocole organise les transmissions dans l'espace de manière à ce que leur densité soit maximale lorsque son intensité d'accès au canal est élevée. Ce résultat est encourageant car il montre qu'un protocole basé sur des interactions locales peut atteindre un optimum global de façon complètement décentralisée.

Dans un deuxième temps, nous analysons l'équité du protocole 802.11 idéalisé à long et à court terme. Il existe un net compromis entre le niveau d'équité et la densité spatiale des transmissions obtenus. Quand l'intensité d'accès au canal est faible, le protocole traite les nœuds du réseau de façon équitable mais le nombre de transmissions simultanées est faible. La situation s'inverse quand l'intensité d'accès au canal est élevée. L'intensité d'accès du protocole peut donc être utilisée comme un moyen d'adapter ses performances aux besoins des applications utilisant le réseau. Cependant, l'équité du protocole 802.11 dépend aussi fortement de la topologie du réseau – 802.11 ne fait qu'accentuer les inégalités topologiques

déjà existantes. Dans les topologies régulières, ces inégalités sont présentes exclusivement au bord du réseau où les nœuds ont un nombre réduit de voisins. Nous démontrons que dans les grands réseaux en ligne et pour toute intensité d'accès au canal finie, cet effet de bord ne se propage pas à l'intérieur du réseau et que par conséquent le protocole est équitable. Par contre, nous démontrons que dans les grands réseaux en grille il y a une transition de phase. En dessous d'une certaine intensité d'accès, l'effet de bord disparaît à l'intérieur du réseau, mais au dessus d'une certaine intensité d'accès il persiste et le protocole est très inéquitable.

Enfin, après avoir étendu notre modèle pour considérer des nœuds ayant différentes capacités de détection et de capture, nous comparons les performances du protocole IEEE 802.11 implémenté dans le simulateur ns-2 à celles de sa version idéalisée. Nous observons que les performances pratiques et théoriques diffèrent grandement et nous identifions les trois problèmes responsables de cette différence. Nous proposons un remède à chacun de ces problèmes et montrons qu'un protocole IEEE 802.11 'guéri' peut atteindre le niveau de performance du protocole idéalisé.

Mots clés

Protocole de gestion d'accès au canal, réseaux ad hoc à sauts multiples, modélisation, IEEE 802.11, densité spatiale des transmissions, équité, transition de phase.

Acknowledgments

I want to thank my advisor, Professor Patrick Thiran, for convincing me to do a PhD and for making it such an enriching experience. I am especially grateful to him for giving me the freedom to explore and to find my own research topic. Despite his name I found him an easygoing, pleasant, and always helpful advisor.

In addition, I want to thank all EPFL professors whose teaching and advice I very much appreciated.

It was a great pleasure and a humbling experience to have Professors François Baccelli, Frank Kelly, and Jean-Yves Le Boudec in my jury. I want to thank them for accepting to review this thesis and for the interest they demonstrated with respect to my work.

Next, I would like to thank all the people in my lab for making it a friendly and lively working place, especially the lunch+coffee break crowd. I also want to acknowledge Catherine Boutremans my first office mate, Olivier Dousse and Ruben Merz for their friendship and help in every aspect of life, including eating my lunch!

I am very obliged to the lab's staff, Danielle Alvarez, Holly Cogliati, Angela Devenoge, Philippe Chammartin, Jean-Pierre Dupertuis, and last but not least Marc-André Lüthi for making everything work so smoothly. Special thanks also to Sylviane Dal Mas, the 'mum' of the communication systems section.

Finally, my gratitude goes to my family and to my not so secret admirer for love, support, and encouragement during all my studies.

Contents

Chapter 1

Introduction

1.1 Motivation

Today, most wireless networks are wireless local area networks (WLAN) operating the IEEE 802.11 WiFi protocol in infrastructure mode. The infrastructure consists of a base station (also called an access point), typically owned by a company or a network operator. Such networks are centralized and the base station controls the access to the communication channel. The base station is also used to connect the WLAN to the company network or to the Internet. IEEE 802.11 offers a second mode, the ad hoc mode. It is seldomly used. In this mode, there is no base station, and devices within the same WLAN communicate directly (i.e., in a single hop).

An extension of WLAN operating in ad hoc mode are multi-hop ad hoc networks. They are typically deployed in large areas. In these networks, some devices might not be able to communicate directly to each other because of their limited radio range. In such cases, intermediary devices act as relays. In other words, the communication goes through multiple hops before reaching its final destination.

Multi-hop ad hoc networks do not require any fixed infrastructure, consequently they are easy to deploy. Also, they offer a potential throughput gain. These characteristics make multi-hop ad hoc networks a promising technology.

Because of their decentralized nature, multi-hop ad hoc networks offer additional challenges that were not necessarily envisioned during the initial design of current Medium Access Control (MAC) protocols. The MAC protocol is run locally by each network device. Its role is to regulate the access of the devices to a shared resource, the communication channel. The fundamental access method of IEEE 802.11 MAC is known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). In CSMA/CA a device wishing to transmit first senses the channel to detect other active devices. If no activity is detected on the channel, the device transmits. Otherwise, its transmission is deferred for a random backoff time. The device uses a timer to count down this waiting time. The value of this timer is decremented only when the device does not detect any activity on the channel. This random backoff is used to reduce the collision probability when the

channel becomes idle after a long period of activity. It avoids that all the devices that were waiting to transmit do so at the same time.

In single-hop networks, a device can accurately detect the activity of all other devices and thus appropriately defer its transmissions. This means that only one successful transmission can occur at any given time. In multi-hop ad hoc networks, a device cannot detect the activity of all the other devices. This gives rise to the well-known hidden node problem. A node (or device) is hidden (from a transmitting node) if it cannot detect its transmission but can create a collision at the intended receiver. Yet, in multi-hop networks, it is typically possible and desirable to have several non-colliding transmissions at the same time.

A good MAC protocol should schedule a large number of concurrent successful transmissions, but it should also guarantee an equal access to the communication channel to each network node. In other words, a good MAC protocol should provide a high *spatial reuse* and a *fair* channel access. These two properties are very hard to achieve in practice. Indeed, detailed experimental studies of IEEE 802.11 in multi-hop ad hoc networks (see for example [GGK01] and [CDL05]) reveal poor performances, especially with high traffic loads. As a consequence, the suitability of IEEE 802.11 as a medium access control protocol for multi-hop ad hoc networks has been questioned [XS01, SHS04].

The lack of analytical models makes it difficult to understand the behavior of IEEE 802.11 in multi-hop ad hoc networks. Existing models concentrate on the WLAN (i.e., single-hop) setting and often rely explicitly on the assumption that nodes can detect all transmissions in the network or, in other words, that there is no hidden node. Clearly, the definition of multi-hop wireless networks invalidates this assumption, as most nodes are hidden from a given node. As a result, the quest is still open for an accurate model for IEEE 802.11 in multi-hop ad hoc networks. In particular, there is a need for models that explain the observed performance of the protocol, that identify its key parameters, and that can be used to derive its theoretical limits. Moreover, these models should be flexible enough to take into account the capabilities of the network devices and to provide insight for small, as well as large, multi-hop networks.

1.2 Outline

We first present IEEE 802.11 together with an idealized version of this protocol in Chapter 2. The idealized 802.11 protocol retains the key features of the IEEE 802.11 protocol while abstracting all implementation-specific details.

In Chapter 3, we introduce the Markovian framework that we use to model the dynamics of the idealized 802.11 protocol. The states of the Markov chain are the transmission patterns allowed by the protocol and a transition between two states correspond to the start or the completion of a transmission. We take special care to relate our model to other existing models.

In Chapters 4, 5, and 6, we use our Markovian framework to evaluate the per-

formance of the protocol according to different metrics. We first evaluate its spatial reuse, i.e., its ability to schedule a large number of concurrent transmissions, which increases throughput (Chapter 4). We compare the spatial reuse achieved by the unslotted 802.11 protocol with the spatial reuse of some slotted protocols and show that its unslotted nature helps 802.11 to organize the transmissions in the network. We then evaluate its ability to give an equal channel access to the network nodes, which is important for fairness. We distinguish fairness in the long-term (Chapter 5) from fairness in the short-term (Chapter 6). In particular, we study how the underlying network topology and the access intensity of the protocol affect its fairness performance.

In Chapter 7, we extend our analysis to incorporate network devices with different sensing and capture capabilities. The sensing capability of a device refers to its ability to detect an ongoing transmission, whereas its capture capability refers to its ability to decode a message (or packet) in the presence of interfering transmissions.

In Chapter 8, we validate our analytical model using ns-2, which is probably the most widely used network simulator in the research community. In particular, we identify the different steps needed to bring the performance of the current IEEE 802.11 protocol to the level of performance achieved by the idealized protocol.

Finally, we discuss the limitations of our results, present possible extensions and conclude in Chapter 9.

1.3 Contributions

These are the main contributions of the thesis.

- We propose a class of continuous Markov chains to model an idealized version of IEEE 802.11.
- We observe that the *access intensity* (i.e., the ratio of the average exchange and backoff time) plays a key role in the performance of the protocol.
- We show that an idealized 802.11 protocol organizes the transmissions in space in such a way that it leads to an optimal spatial reuse when the access intensity is large.
- We derive exact formulæ for the spatial reuse of the idealized protocol in large line networks.
- We show that the observed unfairness of 802.11 is essentially due to the unfair advantage of some of the network nodes, those that have a restricted neighborhood and thus a higher probability to access the communication channel. For example, in lattice topologies nodes at the border of the network are favored.

- For large line networks we show that this border effect does not propagate inside the network and that nodes sufficiently far away from the border have an equal access to the channel; as a result the protocol is long-term fair.

- In contrast, we demonstrate that in large grid networks a phase transition occurs. If the access intensity of the protocol is small, the border effect remains local and the protocol behaves similarly as in one-dimensional networks. However, if the access intensity of the protocol is large enough, the border effect persists independently of the size of the network and the protocol is highly unfair.

- We show that the performance of 802.11 also strongly depends on the capture and sensing capabilities of the nodes. In particular, we show that an idealized 802.11 protocol does suffer from severe unfairness (in other words, starvation) when the receiving and sensing ranges are equal, but quite surprisingly that this unfairness is reduced or even disappears when these two ranges are sufficiently different.

- We identify three problems in the contention resolution mechanism of IEEE 802.11, and we show that they account for most of the gap separating the actual and optimal performance of the protocol. For each of the problems, we propose a solution that, once implemented, allows us to quantify the impact of the problem on the performance of IEEE 802.11. Finally, we show that reducing the overhead of the protocol to some negligible quantity brings its spatial reuse to the fundamental limits set by its physical layer.

Chapter 2

IEEE 802.11 Networks

The IEEE 802.11 standard [iee99] specifies a medium access control layer and several possible physical layers. IEEE 802.11 can operate in two modes: infrastructure and ad hoc. The infrastructure mode is a centralized mode where an access point regulates the access to the channel, whereas the ad hoc mode is fully decentralized. In this work we study networks running IEEE 802.11 in ad hoc mode.

2.1 Physical Layer

IEEE 802.11 provides different physical layers: frequency-hopping spread spectrum, direct sequence spread spectrum, infrared, etc. Our main interest lies not in the specificities of these different physical layers but rather on how the MAC layer perceives the physical layer.

2.1.1 Model

We model the physical layer using two parameters: the receiving range and the carrier sensing range.

The *receiving range (RXRange)* is the maximum distance from the source at which a packet can be successfully received in the absence of interfering nodes. The RXRange depends on the rate at which a packet is sent. In IEEE 802.11, the control packets are always sent at the minimum rate, but the data packets can be sent at a higher rate. We use the RXRange to denote the receiving range of control packets.

The *carrier sensing range (CSRange)* is the maximum distance from the emitter at which a transmission can still be detected. The CSRange is always larger than the RXRange. All packets received from nodes within CSRange of a node are forwarded to its MAC layer, but only packets received from nodes within its RXRange can be successfully processed.

Collisions and capture phenomena are detected as follows. Assume a node is currently processing a packet sent by a node in its RXRange and that a new node in

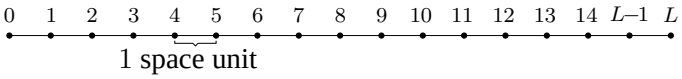


Figure 2.1: Line topology.

its CSRange starts transmitting. The new transmission interferes with the ongoing transmission. If the signal-to-interference ratio is still high enough for the receiver to keep decoding its packet despite the new interferer, we say that there is a *capture effect*, otherwise we say that there is a *collision*.

2.1.2 Network Topologies

We represent the network topology as a graph whose vertices represent the nodes and where there is an edge between two nodes if they are in RXRange of each other. An edge corresponds in fact to two directed links, as a transmission can be initiated by either of its end-nodes.

Throughout this work we consider three topologies: two regular topologies, and one irregular topology. In practice, regular topologies occur when the position of the network nodes can be controlled and is fixed. This is typically the case for indoor networks or more generally for sensor networks. Irregular topologies often occur in situations where a careful deployment of the network nodes is not possible. For example if the network nodes are thrown from a plane during a military operation or if the nodes are mobile. In all the topologies studied the RXRange is equal to 1 space unit.

The first regular topology is a *line* topology (Figure 2.1) where $L + 1$ nodes (numbered from 0 to L) are equally spaced (by 1 space unit) along a straight line. The second regular topology is a *grid* topology where the distance between two neighboring nodes is again 1 space unit. Figure 2.2(a) depicts (as an example) a 17×17 grid topology. The *irregular* topology is represented in Figure 2.2(b). To generate this topology 1065 nodes were uniformly deployed on a 26×26 area, the isolated nodes were then removed to keep a connected component of 970 nodes.

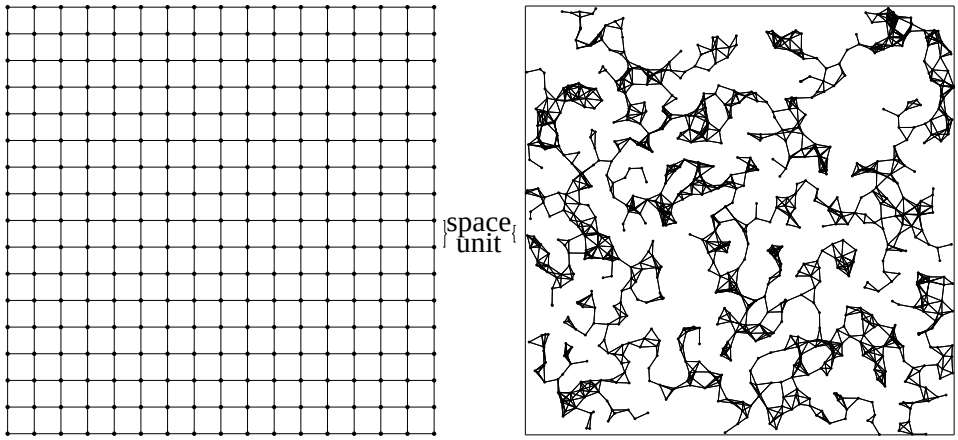
We denote L the number of edges in a topology (or graph). The number of directed links is thus $2L$.

2.2 MAC Protocols

2.2.1 IEEE 802.11

We provide a high-level overview of IEEE 802.11 in ad hoc mode. We concentrate on the features that are relevant to the next chapters. Many details are omitted and can be found in the IEEE 802.11 standard [iee99].

Figure 2.3 illustrates a typical exchange in IEEE 802.11. The sender first transmits a *"request to send"* packet (RTS), to which the receiver answers with a *"clear*



(a) Grid topology (b) Irregular topology

Figure 2.2: Two-dimensional topologies

to send" packet (CTS)[1]. Both packets contain information about the remaining duration of the exchange, so that all overhearing nodes can update the value of their *network allocation vector* (NAV) (i.e., set it to the time at which the exchange will be completed) and refrain from emitting during the exchange. The RTS-CTS handshake is followed by the actual DATA transmission and the exchange is terminated by an acknowledgement (ACK) packet.

The RTS-CTS handshake is used to avoid the so-called *hidden node problem* which is a well-known issue specific to multi-hop networks. Figure 2.4 describes this problem in detail.

In IEEE 802.11, a node intending to transmit senses first the medium. Physical and virtual carrier sensing mechanisms are used to determine the state of the medium. The physical carrier sensing is provided by the physical layer. The virtual carrier sensing is done at the MAC layer. It declares the medium busy if the node's NAV value is larger than the current time. If both carrier sensing mechanisms declare the medium idle, the node transmits, otherwise it backs-off.

To prevent collisions, each node maintains a *backoff timer*, which is initialized to a random value chosen according to some *backoff distribution*. IEEE 802.11 uses an uniform discrete backoff distribution in $\{0, 1, \ldots, cw\} \times 20\mu s$, where the integer cw is the contention window. cw is initially equal to $CW_{min} = 31$, and is doubled every time a packet sent by the node experiences a collision, up to the limit $CW_{max} = 1023$; it is reset to CW_{min} after a successful transmission by the node. The backoff timers run when the channel is idle; when a node senses the

[1]The RTS-CTS handshake is optional but is typically used if the DATA packets are above a certain size. In this work, we assume IEEE 802.11 with RTS-CTS handshake.

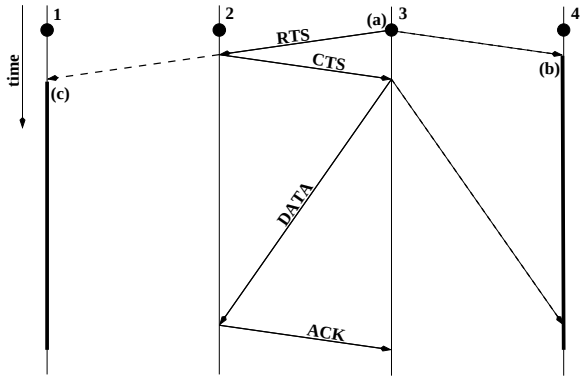


Figure 2.3: A typical exchange between Node 3 and Node 2. At point (a) the backoff timer of Node 3 reaches zero, Node 3 sends a RTS packet to Node 2. Upon receiving the RTS (b) (respectively CTS (c)) Node 4 (resp., Node 1) sets its NAV (bold line) to cover the duration of the exchange between Node 3 and Node 2.

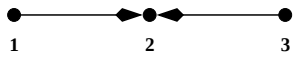


Figure 2.4: Example of a hidden node situation. Node 1 and 3 can communicate with Node 2 (they are in its RXRange). However, if Node 1 and 3 are not in CSRange of each other, Node 1 cannot detect Node 3 transmission and vice-versa. Consequently, a collision at Node 2 occurs. This situation can be avoided by the use of a RTS-CTS handshake prior to the DATA transmission.

channel busy, it temporarily freezes its timer. Nodes may start emitting only when their timer reaches zero. After each transmission, the emitter resets its timer to a new random value. Figure 2.5 illustrates the backoff mechanism for two nodes competing to access the channel.

2.2.2 Idealized 802.11

In most of this work, we consider an idealized version of IEEE 802.11 in order to capture its essential features, and to leave aside the effects due to the imperfection of the real protocol (we refer the reader to Chapter 8 for an overview of those effects). We refer to this idealized version of IEEE 802.11 as *idealized 802.11 protocol* or for short *802.11* (without the IEEE in front).

Exclusion Domain

Before describing our idealized protocol, we need to introduce the notion of exclusion domain.

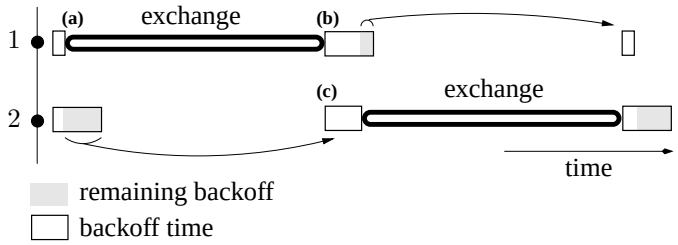


Figure 2.5: Backoff mechanism for two competing nodes. (a) The backoff timer of Node 1 reaches zero, Node 1 transmits. (b) Node 1 resets its backoff timer with a new random value. (c) Node 2 that had frozen its backoff timer during Node 1 transmission, resumes the decrease of its backoff timer.

A link can be in two states, *active* or *idle*. A link is active if there is a data transmission between its two end-nodes. We denote the state of link j by x_j,

$$x_j = \begin{cases} 1 & \text{if link } j \text{ is active} \\ 0 & \text{if link } j \text{ is idle.} \end{cases}$$

During a data transmission, nodes in the RXRanges of the sender and the receiver nodes are silenced by their virtual carrier sensing mechanism (they can neither transmit nor receive). In addition, nodes within the CSRange of the sender are kept from sending by their physical carrier sensing mechanism. We refer to the set of links silenced by an active link by its virtual and physical carrier sensing as its *exclusion domain*. We denote the exclusion domain of link j by $\mathcal{E}_j$.

In this work we concentrate on *symmetric* exclusion domains (Figure 2.6). We call the domain symmetric because the reverse link has the same exclusion domain. In practice, this situation arises when the CSRange and the RXRange have similar values and the network topology is sufficiently regular. For example, indoor environments with their carefully designed network topologies and strong signal attenuation may lead to such a situation. Chapter 7 shows how our approach can be extended to consider *asymmetric* exclusion domains (Figure 2.7), which arise when the CSRange is significantly larger than the receiving range.

Assumptions

The idealized 802.11 protocol assumes that a link can accurately and instantaneously detect activity on the communication channel (using the physical and virtual carrier sensing mechanisms of its two end-nodes) and becomes active if and only if none of the links in its exclusion domain is already active.

In addition, in the idealized protocol, links maintain a separate backoff timer,

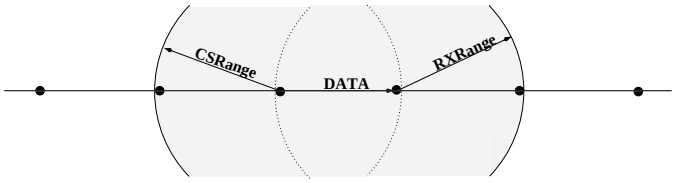


Figure 2.6: CSRange $=$ RXRange $=1$ space unit. All links with an end-node in the gray domain around the active link are silenced (nine directed links are thus silenced). As the nodes in the transmitter's CSRange are in its RXRange, the exclusion domain is symmetric.

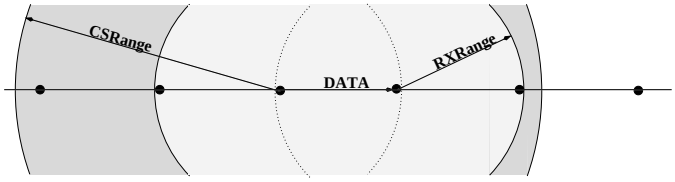


Figure 2.7: RXRange $=1$ and CSRange $=2.2$. As the CSRange is larger than the RXRange, more nodes are silenced around the sender. The exclusion domain is asymmetric.

even if they share the same source node[2]. Moreover, we consider a continuous backoff distribution, instead of the discrete distribution implemented in actual IEEE 802.11 cards, so that two timers have a zero probability of expiring at the same time.

These assumptions imply that the idealized protocol is collision free.

2.3 Conclusion

In this chapter, we have presented the physical layer that we will use throughout this work. Using this physical layer, we defined a network graph. Our main interest is, however, in the medium access control layer. In particular, we introduced the IEEE 802.11 MAC protocol together with an idealized version of this protocol. The study of this protocol is justified by its wide usage in todays wireless networks, and by the lack of existing models especially in the context of multi-hop networks.

[2]Using a backoff timer per directed link was recommended by [BDSZ94] in an attempt to guarantee per link fairness in the context of single-hop networks. Chapter 8 gives further reasons for this assumption in multi-hop networks.

Chapter 3

Markovian Framework

We present a Markovian framework that can be used to model the dynamics of the idealized 802.11 protocol in multi-hop ad hoc networks. This framework relies on two assumptions: (i) the links use an exponential backoff distribution, and (ii) there is always a packet to send on each link (i.e., we operate in saturated traffic conditions).

3.1 Finite Networks

We first describe the dynamics of 802.11 on networks whose graphs have a finite number L of edges. At any given time, the protocol decides which of the $2L$ directed links is active (respectively, idle).

3.1.1 Continuous Markov Chain

If the backoff and the exchange time distributions are exponential, we can model the dynamics of 802.11 by a continuous time Markov chain. Figure 3.1 gives an example of such a Markov chain. The states of the Markov chain are *transmission patterns*.

Definition 3.1 *[Transmission Pattern]*
A transmission pattern x is a vector in $\{0,1\}^{2L}$ that specifies which of the $2L$ directed links is active (resp., idle). A transmission pattern must satisfy the constraints set by the exclusion domains of active links: for every link j such that $x_j = 1$, $\sum_{k \in \mathcal{E}_j} x_k = 0$ (remember that $\mathcal{E}_j$ is the exclusion domain of link j).

We denote by $n(x)$ the number of active links (i.e., the number of ones) in a transmission pattern x.

Let $\mathcal{V}$ be the set of transmission patterns allowed by 802.11 under symmetric exclusion domains. The Markov chain has one state per element of $\mathcal{V}$. Moreover, we say that a state is at *level* i of the Markov chain if the corresponding transmission pattern has exactly i active links, and we denote by $N(i)$ the number of such

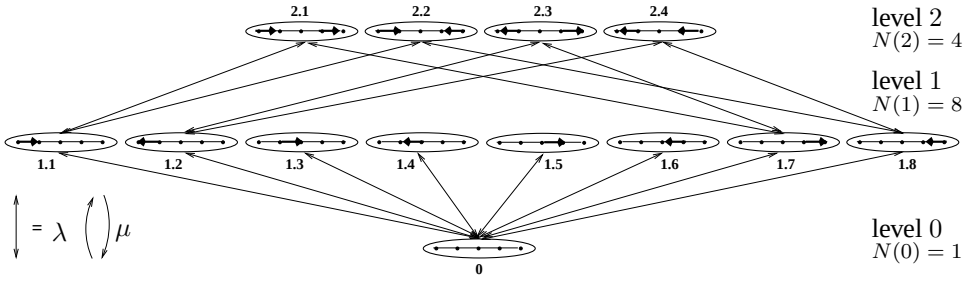


Figure 3.1: Markov chain for a small line topology of 5 nodes. The RXRange and the CSRange cover exactly one neighbor. There are 13 states, $N(0) = 1$ state with no active link, $N(1) = 8$ states with 1 active link, and $N(2) = 4$ states with 2 active links.

states. The states at the highest level are called *patterns of maximal spatial reuse*: they will be of particular interest for our analysis.

Transitions in the Markov chain can only occur between states that are separated by one level: A transition between a state at level i and a state at level $i-1$ corresponds to the completion of a transmission, and a transition between a state at level i and a state at level $i+1$ corresponds to the beginning of a new transmission (which is of course possible only if the new link does not lie in the exclusion domain of already active links).

Denote by μ^{-1} the average exchange time and by λ^{-1} the average backoff time. The transition rate between a state at level i and a state at level $i-1$ (respectively, at level $i+1$) is μ (resp., λ). Moreover, we define the *access intensity* as $\rho := \lambda/\mu$.

In the case of symmetric exclusion domains, the direction (left to right or right to left) of an active link does not change its exclusion domain. Consequently, we can *make abstraction of the link direction* and obtain a Markov chain with a reduced number of states. This *simplified Markov chain* is shown in Figure 3.2 and (for all practical purposes) is equivalent to the Markov chain of Figure 3.1.

3.1.2 Stationary Distribution

A 802.11 network with symmetric exclusion domains is a loss network [Kel91] (see next section for more details). Consequently, the Markov chain is irreducible and reversible. Moreover, its unique stationary distribution has a product form. The stationary probability of a transmission pattern $x \in \mathcal{V}$ is given by

$$\pi(x) = \frac{\rho^{n(x)}}{Z} \tag{3.1}$$

where $Z = \sum_{y \in \mathcal{V}} \rho^{n(y)}$ is a normalizing constant. All the transmission patterns x with the same number $n(x)$ of active links thus have the same stationary probability.

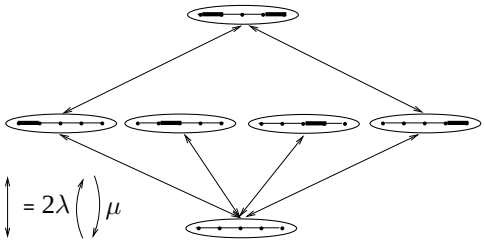


Figure 3.2: Simplified Markov chain that does not account for the direction of active links. In the Markov chain of Figure 3.1 there is one state per transmission pattern. In the simplified Markov chain represented in this figure, a state corresponds to several transmission patterns. All the transmission patterns that have the same pairs of communicating nodes (i.e., that differ only by the direction of their active links) are represented by the same state. For example, the four top level states of Figure 3.1 are grouped into a single top level state. More generally, a state with i active undirected links corresponds to 2^i states with i active directed links. In the simplified Markov chain a transition upward happens when an undirected link becomes active or, in other words, when any of the 2 corresponding directed links becomes active. Such a transition thus arises at rate 2λ instead of λ.

We denote by

$$\pi_i = \frac{1}{Z} \sum_{x \in \mathcal{V} \text{ such that } n(x)=i} \rho^{n(x)} = \frac{N(i)\rho^i}{\sum_k N(k)\rho^k} \tag{3.2}$$

the probability to be at level i of the Markov chain, i.e., the probability to be in a transmission pattern with i active links.

The dependence of $\pi(x)$ and π_i on ρ is omitted on purpose in order to keep the notations as simple as possible.

It is known [BLL84] that (3.1) and (3.2) are insensitive to the exchange time distribution.

For $\rho > 1$, $\pi(x)$ increases with the value of $n(x)$, and the transmission patterns with a high number of active links have an increased probability of appearing, compared to those with only a few active links. *In the limit $\rho \to \infty$, only the transmission patterns with the highest number of active links have a non-zero stationary probability.*

3.1.3 Analogy with Other Models

The idealized 802.11 protocol is analogous to three well studied stochastic models: The *filling and depletion* process [BCJ04] encountered in the packing literature, the loss networks [Lou90, Kel91] already mentioned above, and the hard-core model in statistical physics [vdBS94, GHM98].

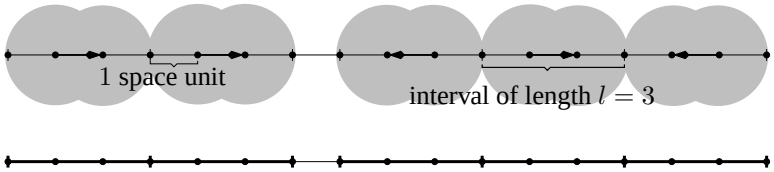


Figure 3.3: Analogy between a medium access control problem and a packing problem. A transmission pattern corresponds to a set of non-overlapping intervals on the line. This specific transmission pattern has five active links or equivalently five packed intervals.

Consider a regular network topology such as the line topology. We define the parameter l as 1 plus the minimal distance separating two active links. This means that there can be an active transmission every l space units. In our setting the RXRange and the CSRange are equal to 1 space unit, which means that transmissions can take place simultaneously every 3 space units, so that $l = 3$ in this case. To each active link, we can associate an interval of length l on the line, which corresponds to the portion of the line 'occupied' by an active link (Figure 3.3). According to the protocol, intervals of length l arrive (are packed) at rate 2λ and depart (are unpacked or removed) at rate μ. An interval is accepted (i.e., one of the corresponding links becomes active), if it does not overlap with intervals already accepted. Consequently, a transmission pattern corresponds to a set of non-overlapping intervals on the line. Despite its simplicity, this *packing formalism*, is quite powerful and we will repeatedly use it throughout this work.

Similarly, the analogy between a network running the 802.11 protocol and a loss network is quite straightforward. The transmissions (active links) correspond to calls and arrive at exponentially distributed times due to the backoff mechanism. Moreover, each transmission 'occupies' some resources in the network due to the exclusion domain constraint that forbids other transmissions in its immediate neighborhood. This corresponds to the reservation of resources along a route in a loss network.

Finally, our model for 802.11 is a hard-core model with activity parameter ρ. The hard-core model is typically used in statistical physics to model a gas whose particles have a non-negligible size. In this model, $x_j = 1$ means that position j is occupied by a particle and the exclusion domain constraints prevent particles from overlapping.

3.2 Infinite Networks

If the number of links in the network is not finite, the normalization constant Z in (3.1) is infinite. To define a measure on infinite networks we use the concept of a *contention graph*.

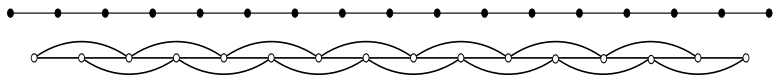


Figure 3.4: Contention graph (bottom) corresponding to a line network (top). The RXRange and the CSRange are both equal to 1 space unit (i.e., $l = 3$). In the contention graph there is one vertex per network link and two vertices are connected if the corresponding links are in one another's exclusion domain.

3.2.1 Contention Graph

In a contention graph, the vertices represent the (undirected) links of the network and the edges mutual exclusion of links. In other words, there is an edge between two vertices if their corresponding links are in one another's exclusion domain. Note that this graph is different from the graph representing the network itself. Figure 3.4 shows the contention graph for a small line network. We say that a vertex is in *state 1* if the corresponding link is active, and in *state 0* otherwise. In this new graph, the exclusion constraints imply that two adjacent vertices cannot be in state 1 simultaneously.

3.2.2 Markov Random Field and Gibbs Measure

A central property of the stationary measure π is that it is a Markov random field[1] on the contention graph. We describe briefly the properties of Markov random fields, which we need in the following chapters. For more details, we refer the reader to [KS80] and to [Dur81] (pages 126-130).

Definition 3.2 *[Markov Random Field]*
Consider a contention graph where each of the L vertices takes its value in $\{0, 1\}$. If B is a finite subset of its vertices, we denote by B^c the vertices outside B and by ∂B all the vertices in B^c that are adjacent to a vertex in B.

A random variable taking values in $\{0, 1\}^L$ with distribution π is called a Markov random field on this contention graph if for each finite subset B of vertices, the conditional distribution of the states inside B given the states of all the vertices in B^c only depends on the state of the vertices in ∂B. In mathematical terms we have

$$\forall \textit{ finite } B \textit{ and } x_B, \quad \pi(x_B|x_{B^c}) = \pi(x_B|x_{\partial B}).$$

More specifically, if $x_B \in \{0, 1\}^{|B|}$ is a transmission pattern in B, one can easily verify from (3.1) that the measure π conditioned on the state of vertices in ∂B is

$$\forall \text{ finite } B \text{ and } x_B, \quad \pi(x_B|x_{\partial B}) = \frac{\rho^{n(x_B)} I(x_B, x_{\partial B})}{Z'}, \tag{3.3}$$

[1] The term "Markov" here refers to the spatial structure of the measure. It has nothing to do with the fact that π is originally the stationary measure of a Markov chain (over time).

where Z' is an appropriate normalizing constant, and where $I(x_B, x_{\partial B}) = 1$ if no adjacent vertices in x_B and $x_{\partial B}$ are both in state 1 and $I(x_B, x_{\partial B}) = 0$ otherwise. As an example, consider the case where the set B contains a single vertex v. We have that $\pi(x_v = 1|x_{\partial v}) = \rho/(\rho+1)$ if all vertices adjacent to v are in state 0, and 0 otherwise. Similarly, $\pi(x_v = 0|x_{\partial v}) = 1/(\rho+1)$ if all vertices adjacent to v are in state 0, and 1 otherwise.

Let us now consider the measures π over the space of all possible transmission patterns that fulfill the specification (3.3). These measures are called *Gibbs measures*.

If the number of links L in the network is infinite, a specification may have more than one Gibbs measure. A standard problem in statistical physics is to determine under which condition such a measure is unique.

Remark

In statistical physics, a Gibbs measure is traditionally defined via a Hamiltonian (also called an energy function) that derives from a Gibbs potential V. The Hamiltonian has the following form

$$H(x) = \sum_{v \sim w} V(x_v, x_w) + \sum_{v} V(x_v)$$

where $v \sim w$ denotes two adjacent vertices in the contention graph. In our model (i.e., the hard-core model), $V(x_v) = -x_v \log \rho$ and

$$V(x_v, x_w) = \begin{cases} \infty & \text{if } x_v = x_w = 1 \text{ and } v \sim w \\ 0 & \text{otherwise.} \end{cases}$$

The Gibbs distribution is then

$$\pi(x) = \frac{1}{Z} e^{-H(x)},$$

where Z is the normalizing constant. In fact there is an equivalence between Markov random fields and Gibbs measures with an energy function deriving from a Gibbs potential (see for example [Bré98]).

3.3 Related Work

For a long time, models for the 802.11 protocol have been limited to the single-hop setting and numerous works have extended the original papers of Bianchi [Bia00] and Calì, Conti and Gregori [CCG00b, CCG00a]. These papers use a discrete time Markov chain to model the operation of the backoff timer at a given node. Each state of the Markov chain is identified by the value of the current contention window cw and by the actual value of the backoff timer. The time spent in a given state of the Markov chain is variable and is equal to the time between two consecutive backoff decrements. These models assume that each transmitted packet

collides with a constant and independent probability p. They express the stationary probability that a node transmits in a generic back-off slot as a function of p and vice-versa. Finally they use numerical techniques to solve the resulting fixed point equation. The recent paper of [KAMG07] simplifies and generalizes the analysis leading to this fixed point equation. It also gives a condition for the uniqueness of the fixed point.

Unfortunately, the models for the 802.11 protocol in the single-hop setting are not easily carried over to the multi-hop setting because they rely explicitly on the assumption that nodes can sense each other's transmissions and thus have the same perception of the activity on the communication channel. In the last few years however, several papers have attempted to model the behavior of the 802.11 protocol in the multi-hop setting.

Most of these models [GK03, WGLA04, CGLA04, MT06, GSK06] are inspired by the work done in the single-hop setting, in the sense that they still use individual (discrete time) Markov chains to characterize the state of each node. In addition, they use topology dependent information to characterize the channel activity. As each station has its own private view of the channel, this approach leads to huge systems of equations that can only be solved via numerical methods and that give little intuition on the behavior of the protocol. To reduce the complexity of these systems [GK03, WGLA04, MT06] consider specific network topologies such as ring, grid or Poisson topologies, whereas [CGLA04, GSK06] present more general techniques to simplify these systems of equations.

The work of [WK05], is the closest to our work. Instead of considering nodes in isolation, it is based on the notion of transmission patterns. It extends the Markov chain formalism used by [Tob87, BKMS87] to model the CSMA protocol, to an idealized 802.11 protocol. The strength of this model is to preserve the dependence between nodes, typical of CSMA protocols, as well as of the 802.11 protocol. As mentioned previously, this model has also strong connections with Kelly's work [Kel91] on loss networks as the stationary distribution of its Markov chain has the same form. Recently, [BMP07] revisited this model using the formalism of interacting particle system.

3.4 Conclusion

In this chapter, we have presented a Markovian framework that captures the dynamics of the idealized 802.11 protocol in multi-hop ad hoc networks. In the four following chapters we will use this framework to derive its properties and performances. We will also show that this framework is quite robust to the different assumptions made in terms of exchange and backoff distributions and that it can be easily extended to asymmetric exclusion domains.

Chapter 4

Spatial Reuse

In this chapter we evaluate the performance of the idealized 802.11 in terms of spatial reuse.

Definition 4.1 *[Spatial Reuse]*
The spatial reuse is the number of active links in the network normalized by L.

In this work, we mainly use the average spatial reuse as a performance metric.

Definition 4.2 *[Average Spatial Reuse σ]*
The average spatial reuse σ is the spatial reuse averaged over time.

Using Equation (3.2) we obtain the following expression for the average spatial reuse of the idealized 802.11 protocol at steady state,

$$\sigma = \frac{1}{L}\sum_i i\pi_i = \frac{1}{L}\frac{\sum_i iN(i)\rho^i}{\sum_i N(i)\rho^i}. \tag{4.1}$$

The spatial reuse measures the ability of the protocol to schedule simultaneous transmissions. Achieving a high number of simultaneous transmissions is a first step towards providing high network throughput.

4.1 The Line Topology

Equation (4.1) is valid for any finite network topology, but it is in general difficult to obtain a closed-form expression for $N(i)$, the number of transmission patterns with i active links. However, we can do so for the line topology.

4.1.1 Analytical Results

We use the packing formalism introduced in Section 3.1.3. To each active link we associate an interval of length $l = 3$. Therefore, finding possible transmission

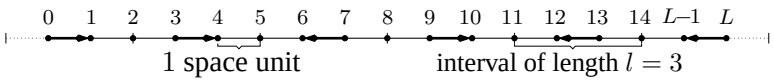


Figure 4.1: A transmission pattern of maximal spatial reuse with symmetric exclusion domains (RXRange = CSRange = 1 space unit).

patterns boils down to packing intervals of length l in a line segment of length $L+(l-1)$ (the additional term $(l-1)$ compensates for the topology border, and is illustrated by the dashed segments in Figure 4.1). The number of valid transmission patterns with i active links is thus

$$N(i) = 2^i \binom{i+v}{i},$$

where $v = L + (l-1) - li$ is the length of the vacant space on the line once i intervals of length l are placed, and where the 2^i factor accounts for all possible permutations of the link directions.

Plugging the expression for $N(i)$ in Equation (4.1), we can compute analytically the spatial reuse on a finite line topology. Note that these values depend on the total length of the network L.

Theorem 4.1 *For large networks, i.e., when $L \to \infty$, the average spatial reuse σ of the protocol under symmetric exclusion domains converges to*

$$\lim_{L\to\infty} \sigma = \frac{2\rho y_1^{l-1}}{1 + 2l\rho y_1^{l-1}}$$

where y_1 is the positive real root of $1 - y - 2\rho y^l$.

Proof: The result is obtained by applying Lemma A.1 (in the appendix) with $k = \lfloor \frac{L+(l-1)}{l} \rfloor$, $m = l$, $r = 2\rho$, and $n = (L + (l-1)) \bmod l$. We then divide by l to obtain the average spatial reuse σ. □

Alternatively, this theorem can also be shown by adapting the approach of [BCJ04] which uses multivariate generating functions combined with the residue method. Although this approach yields the same result, we believe our approach is simpler. Moreover, Lemma A.1 can also be used to obtain the average spatial reuse under asymmetric exclusion domains, as we will see in Chapter 7.

4.1.2 Interpretation of the Results

Figure 4.2 shows the spatial reuse of 802.11 as a function of ρ.[1] As expected, when ρ increases, σ tends to $1/3$ (a little bit more for the finite topology), which is

[1] All the simulation results for the idealized 802.11 protocol were obtained using [ths]. The simulation time was chosen long enough so as to render the effect of the transient negligible (this is confirmed by the perfect match between simulation and analytical results).

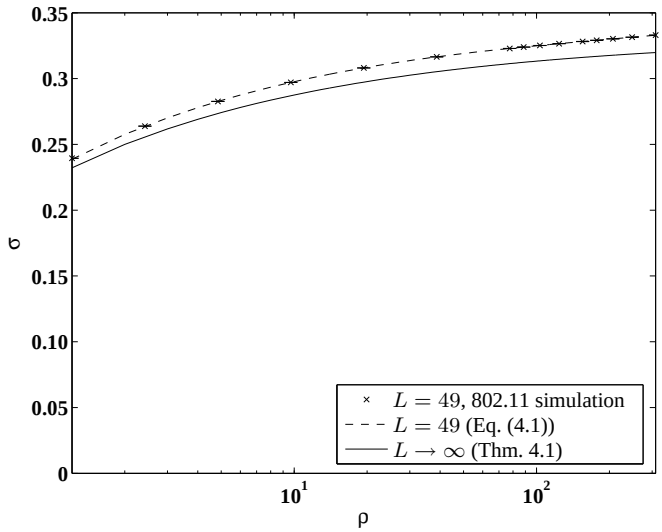


Figure 4.2: Average spatial reuse σ of the idealized 802.11 protocol as a function of the access intensity ρ on line networks of different sizes.

the maximal spatial reuse. Consequently, when the average backoff time is much lower than the average exchange time, 802.11 achieves the maximal spatial reuse. This encouraging result shows that a decentralized protocol such as 802.11 can optimally organize the transmissions in space.

To better illustrate the operation of our protocol, consider the case where the average exchange time is much larger than the average backoff time (i.e., $\lambda \gg \mu$). The protocol first fills the space with active links until no other additional active link can be added. When a transmission ends, it is almost immediately replaced by one or *more* new transmission(s). Once the protocol reaches one of the patterns of maximal spatial reuse, the probability of going back to a transmission pattern with a low spatial reuse is small, as it would imply the termination of several transmissions between two consecutive transmission attempts. This event is unlikely when the average backoff time is much lower than the average exchange time. Figure 4.3 shows the activity of the different links on a line topology of 50 nodes (i.e., $L = 49$) when $\rho = 200$ and $l = 3$. On this topology, as $(L + (l - 1)) \bmod l = 0$, all the transmission patterns of maximal spatial reuse have the same pairs of communicating nodes. In other words, if we make abstraction of the link directions, there is only one pattern of maximal spatial reuse (i.e., there is a single state at the top level of the simplified Markov chain) and 802.11 spends most of its time in this pattern.

The random backoff mechanism creates a local coupling between the links, which is strong enough to spatially organize the transmissions in optimal global patterns. A similar densification phenomenon occurs in the context of the packing

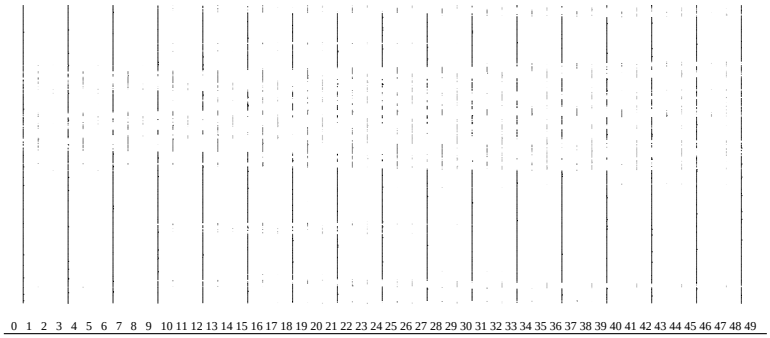


Figure 4.3: Activity of the links in a 50 node line network and over a four second period. On the horizontal axis the coordinates of the network nodes. On the vertical axis the time. We draw a line from $(j-1/2, t_1)$ to $(j-1/2, t_2)$ if either Node $j-1$ or Node j is transmitting between time t_1 and t_2. We observe that the protocol spends most of its time in patterns of maximal spatial reuse.

of granular materials [TTV00]. In essence, the additional randomness introduced by the backoff mechanism produces the same effect as shaking a bucket of sand to achieve a better packing of the sand inside.

Although our model assumes exponential backoff distributions, the ability of 802.11 to organize the transmissions in space does not rely on this assumption. In particular, we obtain the exact same spatial reuse performance (the markers in Figure 4.2) when we simulate the idealized 802.11 protocol with a uniform backoff distribution and a constant exchange time.

4.1.3 Comparison with a Slotted Protocol

It is interesting to compare the spatial reuse of 802.11 that is asynchronous (i.e., that does not require synchronization between the network nodes) with the spatial reuse achieved by its slotted or synchronous counterpart. This is in particular useful to characterize the level of organization achieved by the 802.11 protocol.

The slotted protocol is decentralized and uses the same local information as 802.11. Time is divided into slots of length $1/\mu + \epsilon$ with $\epsilon \ll 1/\mu$. The first ϵ time units of the slot are used to select the transmission pattern that will be active during the remaining of the slot. Consequently, in the slotted protocol, all transmissions taking place in a slot start at the same time and stop simultaneously as well. At the beginning of a time slot, each link picks a random number between 0 and ϵ (for example, according to a uniform distribution on $[0, \epsilon)$). After waiting for this randomly selected time, if the link does not detect any other already active link in its exclusion domain, it sends a busy signal until the end of the ϵ time units and is selected to be part of the transmission pattern for the given slot. If the link does

detect activity on the channel, it remains idle for the remaining of the slot. At the end of a time slot, the process is repeated, and another transmission pattern is selected.

Remarks

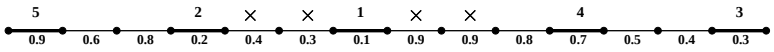
- According to the slotted protocol (described in the previous paragraph), the transmission pattern selected is always maximal, i.e., no additional active link can be added to the pattern.

- Alternatively, instead of waiting for a time t uniformly distributed between 0 and ϵ and then transmitting a busy signal between time t and ϵ, a link could send a busy signal between time 0 and t. The link would then be part of the transmission pattern for the given slot if at time t (i.e., at the end of its busy signal transmission) it senses the channel as idle. This strategy is used by the Hiperlan protocol [ALM98] in the so-called elimination phase (although with a truncated geometric distribution instead of a uniform distribution). It is also reminiscent of the Matérn hard core process (although this process is only defined for Poisson topologies). However, this strategy always yields a lower spatial reuse than the one obtained with the slotted protocol, and it does not guarantee that the transmission patterns selected are maximal (Figure 4.4 illustrates this statement).

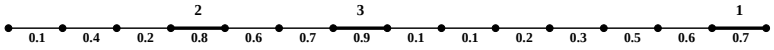
In the slotted protocol, the process for selecting the transmission pattern is identically distributed at each time slot, and independent from time slot to time slot. The average spatial reuse at steady state is thus equal to the average spatial reuse observed in any of the time slots. Consider the process for selecting the transmission pattern for a given time slot. To each active link we associate an interval of length l on the line. The construction of a transmission pattern for a given time slot is then equivalent to the *filling* of the line with as many non-overlapping intervals as possible. However, because of the randomness of the filling process, a part of the line might remain uncovered by the intervals. In 1962, Mackenzie [Mac62] used a recursive approach to compute the average fraction of the line left vacant when the length of the line network goes to infinity. Indeed, each accepted interval, reduces the initial problem to two independent, smaller instances of the same problem.

Exploiting Mackenzie's result, the average spatial reuse achieved by the slotted protocol when $L \to \infty$ can be recast as

$$\sigma_f = \exp\left[-2F(1)\right] \int_0^1 \exp\left[2F(u)\right] du, \tag{4.2}$$



(a) The link with value 0.1 transmits first and by doing so silences the 2 links on its left and right sides (the cross on the figure). Among the links that are not silenced, the link with value 0.2 has the smallest value and transmits next, etc. In this protocol, once a link is silenced (by an active link), it cannot prevent an idle link from becoming active.



(b) In this protocol, a link can only transmit if its value is larger than the values of the 2 links on its left and right sides (4 links in total). Among the links that satisfy this condition, the link with the smallest value is the first to be selected. The transmission pattern selected is a subset of the transmission pattern selected in Figure (a). Moreover, it is not maximal.

Figure 4.4: Comparison of a slotted protocol where (a) a link first waits for a random time and then (if it senses the channel as idle) sends a busy signal, and (b) where a link immediately sends a busy signal of a random duration and then senses the activity on the channel to decide whether it should transmit or not in the slot. Here the random values are selected between 0 and $\epsilon = 1$. In (a) the link with the lowest value has the highest chance to access the channel while in (b) it is the opposite. This explains why the values (represented below the links) in Figure (b) are equal to 1 minus the values in Figure (a).

where

$$F(u) = \int_0^u \frac{1 - y^{l-1}}{1 - y} dy = \sum_{i=1}^{l-1} \frac{u^i}{i}.$$

For $l = 3$, $\sigma_f = \int_0^1 \exp(u^2 + 2u - 3)du \simeq 0.275$ so that we are approximately 18% away from the maximal spatial reuse of $1/3$. For larger l, the spatial reuse decreases: the larger the exclusion domain, the more inefficient the slotted protocol. Moreover, a comparison with 802.11 ($l = 3$) shows that for values of $\rho > 5$ the 802.11 protocol achieves a higher spatial reuse than the slotted protocol. This shows that 802.11 starts to organize the transmissions in space already for low values of the access intensity.

4.2 Two-Dimensional Topologies

The organizing properties of 802.11 (i.e., its ability to favor the patterns of maximal spatial reuse) do not depend on the underlying network topology. Indeed the organizing properties of the protocol can be inferred directly from Equations (3.1)

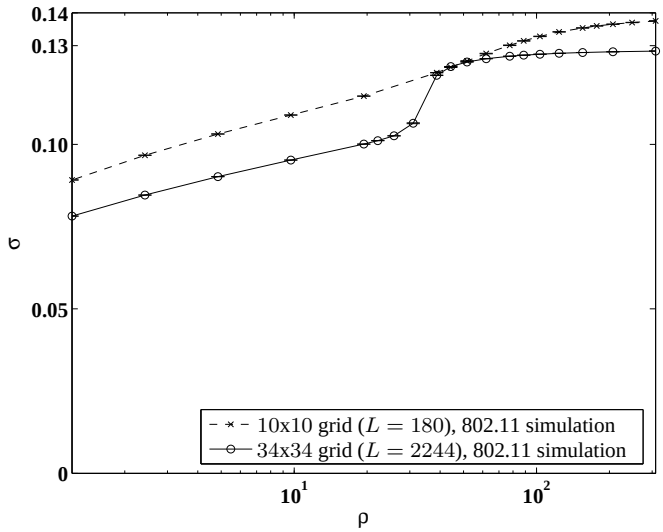


Figure 4.5: Average spatial reuse σ of the idealized 802.11 protocol as a function of the access intensity ρ on grid networks of different sizes. On the 10×10 (resp., 34×34) network the spatial reuse converges progressively towards the maximum of 0.14 (0.13 respectively).

and (3.2). Using these equations, it is easy to see that for $\rho \to \infty$ only the transmission patterns of maximal spatial reuse have a non-trivial stationary probability. Consequently, the 802.11 protocol can achieve the maximal spatial reuse in finite two-dimensional topologies as well. However, contrary to the case of the line topology, we cannot obtain an expression for $N(i)$ in two-dimensional topologies, and as a result we have to rely on simulations to obtain the curve for σ as a function of the access intensity ρ. Figures 4.5 and 4.6 show the simulation results obtained for two grid topologies and for the irregular topology, respectively. In all cases, the spatial reuse of the protocol increases with the access intensity ρ. Observe that in the large grid network the spatial reuse of the protocol increases very rapidly for values of ρ between 30 and 40. This corresponds to a phase transition phenomenon that we will study in detail in the next chapter.

4.3 Related Work

To the best of our knowledge, previous models of 802.11 do not give closed-form formulæ for its spatial reuse, nor do they notice its organizing properties (although [WK05] mentions that the idealized 802.11 protocol achieves maximal spatial reuse when the access intensity goes to infinity). In this work, we borrow tools from statistical physics to predict the performance of 802.11 on large networks.

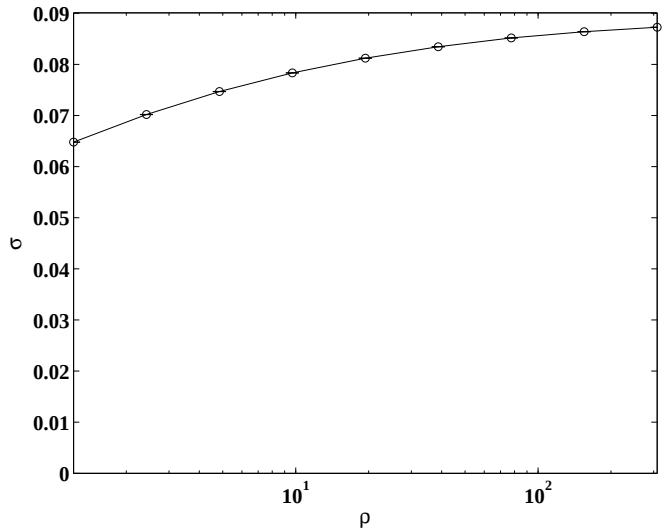


Figure 4.6: Average spatial reuse of the idealized 802.11 protocol as a function of the access intensity ρ on the irregular network ($L = 2435$, see Section 2.1.2 for a detailed description of this topology).

Techniques from statistical physics have been applied first in [Yem83] to analyze large-scale distributed resource sharing mechanisms. In particular, our approach to compute the spatial reuse of the protocol on large line networks is a natural continuation of [PY86] (although we became aware of [PY86] only later). However, our proof techniques differ as they are either based on more recent results [BCJ04] or simpler mathematical tools.

4.4 Conclusion

In this chapter we have presented analytical formulæ for the spatial reuse of 802.11. In particular, we have derived a closed-form formula for its spatial reuse in an infinite line network. Maybe even more important than these quantitative results, is the ability of the idealized 802.11 protocol to organize the transmissions in space. The level of organization, i.e., the spatial reuse of the protocol, increases as the ratio of the average exchange and backoff time becomes large. Asymptotically, when this ratio reaches infinity, only the transmission patterns of maximal spatial reuse have a non-trivial probability. Our results show that the idealized 802.11 can basically find these global patterns in a completely decentralized way. This is quite surprising as the problem of finding these patterns in a general network is NP-Complete (by equivalence with the maximal independent set problem [GJ83]). The idealized 802.11 protocol can thus be viewed as a decentralized optimization algorithm maximizing the spatial reuse.

Chapter 5

Long-Term Fairness

In this section we look at the long-term fairness properties of the idealized 802.11 protocol. The long-term fairness of the protocol is the level of fairness achieved by the protocol after having run for a very long (infinite, in theory) time.

To assess the fairness of the protocol we use *Jain's Fairness Index* [Jai91].

Definition 5.1 *[Jain's Fairness Index]*
Denote by $p(j)$ the probability that a directed link j is active. The link fairness index FI of the protocol is

$$FI = \frac{\left(\sum_j p(j)\right)^2}{2L \sum_j p(j)^2} \tag{5.1}$$

where $2L$ is the number of directed links in the network.

The maximum fairness index is FI $= 1$. It corresponds to a network where all links access the channel equally. Yet, if only k directed links have an equal access to the channel and the remaining links have no access to the channel, the fairness index is $k/(2L)$.

To derive the long-term fairness of the protocol, we thus need to know the probability that a specific link is active in steady state. A link j is active if the chain is in a state whose transmission pattern contains link j. Denote by $N(i, j)$ the number of such patterns with a total number of active links equal to i. The probability $p(j)$ that link j is active is

$$p(j) = \frac{\sum_i N(i, j)\rho^i}{\sum_i N(i)\rho^i}. \tag{5.2}$$

5.1 Line Topology

To obtain an expression for $p(j)$ and FI, we therefore need to compute the number $N(i, j)$ of transmission patterns with i active links that include a given link j.

Currently, we can only do this for the line topology. We denote by j_{right} the link between node $j-1$ and node j, and by j_{left} the link between node j and node $j-1$. To compute $N(i, j_{\text{right}})$, we have to count the number of ways to place $i-1$ active links around the already active link j_{right}. Assume that we place k of these active links on the left of link j_{right}, and $(i-1)-k$ on its right. We obtain

$$N(i, j_{\text{right}}) = 2^{i-1} \sum_{k=0}^{i-1} \binom{k + v_{\text{left}}}{k} \binom{(i-1) - k + v_{\text{right}}}{(i-1) - k},$$

where $v_{\text{left}} = j - 1 - lk$ and $v_{\text{right}} = L - j - l((i-1) - k) = L + l - j - l(i-k)$. Clearly, as the exclusion domains of link j_{right} and link j_{left} are identical by assumption, the computation of $N(i, j_{\text{left}})$ leads to the same result. By Equation (5.2), this implies that two links in opposite directions have the same probability to be active. Plugging the expressions for $N(i,j) = N(i, j_{\text{right}}) = N(i, j_{\text{left}})$ in Equation (5.2) and the resulting expression for $p(j)$ in Equation (5.1), we can compute analytically the fairness index of the protocol on a finite line topology.

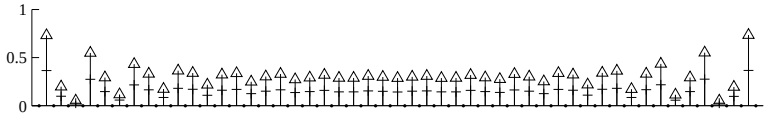
Figure 5.1 illustrates the stationary marginal probabilities of the links in a 50 node line topology for three values of ρ. We observe that as ρ increases, the protocol becomes more unfair. We also observe that the links that belong to the transmission patterns of maximal spatial reuse have an increased access to the channel. Indeed, the ability of 802.11 to organize the transmissions in patterns of maximal spatial reuse can severely impact its long-term fairness on finite topologies. In the limit, when $\rho \to \infty$, only the links that belong to the transmission patterns of maximal spatial reuse get access to the channel; the remaining links get no access to the channel. In the literature this is often referred to as the *starvation* problem of 802.11.

The unfairness of 802.11 is due to the border of the topology. Links at the border have a larger probability of being active than other links, as they compete with a reduced number of links to access the channel. The value of ρ determines how far this border effect propagates into the network. Note that if we join the two boundary nodes of the line topology to form a circle topology (thereby effectively removing the topology border), the protocol becomes completely fair for all values of ρ. Indeed, in a circle topology, $N(i,j)$ does not depend on j, which in turn implies that $p(j)$ is the same for all links.

Figure 5.2 illustrates the stationary marginal probabilities of the links for $\rho = 155$ on line topologies of 5, 50, and 500 nodes, respectively. We see that on large topologies, links in the middle of the topology have an equal access to the channel and that for a given finite value of ρ the protocol is thus fairer on large topologies than on small topologies.

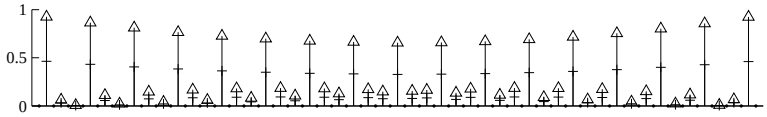
Theorem 5.1 *Consider an L-link[1] line network. When $L \to \infty$ and for all finite values of ρ, the probabilities of links in the center of the network to be active are asymptotically all equal and independent of the border condition.*

[1]Here a link corresponds to two directed links. We already know that a directed link and the corresponding reverse link have the same probability of accessing the channel.



(a) $\rho = 20$, $\sigma = 0.31$, FI $= 0.85$

(b) $\rho = 155$, $\sigma = 0.33$, FI $= 0.71$



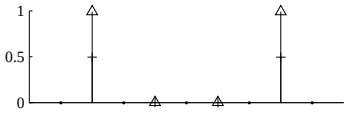
(c) $\rho = 620$, $\sigma = 0.34$, FI $= 0.54$

Figure 5.1: Border effect on a 50 node line network, for different values of the access intensity ρ. The line network is represented on the horizontal axis. Each of its edges corresponds to two directed links j_{right} and j_{left}. There is one vertical bar per network edge and each bar is divided in two equal parts, the bottom part measures $p(j_{\text{right}})$ and the top part $p(j_{\text{left}})$. We observe that as ρ increases 802.11 becomes more unfair and the border effect is felt deeper and deeper inside the network.

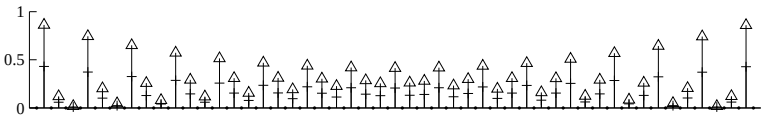
Proof: We first show that in the case of an infinite line network, there is a unique measure satisfying the specification (3.3). As the equations in (3.3) are translation and reflection invariant, this in turn implies that the unique measure is uniform – no link is favored.

We start with the observation that the contention graph in this case is an infinite string of vertices, which are connected by an edge to their neighbors and second nearest neighbors (see Figure 3.4). Let us define a site percolation process on this graph by declaring each vertex "open" with probability $\rho/(1+\rho)$ and "closed" with probability $1/(1+\rho)$, independently of each other.[2] An *open path* is a sequence of connected open vertices that are distinct. Hence, in our i.i.d. site percolation model, there is an open path between open vertices j_1 and j_2 if among the vertices located

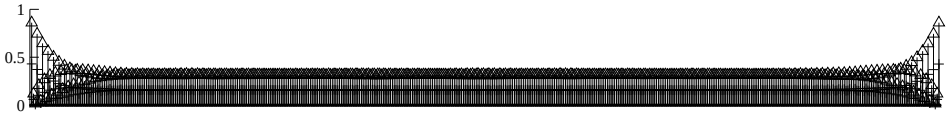
[2]The value $\rho/(1+\rho)$ is equal to the $\max_{y,y' \in \{0,1\}^{|\partial j|}} \{\max_{s \in \{0,1\}} |\pi(x_j = s | x_{\partial j} = y) - \pi(x_j = s | x_{\partial j} = y')|\}$.



(a) 5 node line network



(b) 50 node line network



(c) 500 node line network

Figure 5.2: Border effect on line networks of different sizes for $\rho = 155$. The format is the same as in Figure 5.1. We observe that for a given finite value of ρ the border effect impacts less the fairness of 802.11 on large topologies. Indeed, on such topologies, links in the middle of the network have an equal access to the channel.

between j_1 and j_2 one cannot find two neighboring closed vertices. If j_1 and j_2 are separated by k vertices on the line, the probability of not finding two neighboring closed vertices among these k vertices is no more than $(1 - 1/(1+\rho)^2)^{\lfloor k/2 \rfloor}$. This probability tends to zero when k goes to infinity. Therefore, it is almost surely (a.s.) impossible to find an infinite open path. We then apply Corollary 2 in [vdBM94], which states that the measure is unique if the probability to find an infinite open path is zero in the site percolation process described above. It follows from the uniqueness of the measure that all links have the same marginal probability of being active in an infinite network.

We finish the proof by taking a finite segment of this infinite network and by applying a border condition at the two extremities of the segment. Consider two possible different border conditions and look at the probability that a link is active at the center under each border condition. Corollary 1 in [vdBM94] states that the difference between the two probabilities is less than the probability of finding an open path joining that link to any of the two extremities of the network. As the border moves far away from the considered link, this latter probability tends to

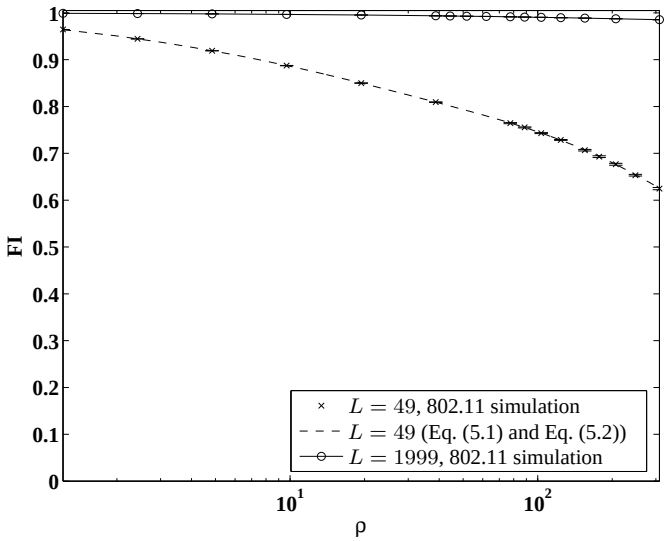


Figure 5.3: Jain's Fairness Index (FI) computed at the end of a very long simulation of 802.11 on a 50 and a 2000 node line network. The Jain's Fairness Index is plotted as a function of the access intensity ρ. The agreement between the simulation results and the values given by Equations (5.1), (5.2) is excellent.

zero and the link has therefore the same probability of being active under the two border conditions.

Moreover, as any border condition leads asymptotically to the same probability distribution in the center of the network, this distribution is also equal to the distribution in the case of an infinite network . Therefore, the marginal probabilities that links are active in the center of the network are asymptotically equal. □

Theorem 5.1 implies that if we move sufficiently far away from the border of the network, the border effect ultimately disappears, as illustrated by Figure 5.2(c). Consequently, in the limit when the number of links goes to infinity, *the idealized 802.11 protocol is long-term fair and all links have the same probability to access the channel*.

Figure 5.3 summarizes the long-term fairness performance of 802.11 on the line topology using the Jain's fairness index.

5.2 Grid Topology

In this section we study the long-term fairness of 802.11 on the grid topology. We start with illustrative simulation results.

Figure 5.4 shows the fairness index of the protocol as a function of the access intensity ρ for two different sizes of grid topologies. On the small topology, the

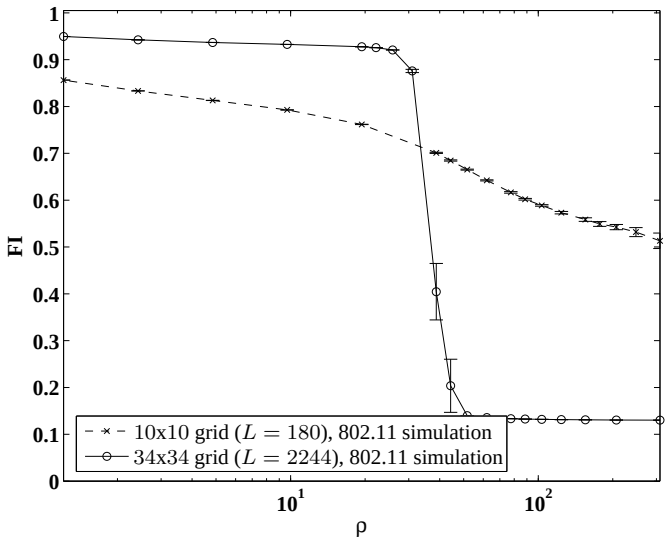


Figure 5.4: Jain's Fairness Index (FI) computed at the end of a very long simulation of 802.11 on a 10×10 and a 34×34 grid network. In the small topology the fairness degrades progressively as ρ increases. On the large topology we observe the sharp drop in the fairness of the protocol for values of ρ between 30 and 45, from values close to 1 to values slightly above $1/8$.

fairness index decreases progressively as the value of ρ increases. On the large topology, the fairness index is very high for values of ρ below 30, drops sharply for values of ρ between 30 and 45 and stabilizes at a value slightly above $1/8$ for values of ρ above 45. Figure 5.5 confirms this behavior. It shows the histogram of the link access probabilities for two values of ρ. At $\rho = 26$ (a value of ρ below 30) all links have a very similar access probability to the channel. In contrast, at $\rho = 78$ (a value of ρ above 45), two categories of links clearly appear: approximately $1/8$ of the links have a very high channel access probability, whereas the remaining $7/8$ of the links have almost no access to the channel.

We next show that in large grid networks, 802.11 exhibits a phase transition. For values of ρ below some constant ρ_{c_1}, the protocol becomes long-term fair as the number of links goes to infinity. It thus has a similar behavior as on the line topology. In contrast, above some constant ρ_{c_2} (which we conjecture to be equal to ρ_{c_1}), the effect of the border always propagates inside the network, independently of its size, and the protocol is not long-term fair. Figure 5.4 illustrates clearly this phase transition phenomenon.

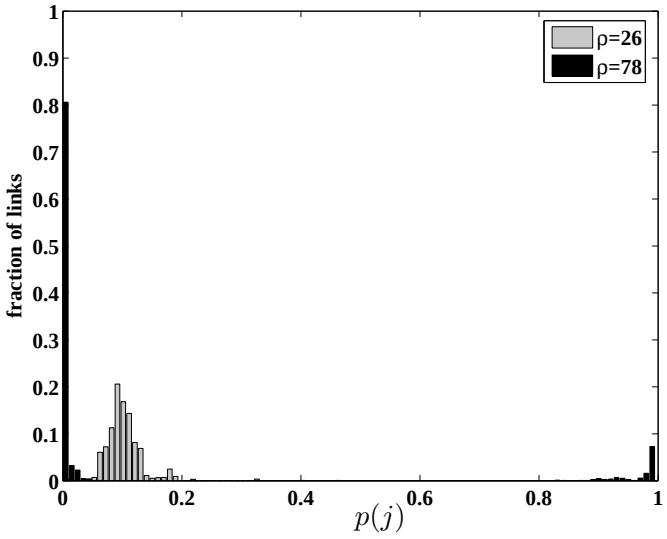


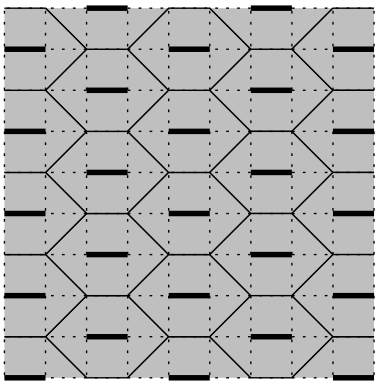
Figure 5.5: Border effect on a 34×34 node grid network. This figure represents the fraction of links with a given channel access probability in two simulations, one with $\rho = 26$ and one with $\rho = 78$. For $\rho = 26$ all links have similar access probabilities; the protocol has a high level of fairness. However, for $\rho = 78$, a small fraction of the links monopolize the access to the channel; the protocol has a low level of fairness.

5.2.1 Sub-Critical Regime

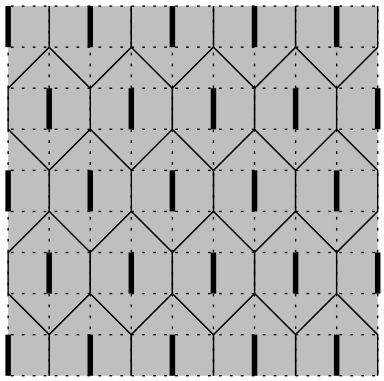
Theorem 5.2 *Consider an L-link grid network. There is some $0 < \rho_{c_1} < \infty$ such that when $L \to \infty$, the border effect disappears for values of $\rho < \rho_{c_1}$.*

Proof: This theorem is proved in the same way as Theorem 5.1. We consider first an infinite grid network. On its contention graph, we define a site percolation process by declaring each site open with probability $\rho/(1+\rho)$ independently. This site percolation process is less trivial than in one dimension: for sufficiently large values of ρ, one might find infinite open paths. However, if ρ is below some threshold ρ_{c_1}, which is larger than or equal to the site percolation threshold of the contention graph (which is not known analytically, but is finite and non-zero, see Theorem 1.33 in [Gri99]) the probability to find an infinite open path is zero. Therefore the same argument as in Theorem 5.1 can be used. We conclude that if ρ is sufficiently small, when the border moves away from the center, the probability of a link of being active is asymptotically independent of the location of the link and of the border condition. □

As a result, *in the sub-critical regime the protocol is long-term fair on large grid networks*.



(a) A transmission pattern of maximal spatial reuse with horizontal active links



(b) A transmission pattern of maximal spatial reuse with vertical active links

Figure 5.6: Transmission patterns of maximal spatial reuse on the grid network. In the packing formalism (see Section 3.1.3) the space occupied by an active link is an hexagon of area four.

5.2.2 Super-Critical Regime

Theorem 5.3 *Consider an L-link grid network. There is some* $\rho_{c_1} \leq \rho_{c_2} < \infty$ *such that when* $L \to \infty$*, the border effect* does not *disappear for values of* $\rho > \rho_{c_2}$.

To prove this theorem, we use Peierls' argument [KS80]. Peierls introduced this argument in 1936 to show that the two-dimensional Ising model exhibits a phase transition. It is now a classical argument in statistical physics. The remainder of this section is devoted to the proof of Theorem 5.3 and is structured as follows: We begin with a few definitions; next, we present the outline of Peierls' argument; finally, we present the main steps necessary to apply the method of [Hei74] to our setting.

Contours and their Properties

In an infinite grid network, if we abstract the link directions, there are eight transmission patterns of maximal spatial reuse, four with horizontal active links and four with vertical active links. The four horizontal (resp., vertical) transmission patterns can be obtained from each other by translation. In these patterns the space occupied by an active link is a hexagon of area four. Figure 5.6 shows two transmission patterns of maximal spatial reuse. Each hexagon or link[3] belongs to exactly one

[3] As each undirected link corresponds to one hexagon and vice-versa, we use interchangeably link and hexagon

of these eight patterns, hereafter called the *phase* of the hexagon. Moreover, each transmission pattern corresponds to a set of non-overlapping hexagons.

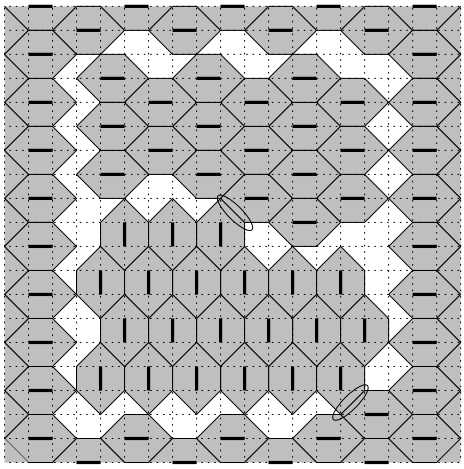
Peierls' argument relies on the notion of *contour*. For the following definition it is helpful to consider the examples depicted in Figures 5.7(a) and 5.9(a).

Definition 5.2 *Consider an arbitrary transmission pattern (i.e., a set of non-overlapping hexagons). An element (of contour) is either an area not covered by hexagons, or a common edge of two hexagons of two different phases. Two elements are connected if they share at least one vertex. A contour is a set of connected elements, which is maximal.*

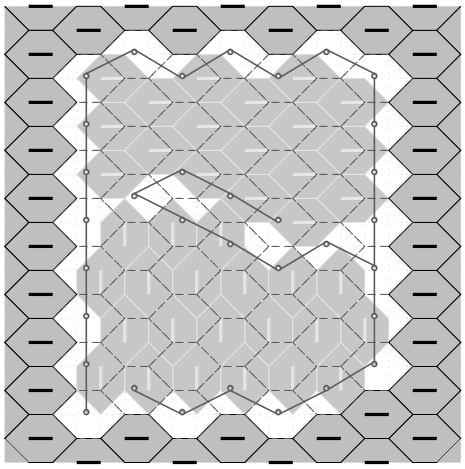
Let C be a contour. C is delimited by an outer polygon and possibly several inner polygons (for example, in Figure 5.7(a) there are two inner polygons). All edges of the outer polygon necessarily belong to hexagons of the same phase, called the *phase of the contour* C. The interior of C is the area contained inside its outer polygon. This area can be tiled with hexagons of the phase of C (Figure 5.7(b)). If such a hexagon intersects a contour element, we say that this hexagon *belongs* to the contour or that the contour *contains* the hexagon. The length of a contour is defined as the number of hexagons belonging to the contour and is denoted by $L(C)$.

Lemma 5.1 *The number of contours of length k containing a given hexagon is upper bounded by q^k, with $q = 15 \cdot 2^{16}$.*

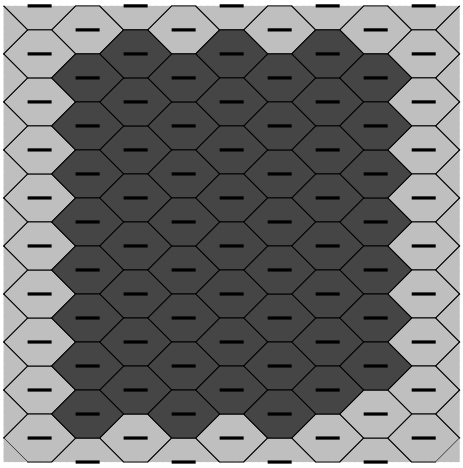
Proof: The given hexagon is necessarily of the phase of the contour. Moreover, we know that, by definition, a contour of length k is contained in k connected hexagons of its phase. In the adjacency graph of these hexagons, we can always find a spanning tree (see Figure 5.7(b)). We can thus associate to each contour of length k a spanning tree with k vertices. The first step to obtain our bound is thus to enumerate all the possible spanning trees having as a root the given hexagon. We do so using a depth-first search approach. The given hexagon has 6 adjacent hexagons, i.e., there are at most 6 possible directions for the first move of our search. However, after the first move there are only at most 5 possible directions (as we are building a tree). At each move, we choose one of these possible directions, and choose between 3 actions: (1) do nothing, (2) put a mark to remember a branching hexagon in the spanning tree, and (3) go back to the last mark, which means that we met a leaf hexagon and need to backtrack. In total we thus have $3 \cdot 5 = 15$ possibilities per move (except for the first move where we have $3 \cdot 6 = 18$ possibilities). However, a spanning tree does not uniquely determine a contour. The second step to obtain our bound is to upper bound the number of contours per spanning tree. For each hexagon in the spanning tree, i.e., each hexagon belonging to the contour, the actual shape of the covered region must be specified and we observe that we have at most 2^{16} possibilities (see Figure 5.8). An upper bound on the number of contours of length k using a specific hexagon is thus $18 \cdot 15^{-2} \cdot q^k$, which is smaller than q^k. □



(a) Transmission pattern with a single contour C. The contour C corresponds to the white area on the figure but also contains the two encircled edges.



(b) The interior of the contour is tiled with hexagons in the phase of C. The contour is contained in $L(C) = 33$ of these hexagons which can be connected by a spanning tree.



(c) Contour removal operation. The number of additional hexagons (compared to the original transmission pattern in (a)) is 16.

Figure 5.7: Removal of a single contour C.

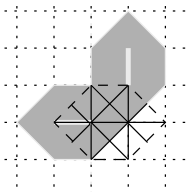


Figure 5.8: An hexagon (the dashed hexagon on the figure) can be divided into 16 triangular parts of identical size. Each of these parts can be covered or not. In the figure, 11 of the 16 parts are covered. The covered region of the hexagon is specified by these covered parts.

Before we begin with Peierls' argument we need one last notion, the notion of C-inner contour.

Definition 5.3 *A C-inner contour is a contour inside C but not inside any other contour.*

Figure 5.9(a) gives an example of a transmission pattern with a contour C and two C-inner contours.

Outline of Peierls' Argument

Assume that the border of the network has one fixed phase denoted by A. As the border of the network is set to A, an active link that does not belong to phase A must be inside at least one contour. An upper bound on $P(i \notin A)$, the probability of finding an active link i in another phase than A, is therefore the probability of having an outer contour (i.e., a contour that is not inside another contour and thus necessarily in phase A) around link i.

Let y denote an arbitrary transmission pattern on the network. The probability of a given contour C is

$$P(C) = \frac{\sum_{y \ni C} \rho^{n(y)}}{\sum_y \rho^{n(y)}}. \tag{5.3}$$

Assume that we can find an upper bound on $P(C)$ of the form

$$P(C) < \rho^{-\varepsilon L(C)} \tag{5.4}$$

where $\varepsilon > 0$. We would then have

$$P(i \notin A) = \sum_{C \ni i} P(C) < \sum_{C \ni i} \rho^{-\varepsilon L(C)} < \sum_l l^2 q^l \rho^{-\varepsilon l}$$

where l^2 is an upper bound on the number of possible starting hexagons for a contour of length l, and thus $l^2 q^l$ is an upper bound on the number of contours C

of length l surrounding link i. By choosing ρ large enough we can certainly have

$$P(i \notin A) < 1 - \frac{1}{8} = \frac{7}{8}.$$

As 8 is the number of phases, this implies that phase A is strictly more probable at link i than at least one of the other phases. Therefore the effect of the border is felt at link i no matter how far the border is.

Upper Bound on $P(C)$

The main difficulty of the proof is to obtain the upper bound (5.4) on the probability $P(C)$ of a given outer contour C. In principle, the idea is quite simple. Let $\bar{y}$ denote an arbitrary transmission pattern with contour C, Equation (5.3) can be rewritten as

$$P(C) = \frac{\sum_{\bar{y}} \rho^{n(\bar{y})}}{\sum_{y} \rho^{n(y)}}.$$

If to each $\bar{y}$ we can map another transmission pattern y' with $\epsilon L(C)$ more hexagons and if this mapping is one-to-one, then, to each term in the numerator we have associated a term in the denominator such that $n(y') - n(\bar{y}) = \epsilon L(C)$. As reducing the number of terms at the denominator can only increase the fraction, we have

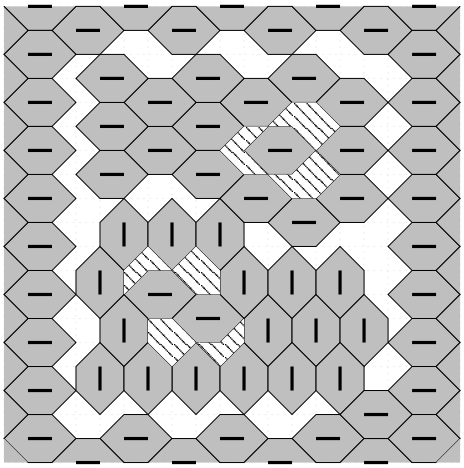
$$P(C) < \frac{\sum_{\bar{y}} \rho^{n(\bar{y})}}{\sum_{y'} \rho^{n(y')}} = \frac{\sum_{\bar{y}} \rho^{n(\bar{y})}}{\sum_{\bar{y}} \rho^{n(\bar{y})+\epsilon L(C)}} = \rho^{-\epsilon L(C)},$$

which is exactly the upper bound (5.4) with $\varepsilon = \epsilon$.

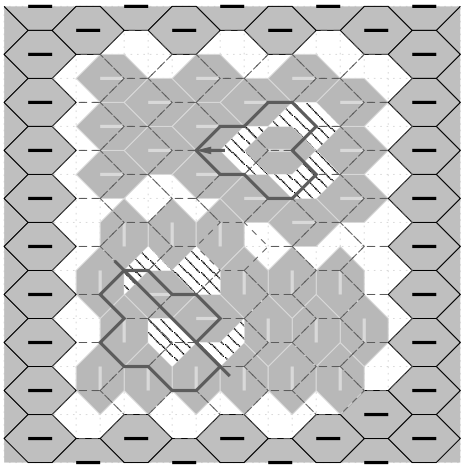
Contour Removal and Transformation: The challenge is thus to find the one-to-one mapping $\bar{y} \leftrightarrow y'$. The mapping is the simplest when the original transmission pattern $\bar{y}$ includes only the outer contour C and no C-inner contour. Figures 5.7(a) (for $\bar{y}$) and 5.7(c) (for y') illustrate this mapping. To obtain the resulting pattern y', we start from $\bar{y}$, and we fill the interior of C with hexagons of phase A, which is the phase of the outer contour C. In y', the contour C is no longer visible; this is why the mapping between $\bar{y}$ and y' is often called a *contour removal* . It is easy to see that the number of additional hexagons in y' is proportional to $L(C)$ and is at least equal to $\epsilon L(C)$ where ϵ is a lower bound on the average uncovered area per unit of contour length.

In general, however, the contour C contains C-inner contours (Figure 5.9(a)), which can be divided in three categories (we assume without loss of generality that the phase of C is horizontal):

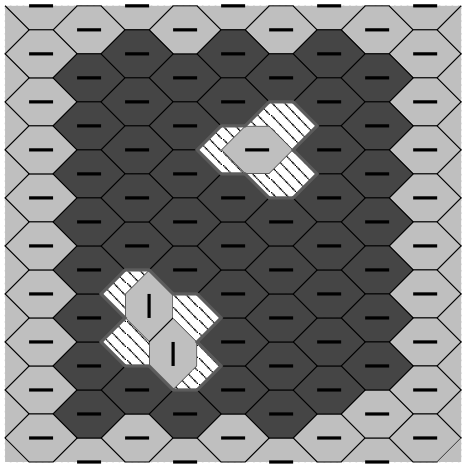
1. C-inner contours in the same phase A as C
2. C-inner contours in a horizontal phase different from the phase A of C
3. C-inner contours in a vertical phase.



(a) A transmission pattern with a contour C (white area) and two C-inner contours (dashed areas).



(b) Transformation of the C-inner contours: the C-inner contour in the horizontal phase (top) is translated, whereas the one in the vertical phase (bottom) is reflected. The outside polygon of the transformed contours together with the translation direction and the reflexion axis used to obtain them are represented by thick gray lines.



(c) Once the C-inner contours are transformed into contours in the phase of C, the contour C is removed.

Figure 5.9: Contour C with two C-inner contours. The two C-inner contours are transformed and the contour C is removed.

To obtain y' from $\bar{y}$, we first *transform* the C-inner contours into contours of phase A. The C-inner contours in the first category are already in phase A and thus do not need to be transformed. The C-inner contours in the second category are translated to become contours of phase A[4]. Finally, to transform the C-inner contours in the third category into contours of phase A, we reflect them in a diagonal axis of the grid lattice[5]. Figure 5.9(b) illustrates the details of this mapping on a small example.

Unfortunately, the reflected contours might touch[6] the contour C. The contours touching C are removed together with C. Figure 5.10 gives an example of such a situation. Moreover, the reflected contours might touch each other, in which case they are reflected in a common axis, but their joint reflection might also touch C, in which case the contours need to be removed, and so on.

Heilmann [Hei74] describes an iterative process to determine which contours should be removed and which contours should be transformed: the same process can be applied in our setting. The result of this process is a set K of contours to transform and a set T of contours to remove. To obtain y' from $\bar{y}$, we first transform all the contours in K and we then remove all the contours in T. In y', the interior of C (as well as the interior of all the other contours in T) is thus filled with hexagons of phase A, whereas the interior of the transformed contours is left unchanged (see Figure 5.9(c)).

Enumeration of the contours to be removed: In the case where the set T of contours to remove contains at least one C-inner contour, the mapping between $\bar{y}$ and y' may no longer be one-to-one, unless we specify the set T. We have therefore

$$P(C) < \sum_{T \ni C} \rho^{-\epsilon L(T)}$$

where $L(T)$ is the sum of the lengths of the contours in set T. We rewrite this inequality as

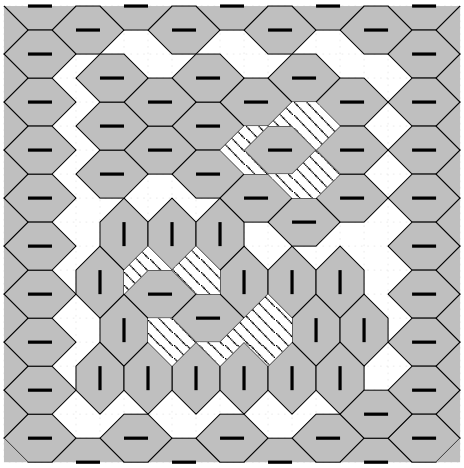
$$P(C) < \rho^{-\epsilon L(C)} \sum_{T \ni C} \rho^{-\epsilon L(T \setminus C)}. \tag{5.5}$$

In the iterative process of [Hei74], the only contour initially in the set T is the outer contour C. C-inner contours are then added to T if and only if their reflection (or their joint reflection) touches a contour already in T. To obtain an upper bound of the type (5.4), we need to enumerate all possible sets T. We do not go into the details of this lengthy and rather complicated enumeration which can be found in [Hei74]. Here we restrict ourself to the enumeration of the sets $\tilde{T} \subset T$ with two

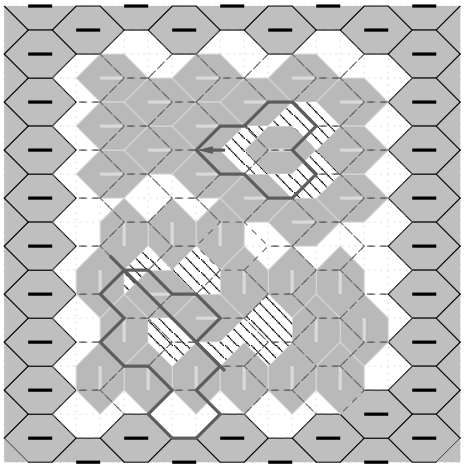
[4] We always choose the shortest possible translation, if there are two possibilities, we favor moving up.

[5] The reflection axis is the (top-left-bottom-right) diagonal that splits the interior of the contour in two areas as equal as possible (if several reflection axis satisfy the above condition we favor the upper most one).

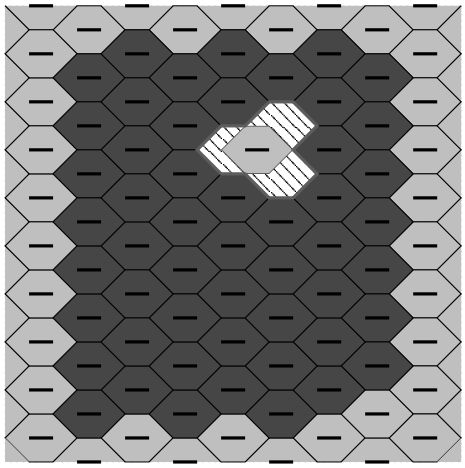
[6] We say that two contours touch if two of their hexagons (one belonging to each contour) overlap or share a common edge.



(a) A transmission pattern with a contour C (white area) and two C-inner contours (dashed areas). The C-inner contour in the vertical phase contains one more hexagon than in Figure 5.9(a).



(b) Transformation of the C-inner contours. The reflection of the vertical C-inner touches the contour C.



(c) The resulting transmission pattern. The horizontal C-inner contour has been transformed and the vertical C-inner contour has been removed together with the outer contour C.

Figure 5.10: Contour C with two C-inner contours. One C-inner contour is transformed while the other one is removed.

contours: the outer contour C and a C-inner contour of length k whose reflection touches C at a specific contact point. This is sufficient to put our setting in gear with the recursive procedure in [Hei74]: Using the contribution of the set $\tilde{T}$ to the sum in (5.5) as an initial seed for this recursive procedure will allow us to directly apply the bound of [Hei74] for $P(C)$.

Before we start to enumerate the sets $\tilde{T}$, we need to make the notion of contact point precise. Let us call D the C-inner contour in $\tilde{T}$. The contour C (resp., D) is contained in $L(C)$ (resp., in $L(D) = k$) hexagons in the phase of C (resp., of D). A necessary condition for contour C and the reflection of D to touch each other is that two of their hexagons touch each other (i.e., overlap or share a common edge). A contact point between a contour C and a contour D can thus be specified as two hexagons (one from C and one from the reflection of D) that touch each other. Given the first hexagon, there are exactly 33 possible relative positions for a second hexagon. Therefore, per hexagon of C, or equivalently per unit of length of C, there are 33 possible points of contact with another contour. The enumeration method in [Hei74] requires that there is only one contact point per unit of contour length. To satisfy this constraint, we redefine the length of a contour as 33 times the number of its hexagons. We use a symbol prime to denote this new length, e.g., $L'(C) = 33L(C)$.

Using Lemma 5.1, we find that an upper bound on the number of contours of length k' touching C at a specific contact point is $q^{k'/33}$ and their contribution to $P(C)$ is less than $(q\rho^{-\epsilon})^{k'/33}$.

We then use the same recursion as in [Hei74] to compute the contribution of all sets T. Adapting relations (3.27) and (3.28) of [Hei74], we finally obtain that for $n = \frac{1}{33}(\epsilon \log \rho - \log q) > f_0$ where f_0 is a constant

$$P(C) < \rho^{-L(C)\left(\epsilon - 33B(\log \rho)^{-1}66^{-n}\right)}$$

with $B = \frac{1}{256}(1 + \frac{66}{f_0 - 1})$. For ρ large enough, $\varepsilon = \epsilon - 33B(\log \rho)^{-1}66^{-n}$ is positive. Consequently, we can use this upper bound on $P(C)$ to conclude Peierls' argument as presented in the outline.

As a result, in the *super-critical regime 802.11 is not long-term fair even if the size of the network goes to infinity*. The effect of the border does not disappear, and a positive fraction of the links get better access to the channel.

5.3 Related Work

The tendency of IEEE 802.11 to starve some of the network links while giving good access to the others is well known. It has been observed experimentally on small testbed topologies [CDL05] and it has also been investigated analytically in several papers [WK05, GSK05, GSK06, MT06]. The models of [GSK05, MT06] predict the unfairness of 802.11 on small (four and six node respectively) topologies. In particular [MT06] notices the role played by the value of the contention window

in the level of unfairness observed. [GSK06] extends the approach of [GSK05] to consider more general network topologies. Although, these papers predict the starvation phenomenon, they fail to really explain it. [WK05] is really the first to notice the special role played by the patterns of maximal spatial reuse and to state that at high access intensities only the links belonging to these patterns get a good access to the channel. Yet, [WK05] provides quantitative results only for the same 6 node topology as [MT06] and does not investigate the long-term fairness properties of the protocol when the size of the network goes to infinity. We used results from the theory of Markov random fields (see [KS80] for a nice introduction) to do so. In particular, we applied two types of arguments. A percolation argument [vdB93, vdBM94] to prove that 802.11 can be long-term fair in large network topologies and the so-called *Peierls argument* to prove that this is not always the case. The simplest version of Peierls' argument can be found in [KS80]. Unfortunately, to characterize the behavior of 802.11 on large two-dimensional topologies, we had to resort to a more elaborate version of this argument. We adapted some of the techniques of [HP73, Hei74] to achieve this goal. Several works [Lou90, KRM02, LRZ06] already showed the existence of a phase transition phenomenon in loss networks. However, these papers only consider networks with relatively simple structures (i.e., tree or bipartite networks) and consequently their proof techniques cannot be used in our setting.

5.4 Conclusion

Topological inequalities in the network are the cause of the observed channel access unfairness in 802.11. In particular, in regular topologies, nodes at the border are typically favored. This border effect also affects, to a certain extent, the nodes inside the network, depending on the access intensity of the protocol and on the dimension of the network. In some cases, the border effects propagate arbitrarily far inside the network without vanishing; this leads to severe unfairness of the protocol. In the other cases, the protocol is long-term fair. In irregular topologies, the topological inequalities are not restricted to the border of the network and the level of competition to access the channel varies from link to link. As a result, in such topologies, 802.11 is typically very unfair at all values of the access intensity even though its unfairness is more pronounced at high values, as illustrated by Figure 5.11.

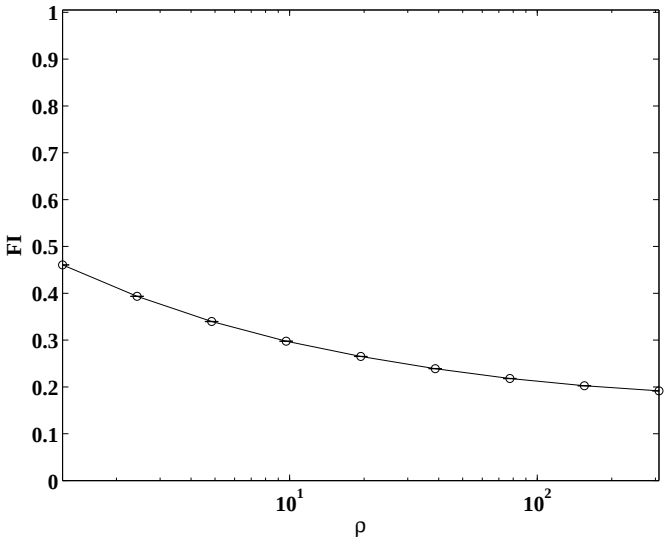


Figure 5.11: Fairness Index (FI) computed at the end of a very long simulation of 802.11 on the irregular network.

Chapter 6

Short-Term Fairness

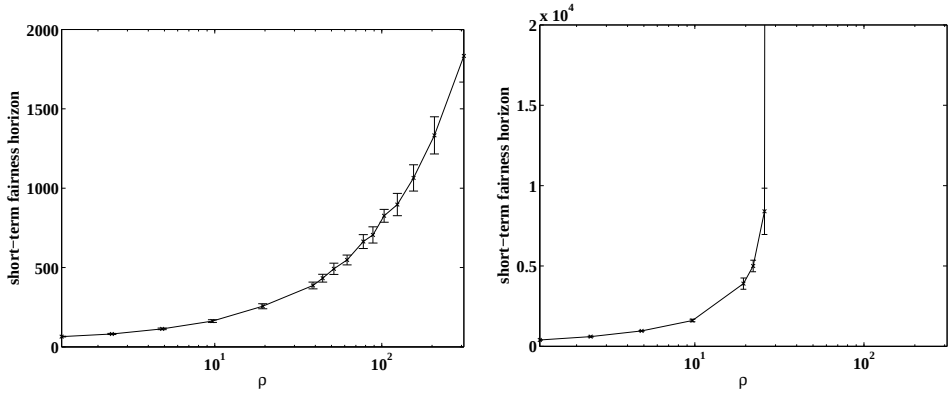
In the previous chapter, we saw that the idealized 802.11 protocol gives in many cases an equal share of the channel to each link in the center of the network in the long term. However, this does not guarantee that certain links are not starved for long periods (and then keep the channel for an equally long period, preventing the neighboring links to be activated).

In this chapter we attempt to characterize the short-term fairness of the protocol in settings where it is long-term fair. Short-term fairness is important for delay sensitive applications such as real-time audio and video. Several studies (see for example [KKB00]) also show that the lack of short-term fairness at the MAC layer can severely impact the performance of a reliable transport protocol such as TCP .

We consider different short-term fairness metrics. We start by evaluating the short-term fairness horizon of 802.11. This metric evaluates the time it takes for the protocol to reach a given level of fairness. We then consider a link-centric metric, where we compute how long on average a link has to wait to access the channel. Finally, we consider the network-wide equivalent of the previous metric and compute the time needed to switch among patterns of maximal spatial reuse.

6.1 Short-Term Fairness Horizon

We begin our study of 802.11 short-term fairness by a few illustrative simulation results. We consider only topologies where the protocol is long-term fair. Figure 6.1(a) shows how long on average we need to operate 802.11 to reach a fairness index of 0.9 in a large line network. As expected, this time, also called the *short-term fairness horizon* [KKB00] of the protocol, increases progressively with the value of the access intensity ρ. Figure 6.1(b) shows that in the grid network the short-term fairness horizon of the protocol increases progressively with ρ and suddenly diverges at a finite value of ρ, which marks the phase transition described in the previous chapter. Indeed, above the critical value of ρ, the protocol is not even long-term fair. Just below the threshold, the protocol is still long-term fair, but one has to wait a very long time to have a fair channel access among the network links.



(a) Short-term fairness horizon of 802.11 on a 2000 node line network.

(b) Short-term fairness horizon of 802.11 on a 34×34 node grid network. The value of the short-term fairness horizon explodes just before the phase transition threshold.

Figure 6.1: Short-term fairness horizon of 802.11. We measure the average time needed by the protocol to reach a fairness index of 0.9. The average exchange time μ^{-1} is equal to 1 time unit.

6.2 Average Link Access and Waiting Times

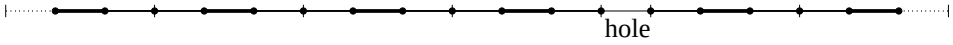
We now characterize the short-term fairness of 802.11 from a link-centric viewpoint. Our goal is to answer two questions: (i) How long on average does a link wait to get access to the channel, and (ii) how long on average does a link keep the access to the channel? Note that we say that a link has access to the channel if it is active or if it does not see the channel as busy.

In a large (infinite) line or circle network, when the access intensity ρ is finite but sufficiently large, the link access and waiting times can be estimated by describing the dynamics of holes, where a *hole* is defined as a unit space not covered by an interval corresponding to an active link (see Figure 6.2(a)). Holes separate links in different phases and are the exact one-dimensional equivalents of contours (Section 5.2.2) in the two-dimensional grid network. The density of holes (i.e., the fraction of holes per unit length) relates directly to the spatial reuse of the protocol: its average is $1 - l\sigma$.

The following proposition gives analytical expressions for the average link access and waiting times based on the dynamic motion of holes along the line. Here again, a link corresponds to two directed links. Although these expressions rely on several approximations, Figure 6.3 shows an extremely good fit with the simulation results.

(a) Hole before the jump

(b) The link (marked by a cross) next to the hole (of Figure (a)) switches from the active to the idle state (the corresponding interval is unpacked)



(c) Hole after the jump

Figure 6.2: Transmission pattern on the line network with a hole. The succession of the 3 figures represent the jump of a hole by $l = 3$ space units (event (ii)).

Proposition 6.1 *Consider an L-link line network with ρ large. When $L \to \infty$, a link waits to get access the channel for on average*

$$\frac{1}{\mu}\left(\frac{1}{\sigma} - 1\right)\left(1 + \frac{\sigma}{1 - l\sigma}\right) \tag{6.1}$$

time units, and keeps the access to the channel for on average

$$\frac{1}{\mu}\left(1 + \frac{\sigma}{1 - l\sigma}\right) \tag{6.2}$$

time units where σ is given by Theorem 4.1.

Derivation of formulæ (6.1) and (6.2): In this computation, we assume that ρ is large so that the density of holes in the network is small. Moreover, we assume a fixed density of holes equal to $1 - l\sigma$.

A link switches from active to idle state at rate μ. When a link next to a hole becomes idle (Figure 6.2(b)), we have two equally likely events: (i) the same link becomes active again (we are back to the pattern in Figure 6.2(a)) or (ii) another link becomes active (Figure 6.2(c)). In terms of holes, event (i) means that the hole does not move, whereas event (ii) means that the hole jumps by l space units; we call the latter a *useful jump*. A hole makes therefore a useful jump at rate $\mu/2$, where the factor $1/2$ is the probability of event (ii).

Denote by $n = L\sigma$ the number of active links and by $k = L(1 - l\sigma)$ the number of holes in the network (n and k are temporary notations used for this derivation). The small number of holes makes it reasonable to assume that they move independently of each other at rate μ (i.e., $\mu/2$ in a given direction). In particular, a hole that is next to another hole might jump above it. We call this one space unit jump event (iii). Event (iii), similarly to event (i), does not change the transmission pattern.

The k holes make a jump at a total rate of $k\mu$. However, only events of type (ii), holes which make a useful jump over an active link, cause a change of transmission

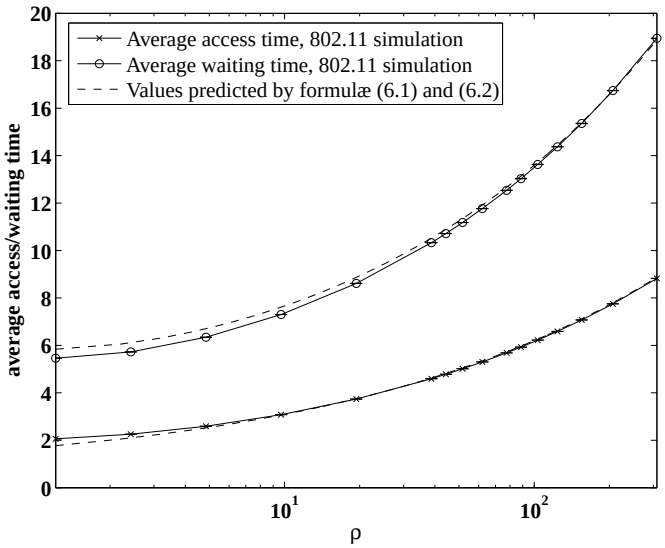


Figure 6.3: Average time a link waits to get access (resp., keeps the access) to the channel in a 2000 node line network with $\mu = 1$. The simulation results are presented with the 95% confidence intervals.

pattern. The probability of a useful jump is $q = \frac{n}{k-1+n}$ and consequently useful jumps occur at rate $k\mu q$.

Given that we have n active links in the network, the rate of useful jumps per active link is thus $k\mu q/n$. This implies that a link gets access to the channel for on average $n/(k\mu q)$ time units. Replacing n by $L\sigma$ and k by $L(1 - l\sigma)$ gives (6.2).

Similarly, the rate of useful jumps per inactive link is $(k\mu q)/(k + n(l - 1))$, from which we easily obtain (6.1). □

This reasoning, however, does not apply to the grid network: As we have seen in the previous section, phases are delimited by contours, not by holes. The dynamics of contours is too complicated to allow a simple method as the one above to give reasonable approximations of the waiting times. In any case, such a characterization would not make sense in the super-critical regime where we have two categories of links with very different access and waiting times. Figure 6.4 illustrates the mean access and waiting times obtained by simulation for the grid network when the protocol operates in the sub-critical regime.

6.3 Rotation between Patterns of Maximal Spatial Reuse

The 802.11 protocol is not long-term fair on small line or grid topologies because of the border effect. However, it is long-term fair on small circle topologies. In such topologies, if L is a multiple of l and if we make abstraction of the link directions, there are l transmission patterns of maximal spatial reuse (Figure 6.5). Each

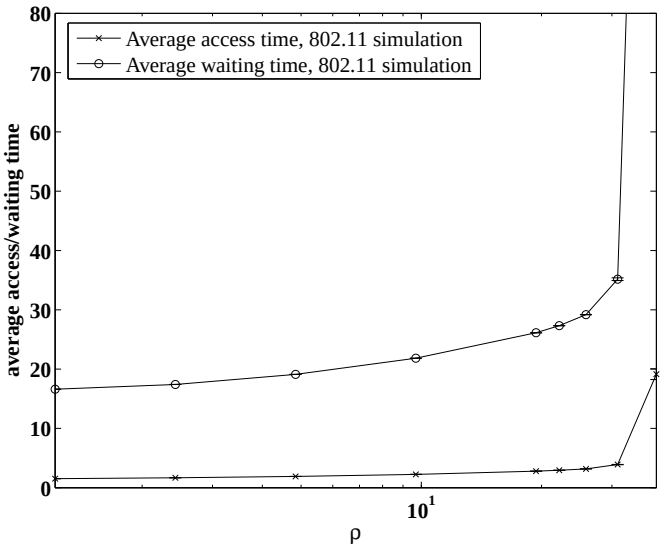


Figure 6.4: Average time a link waits to get access (resp., keeps the access) to the channel in a 34×34 node grid network. The format is the same as in Figure 6.3 but we consider only the values of ρ for which the protocol operates in the sub-critical regime.

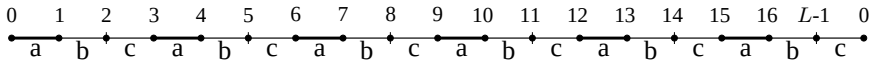


Figure 6.5: On a circle topology with $l = 3$ and where L is a multiple of 3 there are 3 transmission patterns of maximal spatial reuse. Links marked with letter a (resp., b and c) belong to the first (resp., second and third) pattern of maximal spatial reuse.

link (or the corresponding interval of size l) belongs to one of these transmission patterns. To characterize the short-term fairness of the protocol, we can thus look, in addition to the average link access and waiting times, at the average time needed to switch from a pattern of maximal spatial reuse to another.

Figure 6.6 shows this time for a circle topology of size $L = 30$. For small values of ρ the protocol is seldom in patterns of maximal spatial reuse, which explains why the time to switch between them is large. The curve then has a clear minimum for a value $\rho_{min} \simeq 10$. As ρ increases above this value, the protocol spends more time in the patterns of maximal spatial reuse but it becomes increasingly hard to alternate between these patterns. For large values of ρ, we observe that the expected time to switch between patterns of maximal spatial reuse increases linearly with ρ.

In practice, the value ρ_{min} might thus be an interesting operating point for the protocol. Indeed, access intensities $\rho < \rho_{min}$ lead to a poor spatial reuse which

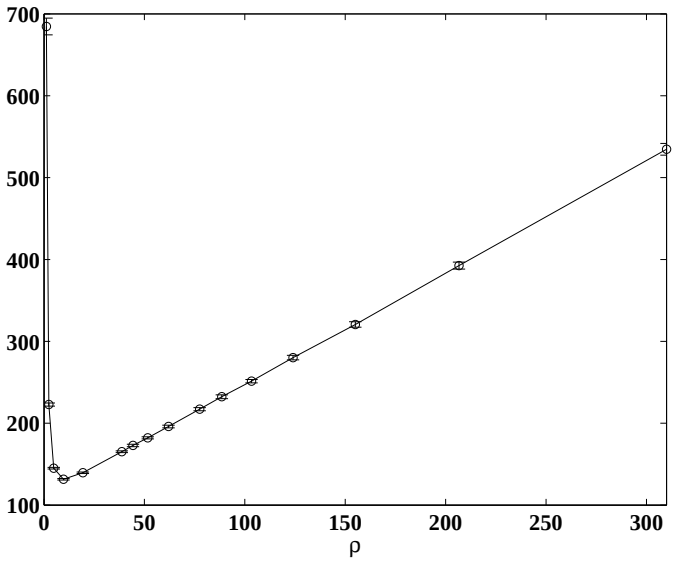


Figure 6.6: Average time needed to switch from a pattern of maximal spatial reuse to another one on a $L = 30$ circle topology with $\mu = 1$. Observe that the minimum is obtained for $\rho_{min} \simeq 10$.

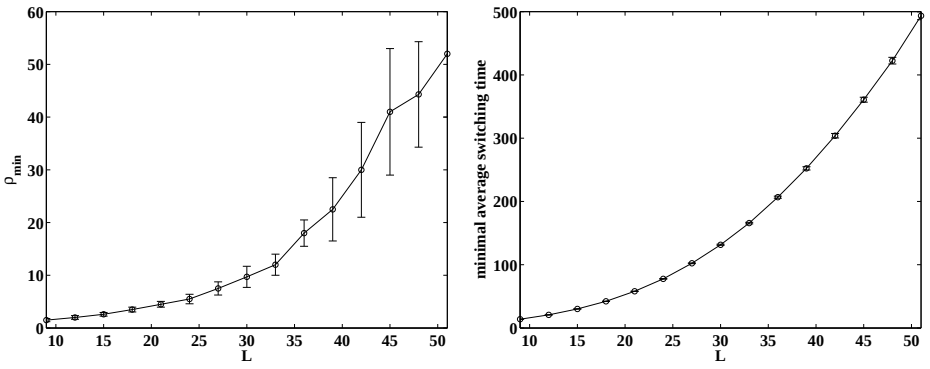
can be increased without prejudice on the fairness. Conversely, values of $\rho > \rho_{min}$ correspond to a good spatial reuse, but at the cost of a reduced short-term fairness. Figure 6.7(a) illustrates how ρ_{min} and the corresponding value of the expected switching time change with the size of the circle topology.

In addition, at larger ρ, the slope of the linear increase characterizes the penalty incurred in terms of fairness for a unit increase of the access intensity ρ. We now present an approach that allows us to compute this important quantity analytically.

We consider a circle topology with a finite number L of links (where L is a multiple of l), and let $\rho \to \infty$ (in practice it is enough to operate at large values of ρ, so that the protocol spends most of its time in patterns of maximal spatial reuse). More precisely, we fixed μ to a strictly positive value and let $\lambda \to \infty$. We attempt to characterize the linear increase of the expected switching time between maximal spatial reuse patterns.

Assume (without loss of generality) that the Markov chain is currently in one of the transmission patterns of maximal spatial reuse, and denote it by x^*. In the state transition diagram of the simplified Markov chain[1], x^* is connected to $n(x^*) = L/l$ states at the level $n(x^*) - 1$. As $\lambda \gg \mu$, a transition to one of these states is most likely immediately followed by the reverse transition (back to x^*). Yet, to switch to another pattern of maximal spatial reuse, *two* consecutive active

[1]Remember that in the simplified Markov chain we abstract the link directions (see Section 3.1.1, page 12).



(a) Value of ρ_{min}

(b) Average time to switch between two patterns of maximal spatial reuse at $\rho = \rho_{min}$

Figure 6.7: Value of (resp., at) the operating point that guarantees a good trade-off between fairness and spatial reuse ($\mu = 1$). As the size L of the network increases, so do the confidence intervals in Figure 6.7(a). This is due to the fact that the curve of Figure 6.6 becomes flatter around ρ_{min} for larger values of L.

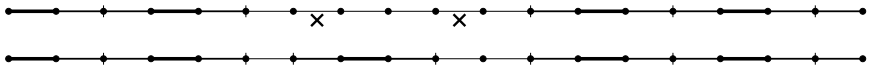


Figure 6.8: First replacement. We start with the transmission pattern of maximal spatial reuse represented in Figure 6.5. Two consecutive active links (marked by a cross) in this pattern become idle (the corresponding intervals are unpacked) and are replaced by a different active link.

links in x^* (or their corresponding intervals), must be replaced by one active link that does not belong to x^*. We call this event *the first replacement*. A possible first replacement for the pattern of maximal spatial reuse of Figure 6.5 is represented in Figure 6.8.

Lemma 6.1 *For $\lambda \to \infty$ ($\mu > 0$ finite), the replacement of two consecutive intervals in a pattern of maximal spatial reuse by an interval that does not belong to this pattern happens at rate*

$$r(L,l) \cdot \frac{\mu}{\rho} \quad \textit{where} \quad r(L,l) = \frac{L(l-1)}{l(l+1)}. \tag{6.3}$$

Proof: Denote by h_{x^*} the expected time to reach, from state x^*, a state that contains $n(x^*) - 2$ adjacent intervals present in x^*. Using a first step analysis (see for example [Bré98], p. 65) we find:

$$h_{x^*} = \frac{1}{n(x^*)\mu}\left(1 + n(x^*)\mu\left(\frac{1}{2\mu + \lambda'}\left(1 + 2\mu \cdot 0 + \lambda' h_{x^*}\right)\right)\right)$$

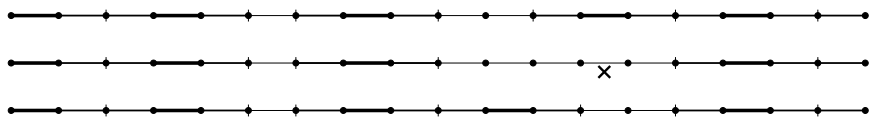


Figure 6.9: Example of a subsequent replacement. An active link next to a hole becomes idle and is immediately replaced by another active link.

where $\lambda' = 2\lambda$ since a link corresponds to two directed links. Solving the equation and using $n(x^*) = L/l$ we obtain

$$h_{x^*} = \frac{1}{2\mu} + \frac{l(2\mu + \lambda')}{2L\mu^2}.$$

This means that for $\lambda \to \infty$ (and $\mu > 0$ finite) the expected time needed to remove 2 adjacent intervals from x^* grows like $\frac{2l\lambda}{2L\mu^2} = \frac{l}{L}\frac{\rho}{\mu}$. Once we have removed 2 adjacent intervals from x^* we have created a vacant space of size $2l$ (see Figure 6.8). In this vacant space, there are $l + 1$ ways to add an interval of length l,[2] but only $l - 1$ ways to add an interval different from the 2 intervals that were removed. As the position of the new interval is selected uniformly at random among the set of feasible positions, we obtain

$$r(L, l) \cdot \frac{\mu}{\rho} = \frac{l - 1}{l + 1}\left(\frac{L}{l}\frac{\mu}{\rho}\right)$$

which is the desired result. □

Once two consecutive active links in x^* have been replaced by an active link that does not belong to x^* (first replacement), any active link that becomes idle is instantly replaced by another (or by the same) active link (subsequent replacements, Figure 6.9). Consequently, it takes some random finite time to reach a pattern of maximal spatial reuse, but this time is negligible for the asymptotics we seek. What we do need, is the probability $q(L, l)$ to reach another pattern of maximal spatial reuse than x^*. Suppose we can calculate this probability, then we have the following result:

Theorem 6.1 *For $\lambda \to \infty$ ($\mu > 0$ finite), the expected time to switch from a transmission pattern of maximal spatial reuse x^* to another pattern of maximal spatial reuse is*

$$\frac{1}{q(L, l) \cdot r(L, l)}\frac{\rho}{\mu} + O(1)$$

where $r(L, l)$ is defined by (6.3) and where $q(L, l)$ is the probability to reach another pattern of maximal spatial reuse than x^* once two consecutive intervals of x^* have been replaced by an interval that does not belong to x^*.

[2] $\lambda \gg \mu$ implies that the probability of removing one more interval is negligible when compared with the probability of adding an interval.

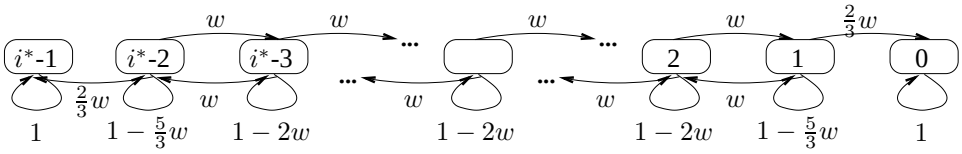


Figure 6.10: Random walk describing the sequence of replacement operations. The random walk is in state k if the current transmission pattern shares k intervals with pattern x_1^* (and $i^* - 1 - k$ intervals with x_2^*). The transitions in the random walk correspond to replacements. $w = \frac{1}{i^*-1}$ is equal to the probability of removing an interval from x_1^* that is next to a hole (i.e., $\frac{2}{i^*-1}$) and replacing it by an interval from x_2^* (which happens with probability $\frac{1}{2}$) or vice-versa.

The problem then reduces to the calculation of $q(L, l)$, which unfortunately is not an easy task. We briefly explain how this can be done for the case $l = 2$, by far the simplest, and comment on how to extend this approach to other (more realistic) values of l.

In the case where $l = 2$, we have only two patterns of maximal spatial reuse (x_1^* and x_2^*) with $i^* = L/2$ active links. Assume the Markov chain is currently in the state x_1^* and that we replace two consecutive interval in x_1^* by an interval from x_2^*. We want to compute the probability $q(L, 2)$ that starting from such a pattern we reach x_2^* before x_1^*.

As a result of the first replacement operation, the Markov chain is in a transmission pattern with $i^* - 1$ intervals, $i^* - 2$ intervals from x_1^* and 1 interval from x_2^*. The subsequent replacements can either (i) remove an interval from x_1^* to add an interval from x_2^*, (ii) remove an interval from x_2^* to add an interval from x_1^*, or (iii) leave the transmission pattern unchanged. In any case, these replacement operations do not change the number of intervals in the transmission pattern. We can describe the sequence of replacement operations using the random walk of Figure 6.10. The state space of the random walk are integers $0 \leq k \leq i^* - 1$ and the random walk is in state k if the current transmission pattern shares k intervals with pattern x_1^* (and thus $i^* - 1 - k$ intervals with x_2^*). The transitions in the random walk correspond to replacements. The random walk has two absorbing states, state 0 and $i^* - 1$.

To compute $q(L, 2)$ we need to compute the probability that starting from state $i^* - 2$ we are absorbed by state 0 (and thus, since $\lambda \to \infty$, by x_2^*). Denote by a_i the probability to be absorbed by state 0 starting from state $i^* - 1 - i$. We obtain

(after simplification) the following system of equations,

$$\begin{aligned} a_0 &= 0 \\ a_1 &= \frac{2}{5}a_0 + \frac{3}{5}a_2 \\ a_i &= \frac{1}{2}a_{i-1} + \frac{1}{2}a_{i+1} \quad \text{(for } 2 \leq i \leq i^* - 3\text{)} \\ a_{i^*-2} &= \frac{3}{5}a_{i^*-3} + \frac{2}{5}a_{i^*-1} \\ a_{i^*-1} &= 1 \end{aligned}$$

Solving this system gives

$$a_i = \frac{1}{L} + \frac{2}{L}i$$

and as a result we have

$$q(L,2) = a_1 = \frac{3}{L}.$$

Corollary 6.2 *In the case $l = 2$, the expected time needed to switch between the two patterns of maximal spatial reuse grows as*

$$2\frac{\rho}{\mu} + O(1).$$

Proof: We apply the Theorem 6.1 with $r(L,2) = L/6$ and $q(L,2) = 3/L$. □

For other values of l, the systems of equations to solve to obtain $q(L,l)$ are more complicated and it is hard to obtain a closed-form for $q(L,l)$. Moreover, in general the value of $q(L,l) \cdot r(L,l)$ depends on L, which implies that the slope of the linear increase of the expected switching time with ρ depends on the size of the topology. This is apparent on Figure 6.11 that shows the value of $q(L,l)$ together with the value of the slope for $l = 3$.

6.4 Related Work

There are very few models trying to predict the short-term fairness of 802.11 in multi-hop networks. [GSK05, LNG06] present a detailed analysis for the hidden node scenario (where two transmitters hidden from each others compete to send packets to the same receiver). More precisely, they compute the expected time needed to switch from a state where one of the transmitters is active to a state where the other transmitter is active. To the best of our knowledge there is no work extending these results to more general topologies.

We also want to acknowledge that the methodology of Section 6.3 was suggested to us by David Aldous and that his book [AF] was a great source of inspiration to us.

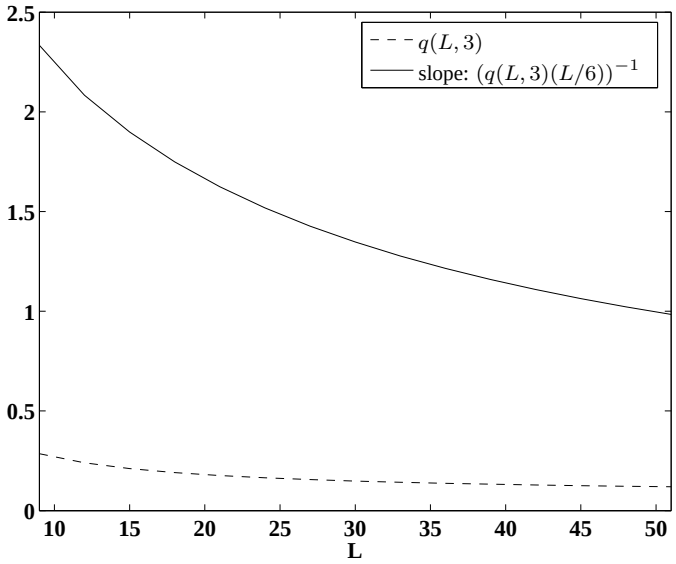


Figure 6.11: For ρ large enough, the expecting switching time between patterns of maximal spatial reuse increases linearly with ρ. For $l = 3$, the slope of this linear increase is equal to $\left(q(L,3) \cdot \frac{L}{6}\right)^{-1}$ where $q(L,3)$ is an absorption probability on a triangular random walk and can be computed by solving a system of equations similar (but more complicated) than the one presented for $l = 2$.

6.5 Conclusion

In this chapter we have proposed several methodologies to study the short-term fairness of the idealized 802.11 in one-dimensional topologies. In addition, we have presented simulation results for regular two-dimensional topologies.

Short-term fairness is an issue at high access intensities where the protocol has a tendency to give repeated access to the same links. Indeed, our study shows that in topologies where 802.11 is long-term fair, operating at high levels of spatial reuse reduces its short-term fairness. One thus needs to carefully select the access intensity of the protocol in order to satisfy the spatial-reuse and the short-term fairness requirements of the application running on top of it.

The work presented in this chapter gives a first idea of the 802.11 short-term fairness performance. However, it is our opinion that a lot of work remains to be done to provide a more complete analysis, especially in two-dimensional topologies.

Chapter 7

Asymmetric Exclusion Domains

In practice, the exclusion domain of a link is not necessarily symmetric around the link. In this chapter, we show to which extent our results can be adapted to deal with *asymmetric exclusion domains* (Figure 7.1). We concentrate on the line topology. Other topologies will be discussed in the next chapters.

The exclusion domain around a link is asymmetric if there are nodes inside the emitter's carrier sensing range (CSRange) that are outside its receiving range (see Node 1 in Figure 7.1). These nodes are unable to decode the RTS packets from the emitter and are silenced only through the physical carrier sensing. Therefore, more nodes are silenced on the emitter's side than on the receiver's side, which is why we call the exclusion domain asymmetric.

Now assume that a node in the receiver's CSRange (for example Node 6 in Figure 7.1) begins transmitting. As the two nodes (Node 4 and Node 6) are within CSRange of each other, the new transmission (from Node 6 to 7) interferes with the ongoing transmission (from Node 3 to 4). In this section, we assume that the signal-to-interference ratio is still high enough for the receiver to be able to decode its packet, despite the new interferer, i.e., there is *capture effect*.

In the case where the strongest signal begins after the weakest one (as in the example of Figure 7.2), we consider two capture models, *full capture* and *limited capture*.

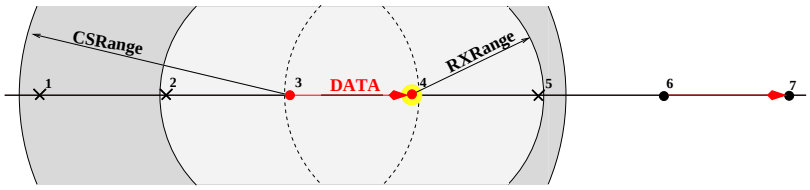


Figure 7.1: The exclusion domain of the directed link between Node 3 and Node 4 is asymmetric. When Node 6 starts transmitting, there is capture at Node 4 and Node 4 can continue to decode the DATA packet from Node 3.

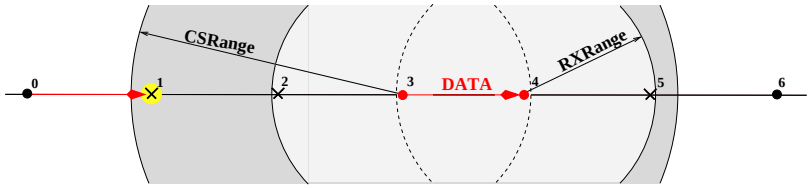


Figure 7.2: Assume there is an ongoing DATA transmission between Node 3 and Node 4. As Node 1 is in the CSRange of Node 3, it can detect this transmission. If Node 0 sends a RTS to Node 1 we have two possibilities. Either Node 1 can resynchronize on this stronger signal; in which case the RTS will be followed by a complete exchange between Node 0 and Node 1; or it cannot resynchronize on the strongest signal and the RTS will be lost. We call the first possibility *full capture* and the second *limited capture*.

7.1 Full Capture

In the full capture model, we assume that the capture effect also occurs when the strongest signal begins after the weakest one. This is a rather optimistic scenario, although it is supported to some extent by experimental evidence. [KVSA04] reports that capture occurs even if the packet with the strongest signal arrives later, but within the physical preamble of the first packet. In essence, *full capture* means that the order of arrival of the transmissions does not matter and it implies that transmission patterns are valid independently of the order of arrival of the transmissions.

As an example, consider a line network topology where the RXRange covers one neighbor, but where the CSRange covers two neighbors. A transmission pattern is valid if two consecutive receivers are separated by at least 2 space units, and if two consecutive transmitters are separated by at least 3 space units (see Figure 7.4). Under this setting, the maximal spatial reuse is still $1/3$, but it is not possible to have two transmitters back to back in a transmission pattern of maximal spatial reuse.

7.1.1 Markov Modeling

Asymmetric exclusion domains and capture effects make the analytical study of the 802.11 protocol more challenging. We now show that the Markovian framework of Chapter 3 can be extended to include this case. The structure of the Markov chain remains essentially unchanged but some states (and the corresponding transitions) disappear. Take, for example, the Markov chain of Figure 3.1 and assume that the CSRange now covers two neighbors instead of one. All transmission patterns remain valid except the Pattern 2.3 at the top level of the Markov chain, where two transmitters are back to back (the resulting Markov chain represented in Figure 7.3). This change in the set of valid transmission patterns $\mathcal{V}$ does not affect

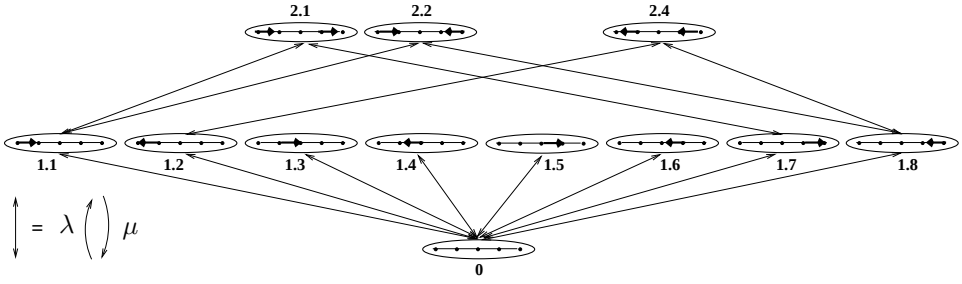


Figure 7.3: The reversible Markov chain used to model the 802.11 protocol under asymmetric exclusion domains and full capture.

the reversibility of the Markov chain. Equation (3.1) still holds, and the organizing properties of the protocol are therefore essentially unchanged. In particular, the long-term fairness behavior of the protocol on infinite topologies is exactly the same. However, the number $N(i)$ of valid transmission patterns with i active links and the number $N(i,j)$ of such patterns including link j have to be recomputed to obtain the exact curves for the spatial reuse and the fairness index of the protocol as a function of ρ. Again, this is only tractable for simple topologies. We compute these numbers explicitly for the line topology in the following paragraph.

We begin with the observation that two consecutive active links along the line can be separated by the least number of space units if and only if their orientation is such that the two senders are not back to back (see Figure 7.4). Again, to each active link we associate an interval of length l on the line. Consider a transmission pattern of maximal spatial reuse and assume that, once we have placed the intervals corresponding to its active links, there is no hole [1]; because of the above observation, all active links must point in the same direction until a certain point, and then point to the other direction (see Figure 7.5). If i is the number of active links, there are clearly $i+1$ possible settings. Now, let us assume that there is one hole. This hole will allow for an additional change of direction, as depicted in Figure 7.6. Therefore, in this case, there is a total of three points where links change direction and there are

$$\binom{i+3}{3} = \binom{i+3}{i}$$

ways to pick these three points[2]. This reasoning can be extended to an arbitrary

[1] Note that this is only possible for certain values of line size L.

[2] Note that these points can be collocated. Two collocated points correspond to a hole without a change of direction. Three collocated points correspond to a hole and two receivers facing each other. The problem is equivalent to counting the number of 3-combinations with repetition of an $(i+1)$-element set.

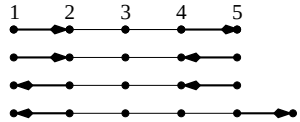


Figure 7.4: Among the four possible configurations of two active links placed as closely to each other as possible (with $l = 3$), only the three first ones maximize spatial reuse.

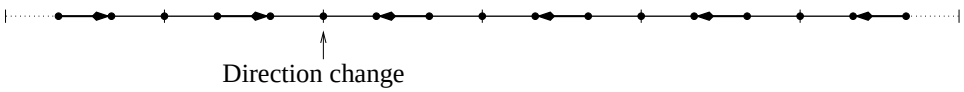


Figure 7.5: Transmission pattern of maximal spatial reuse (with $l = 3$). Active links are sorted in two groups, inside which all links point in the same direction.

number of holes v, and we find

$$N(i) = \binom{i + 2v + 1}{i}$$

where $v = L + (l - 1) - li$.

To obtain an expression for $N(i, j)$ we follow the same procedure as in Section 5.1. If link j_{right} from node $j-1$ to node j is active, we enumerate the possible ways to place k active links on its left side and $(i - 1) - k$ active links on its right side. On the right side, the problem consists simply in packing $(i-1)-k$ intervals of length l in an interval of length $L - j$. On the left side, the problem is slightly trickier, as node $j - l$ cannot emit. This means that in our previous reasoning we have one less change of direction. We obtain

$$\binom{k + 2v_{\text{left}}}{k}$$

possible settings only, where $v_{\text{left}} = j - 1 - lk$. Combining the left and right parts, we get

$$N(i, j_{\text{right}}) = \sum_{k=0}^{i-1} \binom{k + 2v_{\text{left}}}{k} \binom{(i - 1) - k + 2v_{\text{right}} + 1}{(i - 1) - k}$$

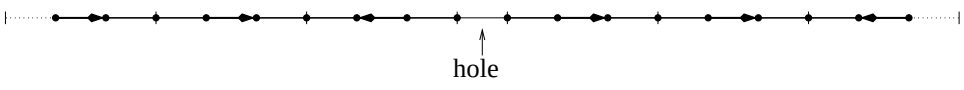


Figure 7.6: Transmission pattern with one hole (with $l = 3$). There are three points where links change direction.

where $v_{\text{right}} = L - j - l((i-1)-k) = L + l - j - l(i-k)$. The same reasoning allows to compute the value of $N(i, j_{\text{left}})$:

$$N(i, j_{\text{left}}) = \sum_{k=0}^{i-1} \binom{k + 2v_{\text{left}} + 1}{k} \binom{(i-1) - k + 2v_{\text{right}}}{(i-1) - k}.$$

Note that in the asymmetric case with full capture links j_{left} and j_{right} do not necessarily have the same probability to be active.

We can then plug the expressions for $N(i)$ and $N(i, j)$ in Equations (4.1), (5.1), and (5.2) to obtain the exact curve for the spatial reuse and the fairness index of the protocol.

The asymptotic spatial reuse for large networks is obtained in a similar way as in the case of symmetric exclusion domains.

Theorem 7.1 *For $L \to \infty$, the average spatial reuse σ of the protocol under asymmetric exclusion domains and full capture converges to*

$$\lim_{L \to \infty} \sigma = \frac{2\rho y_1^{2l-1}}{1 + 2l\rho y_1^{2l-1}}$$

where y_1 is the positive real root of $1 - y - \rho y^{2l}$.

Proof: The result is obtained by applying Lemma A.1 (in the appendix) with $k = \lfloor \frac{L+(l-1)}{l} \rfloor$, $m = 2l$, $r = \rho$, and $n = 2((L + (l-1)) \bmod l) + 1$. We then divide by l to obtain the average spatial reuse σ. □

Again, we verify that σ tends indeed to $1/l$ when ρ tends to infinity.

7.1.2 Simulation Results

We illustrate by simulation the influence of asymmetric exclusion domains on the performance of the 802.11 protocol. Figure 7.7 shows that its spatial reuse increases less rapidly than in the symmetric case. At a given access intensity ρ, the protocol with asymmetric exclusion domains has therefore on average fewer active links than in the symmetric case. This has a positive effect on the fairness of the protocol, especially at high values of ρ, where operating with a slightly lower number of active links greatly improves the fairness. A comparison between Figure 7.9(a) and Figure 7.9(b) confirms that the starvation phenomenon is less pronounced than in the symmetric case and that the links in the middle of the line topology obtain a fairer access to the channel. Yet, we also see that the distribution of the link activities is not symmetric anymore. This is in agreement with our analysis that tells us that two links in opposite directions do not necessarily have the same probability of accessing the channel. This effect has a negative impact on the fairness index of the protocol. We have therefore two competing effects. Figure 7.8 shows that, at low values of ρ, the two effects compensate each other; the fairness index of the protocol is very similar under symmetric and asymmetric

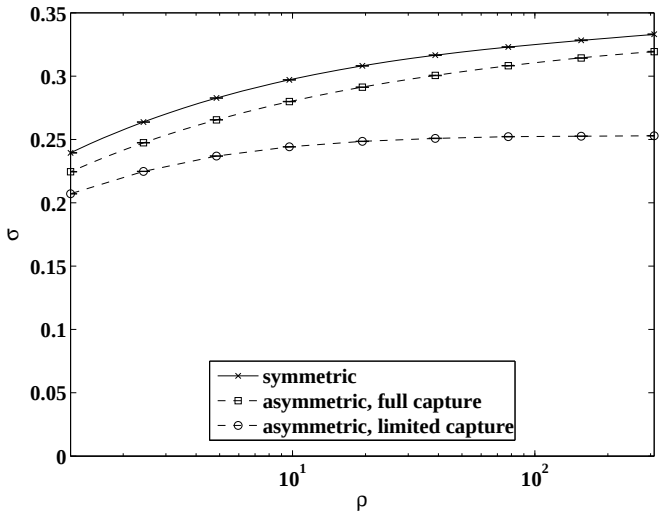


Figure 7.7: Average spatial reuse σ achieved by the idealized 802.11 protocol on a line topology of 50 nodes, as a function of the access intensity ρ. The markers correspond to the results obtained by simulations and the curves to the results obtained analytically (except for the 'asymmetric, limited capture' case where we do not have the exact analytical curve).

exclusion domains. However, at high values of ρ, the positive effect dominates; the fairness index of the protocol in the asymmetric case with full capture is above its fairness index in the symmetric case. In the limit, when $\rho \to \infty$, the behavior of the protocol under asymmetric exclusion domains with full capture is essentially the same as under symmetric exclusion domains. The spatial reuse tends to 0.3469 and approximately $2/3$ of the links get starved. The only difference is that, under asymmetric exclusion domains, some links belong to more patterns of maximal spatial reuse than others, depending on their direction: this reduces the FI from 0.3469 to 0.2676.

7.2 Limited Capture

In this section, we consider asymmetric exclusion domains with limited instead of full capture. Limited capture means that the capture effect occurs only if the strongest signal comes first. This assumption reflects the fact that in practice, many radio circuits cannot resynchronize to a stronger signal if they are already locked on another (weaker) carrier. Consequently, in this capture model, the order of arrival of transmissions does matter.

Consider again the example in Figure 7.4. The first (top) case is possible only if the left most link (from Node 1 to 2) becomes active first. Indeed, suppose instead that the transmission from Nodes 4 to 5 begins first. Node 2 senses this

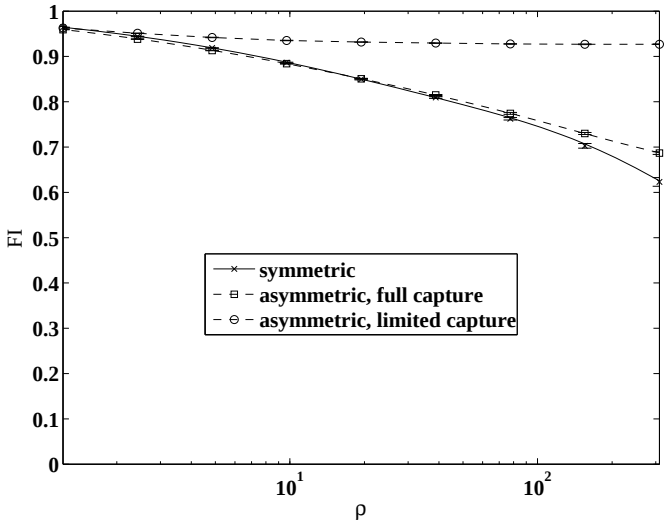


Figure 7.8: Fairness index (FI) of the idealized 802.11 protocol on a line topology of 50 nodes. The format of the figure is the same as for Figure 7.7.

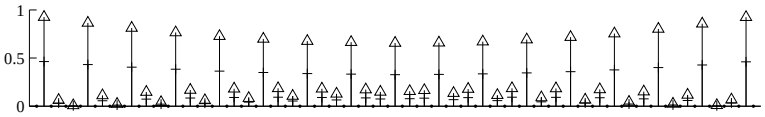
first transmission and locks its synchronization circuit on it. When Node 1 sends a RTS packet to initiate a transmission, Node 2 is unable to resynchronize and decode it. The RTS packet is lost and no transmission follows. This limited capture effect is implemented in the most well-known network simulators, including ns-2 [ns2] and qualnet [qns].

7.2.1 Simulation Results

These simulation results were at first quite surprising to us. Figures 7.7 and 7.8 show the performance of the idealized 802.11 protocol with asymmetric exclusion domains and limited capture. The spatial reuse of the protocol increases to a value much below $1/3$, but its fairness index remains above 0.9, even for large values of ρ. The access probabilities of the different links presented in Figure 7.9(c) illustrate even better the dramatic improvement of the protocol fairness at high access intensities. Except at the border, the situation is now completely fair. These results are good news in practice: They show that the starvation phenomenon is fixed by receivers having a limited capture capability, without requiring any other modifications to the protocol.

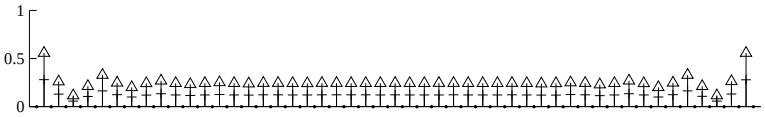
7.2.2 Markov Modeling

The set of the valid transmission patterns (and thus the state space of the Markov chain) is unchanged from the case with the full capture effect. However, some states can now be reached only from a restricted set of lower states. Figure 7.10



(a) Symmetric exclusion domains (RXRange $=$ CSRange $= 1$). $\sigma = 0.34$, FI $= 0.54$.

(b) Asymmetric exclusion domains (RXRange $= 1$, CSRange $= 2.2$) with full capture. $\sigma = 0.32$, FI $= 0.65$.



(c) Asymmetric exclusion domains (RXRange $= 1$, CSRange $= 2.2$) with limited capture. $\sigma = 0.25$, FI $= 0.93$.

Figure 7.9: Link access probabilities in a 50 node line network for $\rho = 620$. On the horizontal axis, the line network. Each of its edges corresponds to two directed links j_{right} and j_{left}. There is one vertical bar per network edge and each bar is divided in two parts, the bottom part measures $p(j_{\text{right}})$ and the top part $p(j_{\text{left}})$.

illustrates the Markov chain used to model the 802.11 protocol with limited capture effect. We see that, compared to the original Markov chain, two transition arrows have been removed: this accounts for the fact that the states at the top level can only be reached if the links become active in the right order.

Removing possible transitions between the states of the Markov chain breaks its regular structure, and the chain looses its reversibility. This makes the analytical study of the chain very difficult as its stationary distribution does not have a nice product form anymore. However, a careful observation of the structure of the chain allows for the explanation of the results observed by simulation.

Let us first look back at the reversible case (corresponding to symmetric exclusion domains and asymmetric exclusion domains with full capture). For large values of ρ, the chain spends most of its time in the states of maximal spatial reuse. Indeed, when a transition occurs from a state at the top level to a state at the lower level (i.e., when an active link becomes idle), as $\lambda \gg \mu$, the next transition is most likely the reactivation of the same (or possibly the reverse) link, bringing the chain back to a top level state. This explains why the protocol almost achieves maximal spatial reuse in the reversible case.

In the non-reversible case, this cannot happen anymore, because many of the

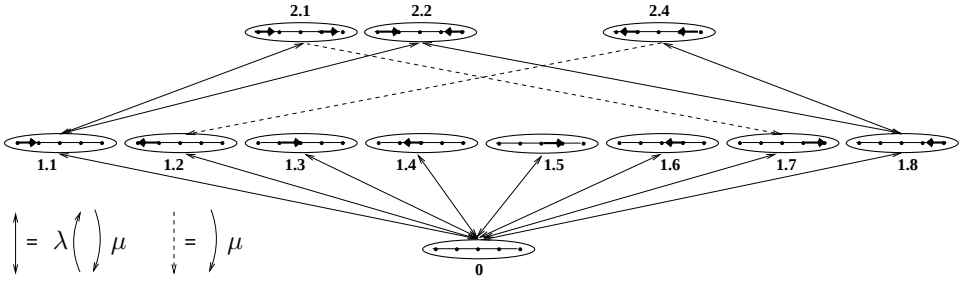


Figure 7.10: Non-reversible Markov chain used to model the 802.11 protocol under asymmetric exclusion domains and limited capture.

transitions from top level states to lower level states cannot be reverted (see Figure 7.10). Therefore, if the chain leaves a top level state, it might have to go down two or more levels before it can climb back to a top level state. The main consequence of this new dynamic is that the time spent in non-maximal spatial reuse states (states below the top level) becomes non-negligible.

In Proposition 7.1, we show that some states below the top level of the Markov chain have a non-trivial (i.e., non-zero) stationary probability when $\rho \to \infty$, contrary to the reversible case. As a consequence, the average spatial reuse *no longer* tends to the optimal value for increasing ρ.

Proposition 7.1 *There exist transmission patterns of non-maximal spatial reuse with a strictly positive stationary probability when* $\rho \to \infty$.

Proof: Consider a transmission pattern of maximal spatial reuse x^*, and another pattern x' obtained by removing an active link from x^* and such that there is no possible transition (in the Markov chain) from x' to x^* (which is only possible in the limited capture case). Assume furthermore that x' has i active links. Now, the balance equation for x' reads

$$\pi(x')i\mu = \pi(x^*)\mu + \sum_{y \in A_{x'}} \pi(y)\lambda,$$

where $A_{x'}$ denotes the set of states y at level $i-1$ such that a transition from y to x' is possible. Dividing both sides of this equation by $i\mu$, we obtain that $\pi(x') \geq \pi(x^*)/i$. Therefore, when ρ tends to infinity, if x^* has a non trivial asymptotic stationary probability, then so does x'. □

However, the loss in the spatial reuse is compensated by an increase in the fairness of the protocol. Indeed, contrary to the two previous cases, links that do not belong to a pattern of maximal spatial reuse are not necessarily starved.

Proposition 7.2 *There exist links that* do not *belong to the patterns of maximal spatial reuse but have a strictly positive probability of being active when* $\rho \to \infty$.

Proof: Take a state x' as above. Remove one additional active link (next to the one removed to obtain x' from x^*) to obtain state y. From state y it is now possible to reach a state z that contains a link that does not belong to any pattern of maximal spatial reuse. We now show that this state has a non-trivial probability. Using the balance equations, we have

$$\pi(y)(\mu(i-1)+\lambda|A_y|) > \mu\pi(z)+\mu\pi(x')$$

where $|A_y|$ is the number of states at level i connected to y. As

$$\pi(z)i\mu > \lambda\pi(y) > \frac{\lambda}{\mu(i-1)+\lambda|A_y|}(\mu\pi(z)+\mu\pi(x'))$$

we obtain

$$\pi(z)(\mu i(i-1)+\lambda i|A_y|-\lambda) > \lambda\pi(x')$$

which implies that for $\rho \to \infty$

$$\pi(z) > \frac{1}{i|A_y|-1}\pi(x').$$

Therefore, if x' has a non-trivial asymptotic stationary probability, so does z, and hence there are links that do not belong to the patterns of maximal spatial reuse but have a strictly positive probability of being active when $\rho \to \infty$. □

Proposition 7.2 implies that, on the line topology, the starvation effect can always be avoided. In addition, we can show that in the asymmetric case with limited capture links in opposite direction have the same probability of being active (this was not the case under full capture). Indeed, given any transmission pattern, a link can become active if and only if there is no transmitting node in the carrier sensing range of its two end nodes. Thus, if link j_{left} can become active, so can link j_{right}.

Note that Propositions 7.1 and 7.2 are not restricted to the line topology and can be applied to any network topology provided that the corresponding Markov chain includes states of type x^*, x', y, and z.

7.3 Related Work

This section reviews the different models for 802.11 highlighting the assumptions made in terms of carrier sensing and capture models.

Among the models presented previously [CGLA04], [GSK06], and [MT06] are the only ones to relax the assumption of equal receiving and carrier sensing ranges, but they still neglect the capture effect. These three papers present results in terms of throughput and fairness, but only the models of [GSK06] and [MT06] are able to predict the starvation phenomenon. The work of [MT06] does not investigate the effect of the carrier sensing range on the performance of the 802.11 protocol, rather, it concentrates on the role played by the minimum contention window in the starvation phenomenon. [GSK06] uses a 50 node two-dimensional topology on a

1000×1000m square area to validate its model. The receiving range is set to 200m and the carrier sensing range to either 200m or 400m. [GSK06] observes that the unfairness of the 802.11 protocol is higher at a large carrier sensing range. Although the model of [CGLA04] does not fully predict the starvation phenomenon, the authors observe it by simulation. Interestingly, they use the same experimental setting as [GSK06], but their simulations show that the unfairness of the 802.11 protocol is now stronger at a small carrier sensing range. These apparently contradicting results are probably a consequence of the limited amount of multi-hopping possible in the topology used in [GSK06] and [CGLA04] (indeed, the receiving and carrier sensing ranges are such that they cover a large part of the network).

7.4 Conclusion

In this chapter, we have extended our Markovian model to allow for different receive and carrier sensing ranges and to account for the capture effect. Our model makes the starvation phenomenon directly apparent in its equations and allows for a very intuitive explanation of 802.11 performance on a finite line topology.

On the positive side, we found that for a given (finite) access intensity ρ, the 802.11 protocol is fairer when the receiving and carrier sensing ranges are significantly different (asymmetric case) than when they have virtually the same values (symmetric case). On the negative side, we showed that the price to pay for this higher fairness is a lower spatial reuse.

In addition, we demonstrated that the capture capabilities of the protocol play a decisive role on its performance when $\rho \to \infty$. In the asymmetric case with full capture (as well as in the commonly adopted symmetric case), the spatial reuse of the 802.11 protocol is maximal but all the links that do not belong to the patterns of maximal spatial reuse are starved. In contrast, in the asymmetric case with limited capture, the spatial reuse of the protocol is not maximal but starvation can be avoided.

Chapter 8

Back to the 'Real' Protocol

In the previous chapters, we obtained a precise characterization of the performance of the idealized 802.11 protocol. The goals of this chapter are, to contrast the performance of the current IEEE 802.11 with the performance of the idealized 802.11 protocol, and to identify all the factors that prevent IEEE 802.11 from achieving the performance of the idealized 802.11 protocol.

8.1 The ns-2 Simulator

We consider the implementation of IEEE 802.11 under the widely used ns-2 simulator [ns2]. The main reason for choosing this specific implementation is twofold: (i) The code is available, contrary to the code contained in the firmware of IEEE 802.11 cards on the market and (ii) it follows closely the IEEE 802.11 standard[1].

The main limitation of the ns-2 simulator is at the physical layer: It does not support cumulative interference. In other words, it computes the signal-to-interference ratio using only one interfering signal at a time instead of the sum of all interfering signals. On a small line topology such as the one of Figure 7.1 this is not a serious limitation as only Node 5 can create a collision at Node 4; if Node 5 does not emit, even the combined interference of all other nodes is not enough to create a collision at Node 4. However, the absence of cumulative interference renders the physical layer of ns-2 not very realistic in the two-dimensional topologies.

We use the default ns-2 parameters, which mimic a 914MHz Lucent WaveLAN DSSS radio interface. The RXRange is 250m, the CSRange is 550m and the capture threshold is equal to 10, which means that a signal can capture the channel if it is 10 times stronger than the interfering signal. Hence, there is capture only if

[1]We noticed two differences between the IEEE 802.11 standard and the ns-2 implementation. The first difference is minor, in ns-2 the contention window is selected in $[0, cw-1]$, whereas according to the standard it should be selected in $[0, cw]$. The second difference concerns the exponential increase of the contention window. According to the standard, the contention window is doubled only before a retransmission and not when a node finds the medium busy on a transmission attempt. We modified the ns-2 implementation to meet the standard requirements before running our simulations.

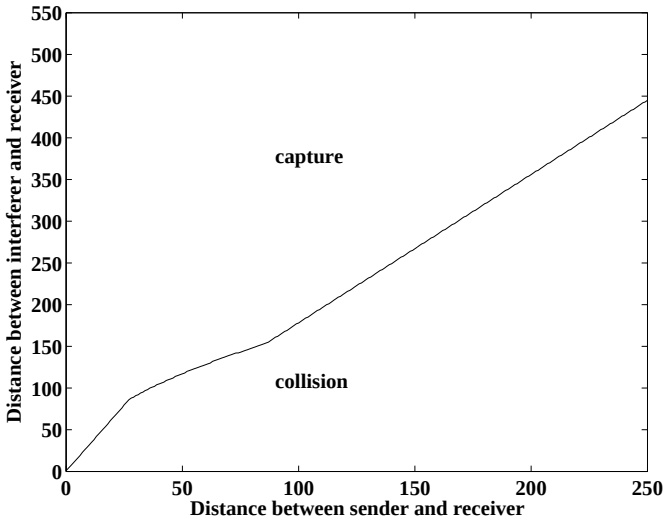
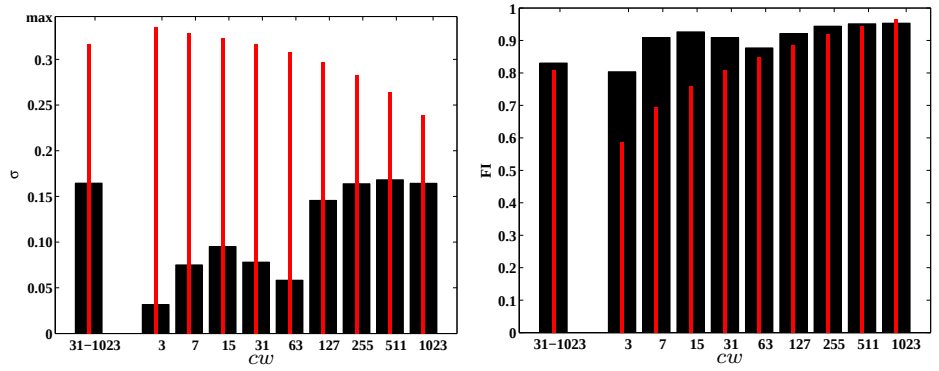


Figure 8.1: Capture versus collision under ns-2. Assume there is an ongoing transmission between a sender and a receiver at distance x (x-axis) and that an interfering node at distance y from the receiver starts transmitting. If the distance between the sender and the receiver is significantly smaller than the distance between the interferer and the receiver, there is a capture effect. Otherwise there is a collision.

the distance between the sender and the receiver is significantly smaller than the distance between the interferer and the receiver. Figure 8.1 illustrates what a capture threshold of 10 means in terms of these respective distances under the ns-2 propagation model[2]. ns-2 assumes the limited capture model (see Chapter 7 for a description of this model).

All simulations assume saturated traffic conditions. The DATA packet size is 1500 bytes (i.e., the average exchange time μ^{-1} is fixed to a value slightly larger than $1500 \cdot 8 \cdot 10^{-6}$). To provide accurate results, each simulation is repeated 50 times (using different random seeds). The simulations run for 50 seconds of real time (a warm-up period of 10 seconds is used). All figures follow the same format. We plot the average spatial reuse σ of the protocol as a function of its contention window cw (i.e., we vary the average backoff time $\lambda^{-1} = 0.5 \cdot cw \cdot 20 \cdot 10^{-6}$). We consider fixed contention window cases (i.e., $cw = CW_{min} = CW_{max}$) and a variable contention window case (i.e., $CW_{min} \leq cw \leq CW_{max}$) where the contention window is doubled after a collision.

[2]We use the TwoRayGround propagation model. In this model the power received at a distance d from the source is proportional to P/d^{α} where $P \simeq 0.28W$ is the transmitting power and where $\alpha = 2$ for $l \leq 86.14$, and $\alpha = 4$ for $l > 86.14$.



(a) Average spatial reuse

(b) Fairness index (long-term)

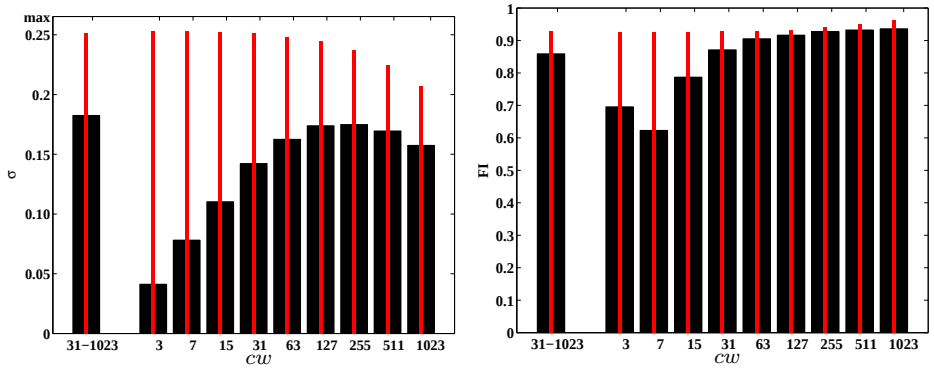
Figure 8.2: Symmetric exclusion domains, CSRange = RXRange = 250m (1 space unit). The theoretical values (idealized 802.11 protocol) are in gray (thin bars), the values for (ns-2) IEEE 802.11 are in black. The left-most bar corresponds to the variable contention window case (i.e., $CW_{min} = 31$ and $CW_{max} = 1023$), the other bars correspond to fixed contention window values (i.e., $cw = CW_{min} = CW_{max}$). The plots show the average spatial reuse (σ) and the fairness index (FI) of the protocol. The 95% confidence interval are not represented on the figure as their small (below ± 0.001) values make them barely visible.

8.2 Performance Overview

We apply our analysis of Chapters 4, 5 and 7 to compute the performance of the idealized 802.11 protocol on a line topology of 50 nodes[3]. We then simulate IEEE 802.11 using the ns-2 simulator on the same topology.

Figure 8.2 shows the performance of the idealized protocol with symmetric exclusion domains together with the performance of (ns-2) IEEE 802.11 with a CSRange of 250m. The average spatial reuse σ of the idealized protocol increases towards the maximal value of 0.3469 as the contention window cw becomes small. This is however not the case for the current IEEE 802.11 protocol. It achieves a spatial reuse of at most 0.17, half the optimal spatial reuse of 0.3469. Yet, we have seen that the high spatial reuse of the idealized protocol comes at the cost of a reduced fairness. On the contrary, the current IEEE 802.11 protocol is relatively fair even at low cw values. Its fairness index (FI) oscillates between 0.80 ($cw = 3$) and 0.95 ($cw = 1023$). Looking more closely at the trade-off between spatial-reuse and fairness, we observe that given a minimum fairness requirement the idealized protocol can guarantee a higher level of spatial reuse than IEEE 802.11. For example, at a FI requirement of 0.95, the idealized protocol can guarantee $\sigma = 0.24$,

[3]In the variable contention window case, the theoretical values are those obtained for $cw = CW_{min}$, as in the idealized 802.11 the absence of collisions means that the contention window is never doubled.



(a) Average spatial reuse

(b) Fairness index (long-term)

Figure 8.3: Asymmetric exclusion domains, CSRange $= 2.2\cdot$ RXRange $= 550$m (2.2 space units). The format is the same as in Figure 8.2.

whereas IEEE 802.11 only achieves $\sigma = 0.17$. The difference is even more striking at lower FI requirements.

Figure 8.3 shows that the spatial reuse of IEEE 802.11 is slightly higher at a CSRange of 550m than at a CSRange of 250m. This is not surprising because at a CSRange of 550m the transmitter can use its physical carrier sensing mechanism (in addition to its virtual carrier sensing mechanism) to determine the state (idle, receiving or sending) of the receiver. This setting thus offers a double protection to the DATA packets and virtually all the DATA packets sent are received correctly. However, the performance of IEEE 802.11 are below the values predicted by the theoretical model both in terms of spatial reuse and in terms of fairness.

8.3 IEEE 802.11: Does it put the correct state in the network?

The emphasis of this section is on identifying situations that happen under the normal operation of IEEE 802.11 and that are responsible for its poor performance. We describe precisely three such situations: the *gagged* node, the *jammed* node, and the *focused* node. To each situation we associate a remedy and implement it in [ns+]. The remedies proposed might not be unique and often consist of minimum effort solutions. However, they are helpful in quantifying the impact of each of these situations on the performance of IEEE 802.11. The remission of the gagged, jammed, and focused node situations brings the performance of IEEE 802.11 very close to the performance of the idealized 802.11.

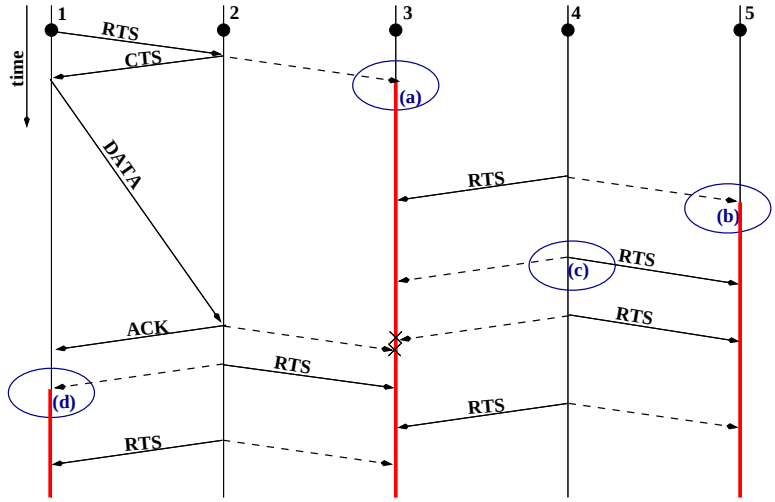


Figure 8.4: The gagged node situation: (a) The NAV of Node 3 is set until the end of the exchange between Nodes 1 and 2. (b) Node 4 sends a RTS to Node 3. Node 3 does not reply with a CTS because of its NAV, however Node 5 sets its NAV to protect a potential transmission between Nodes 4 and 3. (c) Node 4 is now the only node that can attempt to transmit and does so by sending repeated RTSs to Nodes 3 and 5. As a consequence at the end of the exchange between Nodes 1 and 2, Nodes 3 and 5 have been gagged by the repeated RTSs of Node 4. (d) In addition, the RTS from Node 2 to Node 3 silences Node 1. At this point no transmission is possible in the network and the repeated RTSs of Nodes 2 and 4 can maintain Nodes 1, 3, and 5 in a gagged situation for a long time.

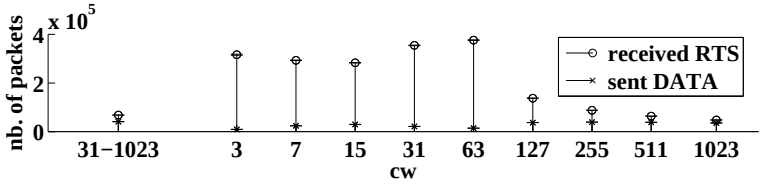
8.3.1 The Gagged Node Situation

Observation

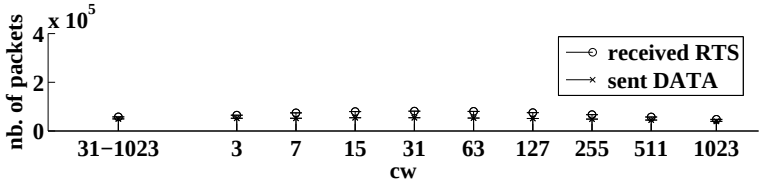
Figure 8.5(a) shows that despite the high number of RTSs received correctly, the number of DATA packets sent by IEEE 802.11 is low. Consequently, IEEE 802.11 is enable to reach a high level of spatial reuse. In addition, it is difficult to provide a logical explanation for the levels of spatial reuse observed at $cw = 31$ and $cw = 63$ under symmetric exclusion domains. Why should the levels of spatial reuse be lower at $cw = 31$ and $cw = 63$, than at $cw = 15$ and $cw = 127$ (Figure 8.2)?

Problem

A node is gagged (silenced) by repeated RTSs (intended to other nodes) that are not followed by a data packet transmission. The gagged node situation arises when a node sets its NAV to protect a DATA packet that is never sent. In this situation, the virtual carrier sensing that is based on the NAV value does not reflect the real state of the channel anymore. Figure 8.4 shows an example of a gagged node situation.



(a) Current IEEE 802.11



(b) IEEE 802.11 without gagged nodes

Figure 8.5: Number of RTS packets received successfully and number of DATA packets sent (symmetric exclusion domains, CSRange = RXRange = 250m).

The gagged node situation is especially likely when the number of RTSs received correctly is high, as nodes need to decode successfully the RTS messages to update their NAV value. Figure 8.5(a) shows that in our simulations with symmetric exclusion domains the number of RTS packets received correctly is the highest at $cw = 31$ and $cw = 63$. This explains why the performances of the protocol are especially bad for these contention window values.

Remedy

The IEEE 802.11 standard [iee99] (p.79) mentions that *'A station that used information from an RTS frame as the most recent basis to update its NAV setting is permitted to reset its NAV if no PHY-RXSTART.indication is detected from the PHY during a period'* slightly larger than the time needed to send a CTS. However, this recommendation is currently not implemented. One possible remedy is thus to implement the NAV reset as proposed by the IEEE 802.11 standard; another, simpler remedy requiring less modification in the code is to modify the NAV on RTS to cover only the CTS exchange. We refer to the first solution as the *reset NAV* solution and to the second as the *reduced NAV* solution. We implemented both solutions [ns+] and found that in terms of performance they are similar. Figures 8.6 and 8.7 show the performance of IEEE 802.11 using the reduced NAV on RTS solution for a CSRange of 250m and 550m respectively. Compared to the current implementation of IEEE 802.11, the number of RTSs received correctly drops but, due to the

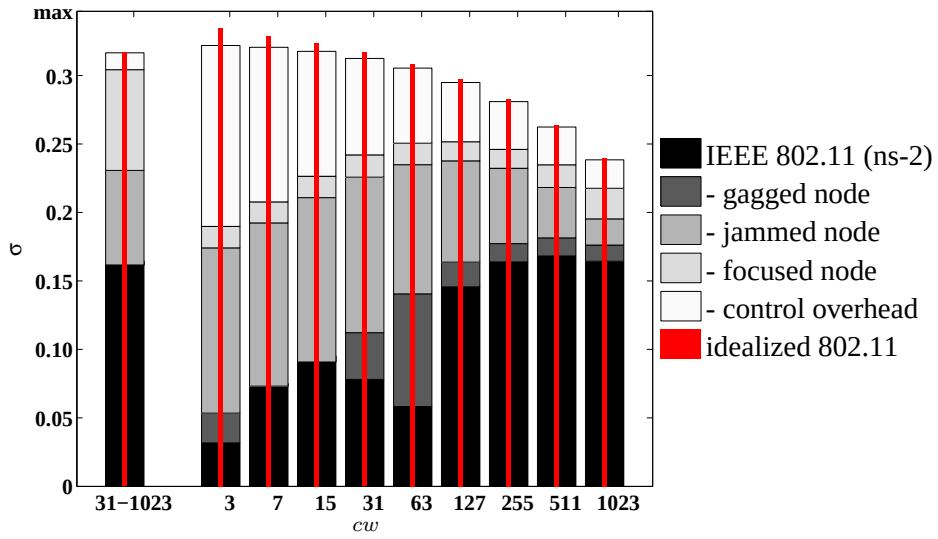


Figure 8.6: Spatial reuse of IEEE 802.11 for symmetric exclusion domains (CSRange = RXRange = 250m). Legend. The theoretical values (idealized 802.11 protocol) are represented by the thin dark gray bars. The values for the current (ns-2) IEEE 802.11 protocol are in black (thick bars). We first solve the gagged node situation (dark gray) by using a reduced NAV on RTS. We then solve the jammed node situation (gray) by using an additional control channel, and the focused node situation (light gray) by using a backoff per link. Finally, as a proof of concept we reduce the size of the overhead (white).

absence of gagged nodes, a successful RTS is now more likely to be followed by a successful CTS and a DATA packet (see Figure 8.5(b)). As a result, the abnormal behavior of the actual IEEE 802.11 protocol disappears completely. The spatial reuse achieved by the modified IEEE 802.11 protocol varies more smoothly and is higher than the spatial reuse of the actual IEEE 802.11 protocol with RTS-CTS. The increase in the level of spatial reuse is especially impressive at a CSRange of 550m.

The third columns of Tables 8.1 and 8.2 report the fairness index of the protocol with reduced NAV on RTS. At a CSRange of 250m, its fairness increases smoothly from FI $= 0.73$ at $cw = 3$ (and cw variable) to FI $= 0.94$ at $cw = 1023$. It is slightly lower than the fairness of the current IEEE 802.11 protocol especially in the variable contention window (where the FI drops by $\simeq 0.1$). At a CSRange of 550m, we observe a similar behavior except at $cw = 7$ where the FI of the protocol without gagged node is higher than the FI of the current protocol. This is not really surprising as the value of IEEE 802.11 fairness index at $cw = 7$ was somewhat illogical (see Figure 8.3).

cw	IEEE 802.11	gagged solved	jammed solved	focused solved	overhead reduced	idealized 802.11
31-1023	0.83	$0.73 \downarrow$	$0.81 \uparrow$	$0.45 \downarrow$	0.62	0.81
3	0.80	$0.73 \downarrow$	$0.75 \simeq$	$0.83 \uparrow$	0.69	0.59
7	0.91	$0.77 \downarrow$	$0.78 \simeq$	$0.82 \simeq$	0.73	0.69
15	0.93	$0.82 \downarrow$	$0.81 \simeq$	$0.82 \simeq$	0.76	0.76
31	0.91	$0.85 \downarrow$	$0.85 \simeq$	$0.82 \simeq$	0.81	0.81
63	0.88	$0.87 \simeq$	$0.88 \simeq$	$0.84 \simeq$	0.85	0.85
127	0.92	$0.89 \simeq$	$0.90 \simeq$	$0.87 \simeq$	0.89	0.89
255	0.94	$0.91 \simeq$	$0.91 \simeq$	$0.91 \simeq$	0.92	0.92
511	0.95	$0.92 \simeq$	$0.93 \simeq$	$0.94 \simeq$	0.94	0.94
1023	0.95	$0.94 \simeq$	$0.94 \simeq$	$0.96 \simeq$	0.96	0.96

Table 8.1: Fairness index of IEEE 802.11 for symmetric exclusion domains, CSRange $=$ RXRange $=$ 250m. We use the symbol $\simeq$ if the variation in the value of the fairness index between two consecutive columns is < 0.05.

cw	IEEE 802.11	gagged solved	focused solved	overhead reduced	idealized 802.11
31-1023	0.86	$0.70 \downarrow$	$0.84 \uparrow$	0.87	0.93
3	0.70	$0.67 \simeq$	$0.80 \uparrow$	0.88	0.93
7	0.62	$0.71 \uparrow$	$0.84 \uparrow$	0.92	0.93
15	0.79	$0.77 \simeq$	$0.87 \uparrow$	0.93	0.93
31	0.87	$0.81 \downarrow$	$0.89 \uparrow$	0.94	0.93
63	0.91	$0.84 \downarrow$	$0.91 \uparrow$	0.95	0.93
127	0.92	$0.86 \downarrow$	$0.92 \uparrow$	0.95	0.93
255	0.93	$0.88 \downarrow$	$0.94 \uparrow$	0.95	0.94
511	0.93	$0.89 \simeq$	$0.95 \uparrow$	0.96	0.95
1023	0.94	$0.92 \simeq$	$0.96 \simeq$	0.96	0.96

Table 8.2: Fairness index of IEEE 802.11 for asymmetric exclusion domains, CSRange $= 2.2\cdot$ RXRange $=$ 550m. We use the symbol $\simeq$ if the variation in the value of the fairness index between two consecutive columns is < 0.05.

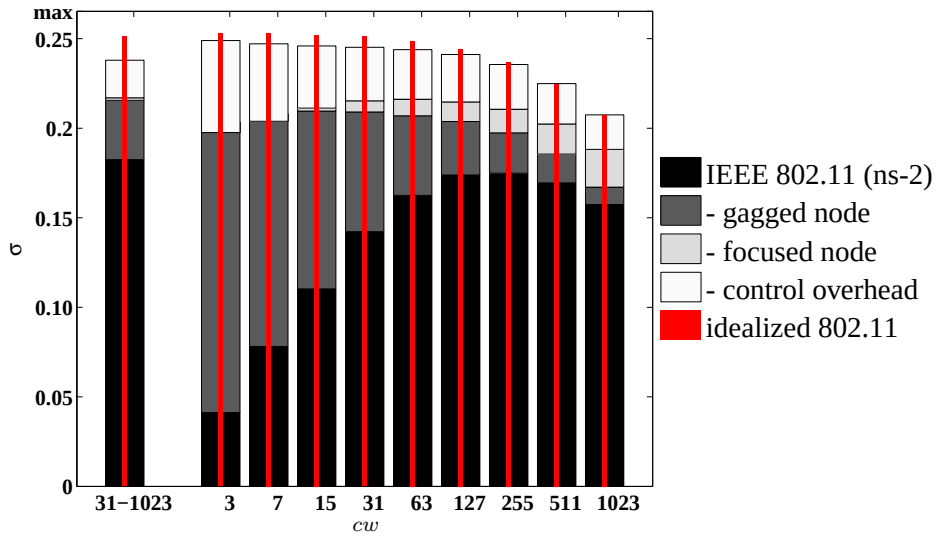


Figure 8.7: Spatial reuse of IEEE 802.11 for asymmetric exclusion domains (CSRange $= 2.2 \cdot$ RXRange $= 550$m). The format is the same as in Figure 8.6.

8.3.2 The Jammed Node Situation

Observation

At a CSRange of 250m, the number of concurrent transmissions is high, but the benefit is reduced due to collisions on DATA packets. Figure 8.9 shows that, at low contention window values, less than 50% of the DATA packets sent are received correctly. Hence, if we could limit the collisions on data packets, IEEE 802.11 would achieve a significantly higher spatial reuse.

Problem

A node is jammed by a data packet (initiated and terminated at two other nodes). As a consequence, it cannot extract information from control messages sent concurrently and cannot update its network allocation vector (NAV). This loss of information or state gives it a wrong view of the network. In particular, such a node cannot rely on its virtual carrier sensing mechanism to determine if the medium is busy, and it is likely to create a collision the next time it transmits. Figure 8.8 shows a typical instance of the jammed node situation. The jammed node situation occurs every time a node cannot decode the content of a CTS that is followed by a DATA packet. A node can be jammed by a data packet or by a control message. However, given that control messages are much shorter than data packets, the first occurrence is more common. Currently, the jammed node situation is responsible for most of the collisions on DATA packets.

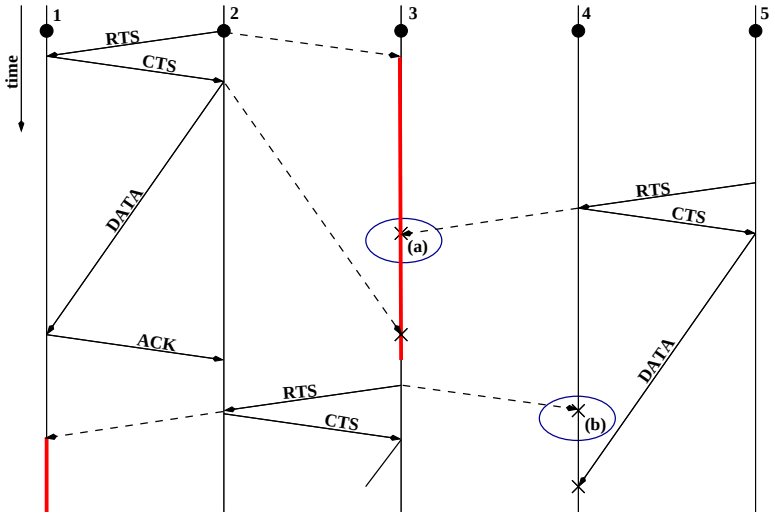


Figure 8.8: The jammed node situation: (a) Node 3 is jammed by the DATA packet sent from Node 2 to Node 1. As a result, it cannot decode the CTS packet sent from Node 4 to Node 5 and cannot update its NAV based on the latest information contained in this CTS. (b) At the end of the exchange between Nodes 1 and 2, Node 3 is not aware of any ongoing transmission, it sends a RTS that collides with the DATA packet sent from Node 5 to Node 4.

Remedy

It is difficult to suppress collisions between control messages as they are part of the contention resolution mechanism. Moreover, a node is more often jammed by DATA packets as they are typically much longer than control messages. A solution to the jammed node problem is thus to create a separate channel for control messages. In fact, IEEE 802.11 has 3 (IEEE 802.11b) to 12 (IEEE 802.11a) non-overlapping channels. Typically only one of these channels is used in a multi-hop ad hoc network. It is thus possible to add a control channel, basically 'for free'. Several works already report that the addition of a control channel increases the throughput of IEEE 802.11 (see for example the recent works of [SV04, RC05]). However, these works do not identify the jammed node situation. Again, we selected the implementation that requires the minimum change in the functioning of IEEE 802.11. Each node has only one transceiver and switches to the data channel only to transmit or receive DATA and ACK packets. The protocol is otherwise unchanged. In Figure 8.8, after the RTS-CTS handshake between Node 1 and Node 2, only these two nodes switch to the DATA channel. Consequently, Node 3, which is listening to the control channel, would be able to decode the CTS sent by Node 4, update its NAV, and the collision at Node 4 would be avoided. Figure 8.6 reports the average spatial reuse σ of IEEE 802.11 with a data and a control channel. At a CSRange of 250m, the spatial reuse of IEEE 802.11 greatly benefits from the

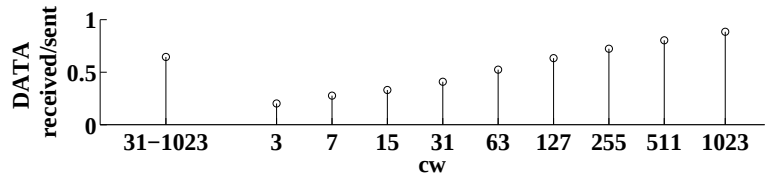


Figure 8.9: Fraction of DATA packets received correctly as a function of the cw. This results are for IEEE 802.11 with a CSRange of 250m and a reduced NAV (i.e., the gagged node problem has been solved)

addition of a control channel. Its fairness index is stable, except in the variable contention window case where it increases back to FI $= 0.81$. At a CSRange of 550m it is not necessary to add a control channel to overcome the jammed node situation as nodes can use their physical carrier sensing to recover the state of the channel.

8.3.3 The Focused Node Situation

Observation

The identification and the resolution of the gagged and jammed node situations are essential steps to obtain a more reliable, efficient IEEE 802.11. Nevertheless, the modified IEEE 802.11 protocol is still unable to achieve the highest levels of spatial reuse. At a CSRange of 250m, it is somewhat surprising that despite a relatively poor fairness (FI $= 0.81$ for $\sigma = 0.23$), IEEE 802.11 using a variable contention window is unable to achieve a higher spatial reuse than its fixed contention window counterpart (at $cw = 127$, FI $= 0.90$ for $\sigma = 0.24$). Similarly, at a CSRange of 550m the gain in spatial reuse achieved by using a variable contention window is negligible compared to the incurred loss in fairness.

Problem

In the current implementation of IEEE 802.11, each node has a backoff counter and a contention window value. The contention window is doubled each time the node needs to retransmit a packet and reset to its minimum value on a successful packet transmission. However, a node has typically several neighbors and can send traffic on each of the corresponding links. Consider a node with two links, one with a high level of collisions and one with a low level of collisions. Assume the node attempts to transmit on the link with the high level of collisions. After each packet collision, the node doubles its contention window, waits for its backoff counter to reach zero and retransmits. Consequently, the node attempts less and less frequently to access the channel even though it might be able to send successfully on its other outgoing link. The node implicitly assumes that all its outgoing links experience the same

contention as the link where it is currently trying to transmit. In other words, a backoff counter per node maintains the node focused on a single neighbor and it cannot take advantage of the link diversity.

Remedy

We propose to use a backoff counter per link instead of per node. Each link has its own contention window. The rules that regulate the decrease of the backoff counter are otherwise unchanged. A backoff counter per link introduces contention among the different links of the same node. In the previous situation, the retransmissions on the link with a high level of collisions would alternate with successful transmissions on the good quality link. Figures 8.6 and 8.7 show the performance of IEEE 802.11 with a per link backoff counter. At a CSRange of 550m, a per link backoff does not change much the spatial reuse of the protocol (except for a slight increase at large cw values), however it improves its fairness. At a CSRange of 250m, we observe a slight increase of its spatial reuse for all fixed cw (its fairness remains essentially unchanged). Yet, a per link backoff is most powerful in the variable contention window case. The modified IEEE 802.11 protocol is now less than 5% away from the spatial reuse of the idealized protocol. However, the fairness index of IEEE 802.11 drops to 0.45 (a value much below the theoretical 0.81). This surprisingly low fairness is due to IEEE 802.11 Binary Exponential Backoff (BEB), which doubles the contention window after a collision. Nevertheless, solving the focused node situation allows the modified IEEE 802.11 protocol to use the additional degree of freedom granted by a variable contention window to trade-off fairness for spatial reuse.

8.4 Closing the Gap

The gagged, jammed, and focused node situations demonstrate how important it is to maintain an accurate network state at each node and show how a wrong perception of the channel state can affect the performance of IEEE 802.11. In the variable contention window case, the suppression of these three situations brings the spatial reuse of IEEE 802.11 within 5% (CSRange = 250m) to 15% (CSRange = 550m) of the spatial reuse of the idealized protocol. As a proof of concept, we now reduce the size of the overhead messages (RTS, CTS, and ACK) and check that the performance of the protocol converges to the theoretical values.

Figures 8.6 and 8.7 show the performance of the modified IEEE 802.11 protocol when the overhead has been significantly reduced; we divided the time needed to send the control packets by 32.

At a CSRange of 250m, we observe that the spatial reuse of the protocol converges to the theoretical limit at all cw values. The small gap that remains at low cw values is due to the remaining overhead and to the discrete nature of the backoff distribution, which makes it impossible to completely avoid collisions. Table 8.1

shows that the fairness of the protocol also converges to the theoretical values. The few unavoidable collisions at low fixed cw slightly improve the fairness of the protocol compared to the theoretical prediction but it decreases its fairness in the variable contention window case. It is legitimate to ask if a sole reduction of the overhead in the current protocol would not bring equally good performances. This is however not the case, if we apply the same reduction of overhead to the current protocol with variable contention window $cw = 31 - 1023$, we obtain a spatial reuse 40% lower than with our modified protocol. At $cw = 31$, where the gagged node situation is a serious problem, reducing the overhead in the current protocol has a devastating effect as the spatial reuse drops essentially to zero. In general, it is easy to see that the gagged, jammed, and focused node situation appear independently of the time needed to send the RTS-CTS packets, it is thus imperative to solve these situations if IEEE 802.11 is to benefit from a reduction of its overhead.

Similarly, at a CSRange of 550m, the modified IEEE 802.11 protocol performs very close to the idealized protocol at all fixed cw. However, notice that in the variable contention window case, the spatial reuse of IEEE 802.11 is slightly below the spatial reuse of the idealized protocol. This is due to the situation depicted in Figure 7.2. When Node 0 sends a RTS to Node 1, Node 1 cannot reply with a CTS because we are in the limited capture case. This triggers a doubling of the contention window at Node 0 which is the cause of the reduced spatial reuse. This behavior can be avoided by using a fixed contention window.

8.5 Related Work

There is a vast body of research related to IEEE 802.11. We concentrate in the sequel on those results that are most closely related to our own work. To simplify the exposition we group them into several categories.

The first category identifies situations where its contention resolution mechanism is vulnerable or unfair. The hidden node situation [TK75] is probably the most (in)famous and the most studied of these situations. A node is hidden (from a sender) if it cannot detect its transmission but can interfere with the reception at the intended receiver. The hidden node situation is partially, but not completely, addressed by the use of the RTS-CTS handshake before the DATA packet transmission [XGB02]. The exposed node situation is often mentioned as the counterpart of the hidden node situation. A node is 'exposed' if it can transmit a packet without interfering with the current transmissions but is forbidden to do so by the protocol. In Figure 7.1, Nodes 1 and 2 are 'exposed' to Node 3 transmission, as they could transmit a packet without interfering with the reception at Node 4. However, if Node 2 transmits, it might interfere with the reception of an ACK at Node 3. The only 'real' exposed node in the case of a bi-directional exchange between Nodes 3 and 4 is thus Node 1. The exposed node problem can be reduced by choosing a smaller carrier sensing range. The jammed and the gagged node situations were only recently studied in the work of Ray et al. [RCS03, RCS05] and deserve more

attention. In this chapter, we proposed an alternative view of these two situations, which complements and extends his work. In particular, we used a ns-2 implementation of IEEE 802.11 (instead of a Matlab implementation) which allows us to (i) give more complete simulation results, (ii) investigate the effect of additional parameters such as the carrier sensing range and the contention window, and (iii) implement remedies to the jammed and gagged node situations.

The second category modifies IEEE 802.11 to improve its throughput or its fairness. Multi-Channel proposals [ND00, JDN01, SV04, RC05] fall in this category. They differ in the number of transceivers needed and the number of channels supported, but they almost all target an increase in IEEE 802.11 throughput. These proposals are in general much more complex (but also maybe more efficient) than the one presented in this chapter, whose only purpose was to solve a specific problem, namely the jammed node situation and as a result reduce the number of collisions on data packets. In addition, several papers study the fairness of IEEE 802.11. We mention a few that relate directly to the focused node situation. Indeed, the use of a backoff counter per link has first been proposed in the context of the MACAW protocol to provide per link fairness in wireless LAN (single-hop networks). More recently, [SHS04] proposed a per (multi-hop) flow scaling of the contention window to achieve a better flow fairness. The disadvantage of this proposal is that it requires the destination of the multi-hop flow to propagate appropriate state information to all intermediate nodes. Finally, the IEEE 802.11e draft standard relies on a backoff counter per traffic class to provide quality of service in the IEEE 802.11 framework. In this chapter, we show that a per-link backoff solves a specific problem, namely the focused node situation and results in a gain in the spatial reuse of the protocol. However, it is not clear that a per-link backoff solves the fairness problem of IEEE 802.11 in the multi-hop setting.

The last category investigates the effect of different parameters of IEEE 802.11 on its performance. Several works [YYS03, YS04, DLV04] study the effect of the carrier sensing range on its spatial reuse and argue that a suboptimal choice of the CSRange impacts the spatial reuse of the protocol. The papers by Calì et al. [CCG00b, CCG00a] also analyzes the effect of the contention window size, and again conclude that if the contention window size is not appropriately chosen the performance of IEEE 802.11 can be severely affected.

8.6 Conclusion

This chapter proposed a global view of IEEE 802.11. We used a systematic approach where we start from the actual implementation of the protocol, highlight an unwanted behavior of the protocol, describe its origin, and propose a solution. The remedies proposed in this chapter are not unique, but they are essential to showing that the resulting protocol can achieve the maximal level of spatial reuse.

The resolution of the gagged, jammed, and focused node situations increases significantly the spatial reuse of IEEE 802.11. To further increase its spatial reuse

and reach the limit imposed by its physical layer, one has to reduce the overhead to some negligible quantity.

All the results presented in this chapter were obtained by simulation. Although the ns-2 implementation of the IEEE 802.11 MAC protocol is very accurate. its physical layer model (wireless channel, interference model, etc.) is quite simplistic. The ultimate goal of this work is thus to be able to reproduce the experiments presented in this chapter on a testbed and to infer the impact of the gagged, jammed, and focused node situations on a real network.

Chapter 9

Closing Remarks

Let us now step back and discuss the strengths and weaknesses of the results presented in this thesis. Also, we will present what we consider to be the most interesting possible extensions. We conclude with a summary of our main contributions.

9.1 Discussion of the Results

In this work we have modeled the MAC layer of 802.11. Indeed, the states of the Markov chain correspond to the transmission patterns allowed by the MAC protocol. Nevertheless, the MAC layer relies on feedback from the physical layer (which is provided through well-defined service primitives) to decide whether a transmission should take place or not. In particular, physical carrier sensing is done at the physical layer and its implementation might vary from one wireless card to another. In our work, we have assumed a rather simplistic physical layer, the same as the one in the ns-2 simulator, and we have derived the performance achievable by 802.11 under this specific physical layer. A different physical layer is likely to change the set of possible transmission patterns and thus the structure of the Markov chain.

In the ns-2 physical layer, a node senses the channel as busy if there is a transmitting node in its carrier sensing range. In other words, its perception of the channel depends only on the activity of nodes in a restricted neighborhood around itself. In practice, this assumption is not necessarily fulfilled[1]. But it significantly simplifies our analysis. It remains to be seen to what extent this assumption is necessary to arrive at an analytically tractable model.

In addition, despite the use of physical and virtual carrier sensing mechanisms, the MAC protocol is often not *ideal*, i.e., it does not guarantee collision-free trans-

[1] According to the IEEE 802.11 standard, physical carrier sensing can be implemented via an energy detection threshold, via the reception of a valid (Physical Layer Convergence Procedure) header, or both. In the first case, all network nodes can contribute to the cumulative level of energy observed (although nodes very far away contribute little), however, in the second case, only nodes close enough can trigger a valid reception.

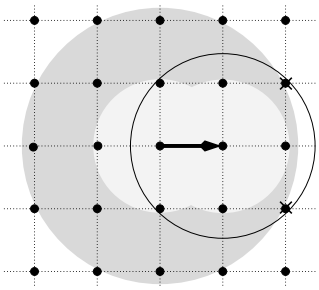


Figure 9.1: Under the ns-2 model (with the default parameters, RXRange = 250m, CSRange = 550m), all nodes in the black circle create a collision with the ongoing transmission if they start transmitting. Two of these nodes are allowed to do so by the protocol. The zone inside the black circle but outside the exclusion domain of the link (gray circles) can thus be seen as a vulnerability zone.

mission patterns. Note that whether or not there is a collision in a given transmission pattern also depends on the physical layer used by the nodes.

Consider the line topology studied in the two previous chapters. In this topology, and according to the ns-2 physical layer, the set of transmission patterns allowed by the protocol consists of all collision-free transmission patterns. This remains true for any other topology with symmetric exclusion domains. In contrast, in the grid topology with asymmetric exclusion domains, it is not the case. Figure 9.1 shows that in this topology the protocol allows an interferer to start transmitting even if this new transmission compromises an existing transmission. Fortunately, in the grid topology (or in other carefully designed topologies) one can often solve this problem by adjusting the physical layer parameters of the protocol. In irregular, random topologies, however, this problem is more difficult to solve.

When the protocol cannot successfully avoid all collisions, our model still predicts (provided that the contention window remains fixed) whether a link accesses the channel or not, but it does not predict whether the transmission on the link will be successful or not.

9.2 Possible Extensions

Following the previous discussion, it would be interesting to extend our model to consider other (more realistic) physical layers. In addition, in topologies where the protocol cannot always guarantee collision-free transmission patterns, one could try to make the distinction between collisions and successful transmissions.

Nevertheless, before extending our model in any way, we strongly believe that an experimental validation of our results is necessary. In particular, it would be interesting to compare the predictions of ns-2 in Chapter 8 with the performance of IEEE 802.11 obtained on a real testbed. This is ongoing work.

We have seen in our study that the value of the access intensity of the protocol determines the trade-off between its spatial-reuse and its fairness. It would be nice to combine these two quantities into a single metric that could be optimized to define an operating point for the protocol.

Another possible extension is to study how the properties of 802.11 affect higher layers. As a first step, one could for example consider an end-to-end flow traversing the line topology and investigate how the organization properties of 802.11 affect the propagation of the packets through the network. The work of [Dou07] goes in that direction.

Finally, on the mathematical side, it would be pleasing to find a simpler proof of the phase transition phenomenon described in Chapter 5. We are currently looking more closely at some of the percolation results presented in [vdBM94, GHM98] to see if they can be extended to fit our setting.

9.3 Conclusion

Throughout this thesis we have used a class of Markov chains to study the performance and properties of an idealized version of IEEE 802.11. In particular, we have shown how to relate the properties of the Markov chain to the properties of the protocol. Most of our analysis considers cases where the protocol can be modeled by a reversible Markov chain with a stationary distribution that has a product form. Networks where nodes have similar sensing and receiving ranges, or where nodes have full capture capabilities fall in this framework.

In this thesis we have considered two asymptotic regimes that reflect well the main characteristics of 802.11.

In the first regime, we let the access intensity ρ of the protocol go to infinity, but we keep the size of the network L finite. In this regime, the spatial reuse of the protocol is maximal. The 802.11 protocol optimally organizes the transmissions in the network. However, all links that do not belong to the transmission patterns of maximal spatial reuse are starved.

In the second regime, we let the size of the network L go to infinity, but we keep the access intensity ρ of the protocol finite. In this regime, we have obtained closed-form formulæ for the spatial reuse of the protocol on a regular one-dimensional topology. We also have shown that in this topology the protocol is long-term fair for all finite values of the access intensity. We then characterized its short-term fairness using different metrics. In contrast, we demonstrated that in regular two-dimensional topologies, the long-term fairness of the protocol exhibits a phase transition. The protocol is long-term fair for values of the access intensity below a threshold but severely unfair for values of the access intensity above it. Just below the threshold the protocol is still long-term fair, however, its short-term fairness is extremely poor. The 802.11 fairness properties thus highly depend on the underlying network topology.

In practice, the size of the network topology and the access intensity of the

protocol are both finite, and the trade-off between the spatial-reuse and the fairness of the protocol depends on their relative values. For a given topology size, the level of of organization achieved by the protocol increases with its access intensity. In general, the higher the level of organization, the higher the spatial reuse of the protocol, but the lower its fairness.

The applications running on a network typically impose their requirements on the MAC layer. Moreover, the network topology is often difficult to change. Our study shows that in such cases, one can tune the performance of 802.11 to meet the need of the applications using its access intensity. This can be done through the size of the DATA packets or by modifying the average backoff duration.

In this thesis, we have also considered cases where the Markov chain modeling the dynamics of the idealized 802.11 protocol is irreversible. For example, we have studied networks where the nodes have limited capture capabilities. In such networks, the trade-off between the spatial-reuse and the fairness of the protocol is less interesting, but starvation can always be avoided.

Finally, we have validated our results using the ns-2 simulator. During this validation process we have identified three situations that occur under normal operations of IEEE 802.11 and that prevent the protocol from operating at the limit set by its physical layer. For each situation we have proposed a remedy and we have shown that a ‘cured’ IEEE 802.11 can achieve the level of performance of the idealized 802.11 protocol.

Appendix A

Appendix

The following lemma is used in the proofs of Theorems 4.1 and 7.1.

Lemma A.1 *Let $\{X(k), k \geq 1\}$, be an infinite sequence of discrete random variables defined by their distribution:*

$$p_k(i) \doteq \mathbb{P}(X(k) = i) = \begin{cases} Z^{-1} r^i \binom{km-(m-1)i+n}{i} & 0 \leq i \leq k \\ 0 & \textit{otherwise} \end{cases}$$

where m and n are fixed finite integer parameters with $m \geq 2$ and $n \geq 0$, r is a positive real number, and where

$$Z = \sum_{i=0}^{k} r^i \binom{km-(m-1)i+n}{i}$$

is a normalizing constant. Then

$$\lim_{k\to\infty} \frac{\mathbb{E}(X(k))}{k} = \frac{mry_1{}^{m-1}}{1 + mry_1{}^{m-1}} \tag{A.1}$$

where y_1 is the positive real root of $1 - y - ry^m$.

Theorems 4.1 and 7.1 are two particular cases of this lemma, where the random variables $X(k)$ are the levels of the Markov chain (with $k = \lfloor (L + l - 1)/l \rfloor$), and whose invariant distribution π_i coincides with the distribution $p_k(i)$ in the lemma, for particular values of m, r, and n that differ for both theorems. As a result, the asymptotic average spatial reuse for large L (and thus k) is given by (A.1), for the appropriate values of m, r, and n.

A.1 Outline of the Proof of Lemma A.1

The proof can be subdivided in three main steps:

1. We first show that there is an integer $0 < i_k^* < k$ such that $p_k(i_k^*)$ is maximal.

2. We then show that
$$\lim_{k\to\infty} \frac{\mathbb{E}(X(k) - i_k^*)}{k} = 0.$$
This implies that $\lim_{k\to\infty} \frac{\mathbb{E}(X(k))}{k}$ exists if and only if $\lim_{k\to\infty} \frac{i_k^*}{k}$ exists.

3. The last step of the proof consists in computing this limit, and showing that it is (A.1).

A.2 Computation

A.2.1 Step 1

We first show that the function $p_k(i)$ is maximum for some $0 < i_k^* < k$, by using functions of the ratios of these quantities. We define the functions $\alpha_k(i)$ for $1 \leq i \leq k$ as

$$\begin{aligned} \alpha_k(i) &\doteq \frac{p_k(i)}{p_k(i-1)} \\ &= r\frac{\binom{km-(m-1)i+n}{i}}{\binom{km-(m-1)(i-1)+n}{i-1}} \\ &= \frac{r(km - m(i-1) + n)_m}{i(km - (m-1)(i-1) + n)_{m-1}} \end{aligned} \tag{A.2}$$

where the notation $(x)_m$ stands for $(x)_m = \prod_{j=0}^{m-1}(x-j)$. Therefore, the ratio between two consecutive terms is

$$\begin{aligned} \frac{\alpha_k(i+1)}{\alpha_k(i)} &= \frac{i}{i+1} \cdot \frac{(km - mi + n)_m}{(km - m(i-1) + n)_m} \cdot \\ &\quad \frac{(km - (m-1)(i-1) + n)_{m-1}}{(km - (m-1)i + n)_{m-1}}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \frac{\alpha_k(i+1)}{\alpha_k(i)} = \frac{i}{i+1} \cdot \frac{km - mi + n - m + 1}{km - mi + n + 1} \cdot \\ \prod_{j=0}^{m-2} \frac{(km - mi + n - j)(km - (m-1)(i-1) + n - j)}{(km - m(i-1) + n - j)(km - (m-1)i + n - j)}. \end{aligned}$$

A close observation of each fraction allows to see that for $m > 1$ each fraction is strictly smaller than one. Thus, α_k is a strictly decreasing function that ranges from $\alpha_k(1) = r((k-1)m+n+1) > 1$ to $\alpha_k(k) = r(m+n)_m/(k(k+m-1+n)_{m-1}) < 1$. This implies that there is a value $0 < i_k^* < k$ such that $\alpha_k(i_k^* + 1) \leq 1$ and $\alpha_k(i_k^*) > 1$. Consequently, $p_k(i_k^*)$ is maximal, and $\alpha_k(i_k^* + 2) < 1$.

A.2.2 Step 2

We now move on to the computation of the expected value of $X(k)$. We will see that the expected value is close to i_k^* for large k; hence we directly compute $\mathbb{E}(X(k) - i_k^*)$ instead of $\mathbb{E}(X(k))$. We also introduce the values $\hat{\imath}_k := i_k^* + f(k)$ and $\check{\imath}_k := i_k^* - f(k)$, where $f(k)$ is an integer function such that $\lim_{k\to\infty} f(k)/\sqrt{k} = 0$ but $\lim_{k\to\infty} f(k)/\log k = \infty$, to decompose the summation in

$$\mathbb{E}(X(k) - i_k^*) = \sum_{i=0}^{k} (i - i_k^*) p_k(i)$$

in three terms

$$\begin{aligned}\mathbb{E}(X(k) - i_k^*) = &\sum_{i=0}^{\check{\imath}_k} (i - i_k^*) p_k(i) + \\ &\sum_{i=\check{\imath}_k+1}^{\hat{\imath}_k-1} (i - i_k^*) p_k(i) + \sum_{i=\hat{\imath}_k}^{k} (i - i_k^*) p_k(i) \end{aligned} \tag{A.3}$$

We show that each of these three terms is of order $o(k)$. For the first one, using the definition of α_k and the fact that it is a strictly decreasing function, we find that

$$\begin{aligned} 0 &\leq \sum_{i=0}^{\check{\imath}_k} (i_k^* - i) p_k(i) \\ &\leq p_k(\check{\imath}_k) \sum_{i=0}^{\check{\imath}_k} (i_k^* - i) \\ &\leq p_k(\check{\imath}_k) \sum_{i=0}^{k} (k - i) \\ &= (k(k+1)/2)\, p_k(\check{\imath}_k) \\ &= (k(k+1)/2)\, p_k(i_k^* - 1) \prod_{j=\check{\imath}_k+1}^{i_k^*-1} (\alpha_k(j))^{-1} \\ &= (k(k+1)/2)\, p_k(i_k^* - 1) \prod_{j=0}^{f(k)-2} (\alpha_k(i_k^* - 1 - j))^{-1} \\ &\leq (k(k+1)/2)\, p_k(i_k^* - 1)(\alpha_k(i_k^* - 1))^{-f(k)+2} \\ &\leq (k(k+1)/2)(\alpha_k(i_k^* - 1))^{-f(k)+2}, \end{aligned}$$

where that last line follows from $p_k(i_k^* - 1) \leq 1$. As $\alpha(i_k^* - 1) > 1$, this expression tends to 0 when k tends to infinity, which shows that the first term in (A.3) is $o(k)$. The third term can be bounded in a similar manner, whereas the computation for

the middle term reads

$$\begin{aligned}\left|\sum_{i=\check{\imath}_k+1}^{\hat{\imath}_k-1}(i-i_k^*)p_k(i)\right| &\leq \sum_{i=\check{\imath}_k+1}^{\hat{\imath}_k-1}|(i-i_k^*)p_k(i)| \\ &= \sum_{i=-f(k)+1}^{f(k)-1}|i|p_k(i+i_k^*) \\ &\leq \sum_{i=-f(k)}^{f(k)}f(k)p_k(i+i_k^*) \\ &\leq (2f(k)+1)f(k) \\ &= o(k).\end{aligned}$$

Combining the three terms, we conclude that (A.3) is of order $o(k)$, and thus that

$$\lim_{k\to\infty}\frac{\mathbb{E}(X(k)-i_k^*)}{k}=0,$$

which implies in turn that $\lim_{k\to\infty}\mathbb{E}(X(k))/k$ exists if and only if $\lim_{k\to\infty}i_k^*/k$ exists.

A.2.3 Step 3

The last step of the proof thus consists in computing the latter limit. We first extend the support of the function $\alpha_k(i)$ from $\mathbb{N}$ to $\mathbb{R}^+$, by observing in (A.2) that $\alpha_k(i)$ is well defined for non-integer values of i. The function $\alpha_k(x)$ over $x\in\mathbb{R}^+$, is a continuous, strictly decreasing function, and we can therefore find the value $x\in\mathbb{R}$ such that $\alpha_k(x)=1$. Clearly, $|x-i_k^*|<1$, so that $\lim_{k\to\infty}x/k=\lim_{k\to\infty}i_k^*/k$. We need therefore to compute the root $x\in\mathbb{R}^+$ of the equation $\alpha_k(x)=1$, which can be expanded from (A.2) as

$$\frac{r\prod_{j=0}^{m-1}(km-m(x-1)+n-j)}{x\prod_{j=0}^{m-2}(km-(m-1)(x-1)+n-j)}=1.$$

Dividing the numerator and denominator by k^m, we obtain

$$\begin{aligned}r\prod_{j=0}^{m-1}(m-m(\frac{x}{k}-\frac{1}{k})+\frac{n-j}{k}) = \\ \frac{x}{k}\prod_{j=0}^{m-2}(m-(m-1)(\frac{x}{k}-\frac{1}{k})+\frac{n-j}{k}).\end{aligned}$$

Taking the limit of both sides for $k\to\infty$ we get

$$r\prod_{j=0}^{m-1}(m-m\frac{x}{k})=\frac{x}{k}\prod_{j=0}^{m-2}(m-(m-1)\frac{x}{k}),$$

and solving for r, we obtain

$$r = \frac{x}{m(k-x)} \left(1 + \frac{x}{m(k-x)}\right)^{m-1}. \tag{A.4}$$

Let

$$y = \left(\frac{x}{m(k-x)r}\right)^{1/(m-1)},$$

or equivalently

$$\frac{x}{k} = \frac{mry^{m-1}}{1 + mry^{m-1}}. \tag{A.5}$$

Using this change of variables in (A.4), we get

$$r = ry^{m-1} \left(1 + ry^{m-1}\right)^{m-1}$$

which we can recast as

$$1 - y - ry^m = 0. \tag{A.6}$$

By Decartes' rule of signs, the difference between the number of positive real roots of a real polynomial and the number of changes of signs of the sequence of its coefficients is always an even, non positive number. The polynomial $1 - y - ry^m$ has one change of sign, and therefore only one positive real root (as we could expect since $\alpha(x)$ is continuous and strictly decreasing for $x \in \mathbb{R}^+$), which we denote by y_1.

Because of (A.5), we have therefore proven that

$$\lim_{k\to\infty} \frac{\mathbb{E}(X(k))}{k} = \frac{mry_1{}^{m-1}}{1 + mry_1{}^{m-1}}$$

where y_1 is the real solution of (A.6). □

Notations

The notations used throughout this thesis are summarized in this table. The notations that are not summarized in this table have only a local meaning.

Expression	Definition	Page
RXRange	The receiving range	5
CSRange	The carrier sensing range	5
L	The number of edges (undirected links) in the network topology	6
RTS	Request to send packet	6
CTS	Clear to send packet	7
ACK	Acknowledgement packet	7
NAV	Network allocation vector. The value of the NAV is used to provide virtual carrier sensing.	7
cw	The contention window used by the backoff mechanism. The backoff value is selected uniformly in $\{0, 1, \ldots, cw\}$.	7
CW_{min}	The minimum value of cw	7
CW_{max}	The maximum value of cw	7
x_j	The state of link j. $x_j \in \{0, 1\}$	9
$\mathcal{E}_j$	The exclusion domain of link j	9
$\mathcal{V}$	The set of transmission patterns feasible under 802.11	11
$n(x)$	The number of active links in transmission pattern x	11
$N(i)$	The number of transmission patterns with i active links	11
λ	One over the average backoff time	12
μ	One over the average exchange time	12
ρ	The access intensity λ/μ (i.e, the ratio of the average exchange and backoff time)	12
$\pi(x)$	The stationary probability of transmission pattern x	12
π_i	The stationary probability to be in a transmission pattern with i active links	13
l	1 plus the minimum distance separating two active links in a line network	14

Expression	Definition	Page
σ	The average spatial reuse	19
FI	The Jain's fairness index	27
$N(i,j)$	The number of transmission patterns with i active links where link j is active	27
$p(j)$	The stationary probability that link j is active	27

Publications

Conference papers

- Mathilde Durvy, Christophe Diot, Nina Taft and Patrick Thiran. Network Availability Based Service Differentiation. In *Proc. International Workshop on Quality of Service (IWQoS)*, Monterey, June 2003.
- Mathilde Durvy and Patrick Thiran. Reaction-Diffusion Based Transmission Patterns for Ad Hoc Networks. In *Proc. IEEE Infocom*, Miami, March 2005.
- Mathilde Durvy and Patrick Thiran. A Backoff Mechanism to Achieve Full Organization. Abstract presented *SpaSWiN*, Boston, April 2006.
- Mathilde Durvy and Patrick Thiran. A Packing Approach to Compare Slotted and Non-Slotted Medium Access Control. In *Proc. IEEE Infocom*, Barcelona, April 2006.
- Mathilde Durvy and Patrick Thiran. Understanding the Gap between the IEEE 802.11 Protocol Performance and the Theoretical Limits. In *Proc. SECON*, Reston, September 2006.
- Mathilde Durvy, Olivier Dousse and Patrick Thiran. Modeling the 802.11 Protocol under Different Capture and Sensing Capabilities. In *Proc. IEEE Infocom (mini-symposium)*, Anchorage, May 2007.
- Mathilde Durvy, Christina Fragouli, and Patrick Thiran. Towards Reliable Broadcasting using ACKs. In *Proc. ISIT*, Nice, June 2007.
- Mathilde Durvy, Olivier Dousse and Patrick Thiran. Border Effects, Fairness, and Phase Transition in Large Wireless Networks. To appear in *Proc. IEEE Infocom*, Phoenix, April 2008.

Journal papers

- Mathilde Durvy, Olivier Dousse and Patrick Thiran. Modeling the 802.11 Protocol under Different Capture and Sensing Capabilities. *Submitted for publication*.

Curriculum Vitæ

Mathilde Durvy was born in 1979 in Luxembourg. She earned a M.S. degree in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland in October 2002. Her global average was 5.89 out of 6, making her one of the top students of her class.

During her undergraduate studies she was selected by EPFL's exchange program and spent one year at Carnegie Mellon University (CMU) in Pittsburgh, Pennsylvania.

Her master's thesis was done at Sprint Advanced Technology Laboratories (California) on Network Availability based Service Differentiation.

In November 2002, she joined the Laboratory for Computer communication and Applications (LCA) and began working on her PhD thesis under the supervision of Professor Patrick Thiran. There, she participated to the National Center of Competence in Research on Mobile Information and Communication Systems (NCCR-MICS).

During her PhD work, she was a teaching assistant for the class of stochastic models for communication.

In 2005 she received an Infocom travel grant and in 2007 she was a finalist of the Google Anita Borg scholarship.

Her current research interests are in multi-hop wireless networks. She is a reviewer for the IEEE/ACM Transactions on Networking, Infocom, and other conferences.

Some invited talks

- "A Packing Approach to Compare Slotted and Non-slotted Medium Access Control", *MICS Scientific Conference*, Zürich, October 2006.
- "A Backoff Mechanism to Achieve Full Organization?"
  Deutsche Telekom Laboratories, Berlin, June 2006
  Inter perf, Pisa, October 2006.
- "Modeling the 802.11 Protocol under Different Capture and Sensing Capabilities"
  Thomson Paris Research Lab, Paris, April 2007
  MICS Workshop, Neuchatel, July 2007.

Bibliography

[AF] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs. http://www.stat.berkeley.edu/ aldous/RWG/book.html.

[ALM98] Giuseppe Anastasi, Luciano Lenzini, and Enzo Mingozzi. Stability and performance analysis of HIPERLAN. In *INFOCOM (1)*, pages 134–141, 1998.

[BCJ04] Yuliy Baryshnikov, E. G. Jr Coffman, and Predrag Jelenkovic. Space filling and depletion. In *J. Appl. Probab.*, volume 42, pages 691–702, 2004.

[BDSZ94] Vaduvur Bharghavan, Alan Demers, Scott Shenker, and Lixia Zhang. MACAW: a media access protocol for wireless LAN's. In *SIGCOMM*, pages 212–225. ACM Press, 1994.

[Bia00] Giuseppe Bianchi. Performance Analysis of the IEEE 802.11 Distributed Coordination Function. In *IEEE J. Select. Areas Commun.*, volume 18, 2000.

[BKMS87] R.R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin. Throughput analysis in multihop CSMA packet radio networks. In *IEEE Trans. Commun.*, volume 35, pages 267–274, 1987.

[BLL84] D. Y. Burman, J. P. Lehoczky, and Y. Lim. Insensitivity of blocking probabilities in a circuit-switching network. In *J. Appl. Probab.*, volume 21, pages 850–859, 1984.

[BMP07] C. Bordenave, D. McDonald, and A. Proutiere. A particle system in interaction with a rapidly varying environment: Mean field limits and applications. *ArXiv Mathematics e-prints*, January 2007.

[Bré98] P. Brémaud. *Markov chains: Gibbs field, Monte Carlo simulation, and queues*. Springer-Verlag, 1998.

[CCG00a] Frederico Calì, Marco Conti, and Enrico Gregori. Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Trans. Networking*, 8(6):785–799, 2000.

[CCG00b] Frederico Calì, Marco Conti, and Enrico Gregori. IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism. *IEEE J. Select. Areas Commun.*, 18(9):1774–1780, 2000.

[CDL05] Claude Chaudet, Dominique Dhoutaut, and Isabelle Guérin Lassous. Experiments of Some Performance Issues with IEEE 802.11b in Ad Hoc Networks. In *WONS*, 2005.

[CGLA04] Marcelo M. Carvalho and J. J. Garcia-Luna-Aceves. A scalable model for channel access protocols in multihop ad hoc networks. In *MobiCom*, pages 330–344. ACM Press, 2004.

[DLV04] J. Deng, B. Liang, and P.K. Varshney. Tuning the Carrier Sensing Range of IEEE 802.11 MAC. In *IEEE GLOBECOM*, 2004.

[Dou07] O. Dousse. Revising buffering in CSMA/CA wireless multihop networks. In *SECON, San Diego*, 2007.

[Dur81] Richard Durrett. An introduction to infinite particle systems. *Stochastic Processes and their Applications*, 11(2):109–150, 1981.

[GGK01] P. Gupta, R. Gray, and P.R. Kumar. An experimental scaling law for ad hoc networks. UIUC, May 2001.

[GHM98] H. Georgii, O. Haggstrom, and C. Maes. The random geometry of equilibrium phases, 1998.

[GJ83] M.R. Garey and D.S. Johnson. Computers and intractability: A guide to the theory of NP-completeness. In *Freeman*, New York, 1983.

[GK03] Nitin Gupta and P. R. Kumar. A performance analysis of the IEEE 802.11 Wireless LAN Medium Access Control. In *Communications in Information and Systems*, volume 3, pages 279–304, 2003.

[Gri99] G. Grimmett. *Percolation*. Springer-Verlag, 1999.

[GSK05] Michele Garetto, Jingpu Shi, and Edward W. Knightly. Modeling media access in embedded two-flow topologies of multi-hop wireless networks. In *MobiCom*, 2005.

[GSK06] M. Garetto, T. Salonidis, and E. Knightly. Modeling Per-flow Throughput And Capturing Starvation in CSMA Multi-hop Wireless Networks. In *INFOCOM, Barcelona*, 2006.

[Hei74] O. J. Heilmann. The use of reflection as symmetry operation in connection with the peierls' argument. In *Commun. math. Phys.*, volume 36, pages 91–114, 1974.

[HP73] O. J. Heilmann and E. Praestgaard. Phase transition of hard hexagons on a triangular lattice. In *Journal of Statistical Physics*, volume 9, pages 23–44, 1973.

[iee99] IEEE802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Aug 1999.

[Jai91] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley and Sons, Inc., 1991.

[JDN01] N. Jain, S.R. Das, and A. Nasipuri. A Multichannel CSMA MAC protocol with Receiver-Based Channel Selection for MultiHop Wireless Networks. In *IEEE IC3N*, 2001.

[KAMG07] A. Kumar, E. Altman, D. Miorandi, and M. Goyal. New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs. In *IEEE/ACM Trans. Networking*, volume 15, 2007.

[Kel91] F. P. Kelly. Loss networks. In *The Annals of Applied Probability*, volume 1, pages 319–378, 1991.

[KKB00] Can Emre Koksal, Hisham Kassab, and Hari Balakrishnan. An analysis of short-term fairness in wireless media access protocols. In *Measurement and Modeling of Computer Systems*, pages 118–119, 2000.

[KRM02] I. Ziedins K. Ramanan, A. Sengupta and P. Mitra. Markov random field models of multicasting in tree loss networks. In *Advances in Applied Probability*, volume 34, pages 58–84, 2002.

[KS80] R. Kindermann and J. Laurie Snell. *Markov Random Fields and their Applications*. American Mathematical Society, Series: Contemporary mathematics, 1980.

[KVSA04] A. Kochut, A. Vasan, A. Shankar, and A. Agrawala. Sniffing out the correct physical layer capture model in 802.11b. In *ICNP*, 2004.

[LNG06] Zhifei Li, Sukumar Nandi, and Anil K. Gupta. Modeling the short-term unfairness of IEEE 802.11 in presence of hidden terminals. *Perform. Eval.*, 63(4):441–462, 2006.

[Lou90] G. M. Louth. *Stochastic networks: complexity, dependence and routing*. (thesis) Cambridge Univ., 1990.

[LRZ06] B. Luen, K. Ramanan, and I. Ziedins. Nonmonotonicity of phase transitions in a loss network with controls. In *Ann. Appl. Probab.*, volume 16, pages 1528–1562, 2006.

[Mac62] J.K. Mackenzie. Sequential filling of a line by intervals placed at random and its applications to linear adsorption. In *J. Chem. Phys.*, volume 37, pages 723–728, 1962.

[MT06] K. Medepalli and F.A. Tobagi. Towards Performance Modeling of IEEE 802.11 based Wireless Networks: A Unified Framework and its Applications. In *INFOCOM, Barcelona*, 2006.

[ND00] A. Nasipuri and S.R. Das. Multichannel CSMA with signal power-based channel selection for multihop wireless networks. In *IEEE Vehicular Technology Conference*, 2000.

[ns+] IEEE 802.11 simulator (based on ns-2) with remedies to the gagged, jammed, and focused node situations. Available at http://icapeople.epfl.ch/mdurvy/research.html.

[ns2] Network simulator 2 (ns-2). http://www.isi.edu/nsnam/ns/.

[PY86] E. Pinsky and Y. Yemini. The asymptotic analysis of some packet radio networks. In *IEEE J. Select. Areas Commun.*, number 6, pages 938–945, 1986.

[qns] Qualnet network simulator. http://www.scalable-networks.com/.

[RC05] A. Raniwala and T.-C. Chiueh. Architecture and Algorithms for an IEEE 802.11-Based Multi-Channel Wireless Mesh network. In *IEEE INFOCOM*, 2005.

[RCS03] S. Ray, J.B. Carruthers, and D. Starobinski. RTS/CTS-induced congestion in ad hoc wireless lans. In *WCNC*, volume 3, pages 1516–1521, March 2003.

[RCS05] S. Ray, J.B. Carruthers, and D. Starobinski. Evaluation of the masked node problem in ad-hoc wireless LANs. In *IEEE Transactions on Mobile Computing*, volume 4, pages 430 – 442, September 2005.

[SHS04] K. Sundaresan, H.-Y. Hsieh, and R. Sivakumar. IEEE 802.11 over Multi-hop Wireless Networks: Problems and New Perspectives. In *Ad Hoc Networks*, volume 2, pages 109–132, 2004.

[SV04] J. So and N. Vaidya. Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using A Single Transceiver. In *MobiHoc*, 2004.

[ths] Discrete event simulator for the idealized 802.11 protocol. Available at http://icapeople.epfl.ch/mdurvy/research.html.

[TK75] F. A. Tobagi and L. Kleinrock. Packet switching in radio channels: Part II - the hidden terminal problem in carrier sense multiple-access modes and the busy-tone solution. In *IEEE Transactions on Communications*, page 23(12):1417 1433, 1975.

[Tob87] F. A. Tobagi. Modeling and performance analysis of multihop packet radio networks. In *Proceeding of the IEEE*, volume 75, pages 135–155, 1987.

[TTV00] J. Talbot, G. Tarjus, and P. Viot. The adsorption-desorption model and its application to vibrated granular materials. *Phys. Rev.*, E 61(5429), 2000.

[vdB93] J. van den Berg. A uniqueness condition for gibbs measures, with application to the 2-dimensional ising antiferromagnet. *Communications in Mathematical Physics*, 152:161–166, 1993.

[vdBM94] J. van den Berg and Chr. Maes. Disagreement percolation in the study of markov fields. *Ann. probab.*, 22(2):91–114, 1994.

[vdBS94] J. van den Berg and J. E. Steif. Percolation and the hard-core lattice gas model. *Stochastic Process. Appl.*, 49:179–197, 1994.

[WGLA04] Yu Wang and J. J. Garcia-Luna-Aceves. Modeling of collision avoidance protocols in single-channel multihop wireless networks. *Wirel. Netw.*, 10(5):495–506, 2004.

[WK05] X. Wang and K. Kar. Throughput Modelling and Fairness Issues in CSMA/CA Based Ad-Hoc Networks. In *INFOCOM, Miami*, 2005.

[XGB02] K. Xu, M. Gerla, and S. Bae. How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks? In *GLOBECOM*, 2002.

[XS01] S Xu and T. Saadawi. Does the IEEE 802.11 MAC protocol work well in multihop wireless adhoc networks? In *IEEE Commun. Mag.*, volume 39, pages 130–137, 2001.

[Yem83] Y. Yemini. A statistical mechanics of distributed resource sharing mechanisms. In *INFOCOM*, pages 531–539, 1983.

[YS04] F. Ye and B. Sikdar. Distance-Aware Virtual Carrier Sensing for Improved Spatial Reuse in Wireless Networks. In *IEEE GLOBECOM*, 2004.

[YYS03] F. Ye, S. Yi, and B. Sikdar. Improving Spatial Reuse of IEEE 802.11 Based Ad Hoc Networks. In *IEEE GLOBECOM*, 2003.