Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

SCUOLA DI DOTTORATO IN INGEGNERIA DELL'INFORMAZIONE

INDIRIZZO IN SCIENZA E TECNOLOGIA DELL'INFORMAZIONE

XXII Ciclo

Cooperative Medium Access Control Policies in Wireless Networks

Dottorando

ANDREA MUNARI

Supervisore:
Chiar.mo Prof. Michele Zorzi

Direttore della Scuola:
Chiar.mo Prof. Matteo Bertocco

Anno Accademico 2009/2010

To my family

Contents

Abstract

Broadly speaking, wireless ad hoc networks are permeated by cooperative behaviors. In such systems, indeed, nodes have to continuously pool their resources to achieve goals that are of general interest, such as routing packets towards a destination that would otherwise be out of reach for an information source, or coordinating medium access so as to successfully share a common spectrum. Recently, however, the idea of collaboration has gathered a renewed and increasing deal of attention in the research community thanks to the development of innovative concepts, most notably the idea of cooperative relaying, that have been shown to unleash significant improvements, going beyond some intrinsic limitations that affect wireless communications systems. While these emerging solutions have been thoroughly studied from a theoretical perspective, the advantages they offer can be reaped in real world implementations only if additional coordination among nodes is provided, so that cooperating terminals are allowed to offer their help obeying the rules that control medium access, and without disrupting the normal activity of the network.

Along this line of reasoning, this thesis focuses on the design and study of link layers that implement cooperation in ad hoc networks. The contribution of our work is twofold. On the one hand, we introduce novel and beneficial collaborative strategies, while on the other hand we investigate how the intrinsic nature of different medium access control policies can be more or less beneficial to cooperative behaviors.

In the first part, we present an innovative approach to cooperation in networks with multi-antenna equipped nodes that rely on directional transmissions and receptions. Our solution proposes terminals to share information on ongoing communications to better coordinate medium access at the link layer, and manages to overcome issues such as node deafness that typically hamper the potential of large scale directional wireless systems.

The central part of this work, conversely, concentrates on the simpler ad hoc scenario where omnidirectional communications are in place, and tackles some inefficiencies of the cooperative relaying paradigm. In particular, we introduce the novel concept of hybrid cooperative-network coded ARQ, which allows a relay to code data of its own together

with a corrupted packet during a retransmission at no additional cost in terms of bandwidth. Such a solution encourages nodes to cooperate, since they are offered the possibility to pursue a goal of their interest while helping surrounding terminals. Moreover, the capability of exploiting retransmissions to serve additional traffic, achieved by smartly taking advantage of network coding techniques, triggers beneficial effects also at a network level in terms of sustainable throughput and reduced congestion. The potential of the proposed approach is first investigated by means of mathematical analysis, while subsequently extensive simulation campaigns test the effectiveness of link layers that implement it in a variety of networking environments.

Taking the cue from a reasoned comparison of the results achieved by relaying schemes in different scenarios, we devote the final part of the thesis to the investigation of the impact that distinct spectrum control policies can have on collaborative behaviors. Combining once again mathematical analysis and simulations, we consider how the characteristics of completely distributed, e.g., carrier sense based, and centralized, e.g., time division based, systems influence the effectiveness of a given cooperative strategy. Not only does our study shed light on the relations that exist between cooperation and medium access, but also it provides important hints on how to efficiently design link layers capable of supporting such techniques.

Finally, the appendix of this thesis reports the outcome of a research activity, carried out in collaboration with the IBM Zurich Research Laboratory (Switzerland), whose focus falls out of the topic of cooperative link layers and covers the design of energy-efficient routing protocols for wireless sensor networks.

Sommario

Le reti wireless ad hoc presentano in generale moltissimi comportamenti di natura cooperativa, nei quali i nodi condividono le loro risorse per perseguire un interesse di utilità comune. Basti pensare, in tal senso, alle procedure di routing per la consegna di traffico multihop, o allo scambio di informazioni tra terminali necessario per riuscire a gestire in modo efficace uno spettro condiviso. Recentemente, inoltre, è progressivamente emerso un rinnovato e crescente interesse nella comunità di ricerca per il concetto di collaborazione tra nodi, grazie allo sviluppo di nuovi paradigmi, tra i quali in primis l'idea del relaying cooperativo, che si sono dimostrati in grado di mitigare brillantemente alcuni problemi tipici dei sistemi wireless, rendendo possibili significativi miglioramenti delle prestazioni. Sebbene tali soluzioni innovative siano state oggetto di notevole attenzione in letteratura, gli studi su di esse si sono concentrati principalmente su trattazioni di natura analitica, atte a dimostrarne le potenzialità e i vantaggi nell'ottica della teoria dell'informazione. Approcci di questo tipo tendono chiaramente a considerare, ai fini della trattabilità matematica, topologie semplificate quali reti a tre soli nodi, e spesso assumono un accesso al mezzo idealizzato. Nel momento in cui queste idee vogliano essere implementate in scenari reali, tuttavia, si rende necessario un profondo raffinamento della coordinazione a livello di rete, dal momento che i nodi cooperanti devono comunque sottostare alle regole che caratterizzano la gestione del canale (link layer), di modo da offrire il loro contributo senza ostacolare la normale attività della rete.

Prendendo spunto da tale riflessione, questa tesi si concentra sulla definizione e l'analisi di link layer che implementino soluzioni cooperative in reti ad hoc. Due sono i principali contributi del lavoro. Se da un lato, infatti, sono introdotti paradigmi innovativi ed efficaci, dall'altro viene presentato uno studio articolato e completo su come diverse politiche di accesso al mezzo possano influenzare tali comportamenti cooperativi.

La prima parte della tesi si focalizza sullo sviluppo di un nuovo approccio collaborativo per reti i cui terminali, dotati di sistemi multiantenna, siano in grado di effettuare trasmissioni e ricezioni direzionali. L'idea proposta prevede che i nodi condividano, tramite scam-

bio di brevi pacchetti di controllo, informazioni sulle comunicazioni attive di cui sono a conoscenza, per poter favorire la maggior distribuzione possibilie di una percezione corretta dello stato del sistema al fine di garantire una migliore coordinazione nell'accesso al mezzo. Studi dedicati dimostrano come tale soluzione sia in grado di superare problemi, quali la sordità di nodo (deafness), che spesso limitano l'efficacia delle trasmissioni direzionali in reti con numero elevato di dispositivi, portando a importanti guadagni in termini di prestazione complessive.

La parte centrale del lavoro, al contrario, prende in considerazione reti ad hoc con comunicazioni omnidirezionali, e affronta alcune inefficienze che caratterizzano il paradigma di relaying cooperativo. In particolare, viene introdotto per la prima volta il concetto innovativo di ARQ ibrido cooperativo-network coded, che permette a nodi che agiscano da relay di utilizzare le ritrasmissione di un pacchetto in vece di una sorgente, non in grado di consegnarlo, al fine di servire anche del proprio traffico. Tale approccio, a differenza del comportamento puramente altruistico richiesto dal relaying semplice, incoraggia i terminali a cooperare, offrendo loro la possibilità di perseguire un loro interesse contingente nell'atto stesso di aiutare altri nodi in difficoltà. Inoltre, la capacità di sfruttare il meccanismo di ritrasmissione per servire traffico addizionale, resa possibile dall'utilizzo di tecniche di combinazione lineare sui dati caratteristiche del network coding, getta le basi per benefici anche a livello di rete, quali un incremento del throughput sostenibile e una riduzione della congestione nell'utilizzo della banda. Le potenzialità della soluzione identificata sono dapprima studiate per mezzo di modelli matematici, seguendo le modalità tipicamente riscontrabili in letteratura. Successivamente sono proposti l'implementazione e lo studio simulativo di diversi link layer in grado di supportare tale forma di ARQ ibrido in contesti differenti, quali reti completamente distribuite e reti maggiormente strutturate.

Traendo spunto da un confronto ragionato dei risultati ottenibili da schemi di relaying in scenari di rete diversi, la parte finale di questa tesi è dedicata alla discussione dell'impatto che politiche di accesso al mezzo distinte possono avere su comportamenti di natura cooperativa. Combinando ancora una volta analisi matematica e studi simulativi, viene affrontato il problema di come le caratteristiche intrinseche di sistemi basati su carrier sensing e su condivisione del mezzo a multiplazione di tempo influenzino l'efficacia di meccanismi di collaborazione tra nodi. Le osservazioni ottenute tramite questo approccio non solo mettono in luce la stretta relazione esistente tra politiche di gestione dello spettro e cooperazione, ma al tempo stesso forniscono importanti suggerimenti sulla progettazione di link layer in grado di supportare in modo efficace tali strategie.

In appendice, infine, vengono riportati i risultati di attività di ricerca svolte in collaborazione con i laboratori di ricerca IBM di Zurigo (Svizzera) e incentrate su tematiche che si discostano leggermente dal fulcro della tesi, quali la progettazione di teniche di routing per reti wireless di sensori, con particolare attenzione all'efficienza energetica.

Chapter 1

Introduction

Cooperative techniques are attracting an ever growing interest in the wireless research community. The basic idea that underpins these strategies is to have nodes help each other with the aim of improving the overall network efficiency, e.g., by providing mutual support in case of weak radio links so as to enhance the reliability of data transfers, or by distributing information that is useful to better coordinate communications. While this paradigm requires devices to willingly and selflessly pool their resources in terms of energy and bandwidth, it also has the capability of enabling significant performance gains at a broader level, as it manages to go beyond some important limitations that typically beset wireless networks. In order to reap the most out of these benefits, however, a level of coordination among nodes higher than the one required by normal operation mode is to be provided, so as to detect when cooperation is actually useful or necessary, as well as to properly organize medium access for all the players involved in a collaborative process. In fact, non-harmonized cooperative actions may trigger higher interference and congestion, ending up being harmful rather than beneficial. In this perspective, then, the design of specifically tailored medium access policies plays a role of pivotal importance. This aspect, however, is often neglected in the literature, as most works concentrate on the proposal of novel collaborative approaches and on the analysis of their performance focusing on toy-topologies and simplistic scenarios with perfectly coordinated access control.

Starting from these remarks, this thesis deals with different cooperative schemes in wireless ad hoc networks, devoting particular attention to the definition of link layers that are capable of effectively supporting them, as well as investigating the performance of the considered solutions in large and realistic networking environments. The contribution of the work presented here is twofold. Indeed, while on the one hand we propose novel paradigms

for cooperation that are capable of enabling important gains over state of the art techniques, we also thoroughly study the interactions between some of the most widespread medium access control policies and existing collaborative schemes, highlighting important trade-offs that arise when such approaches have to be integrated at the link layer, and shedding light on the suitability of different MACs to support cooperative behaviors.

The first part of the thesis concentrates on ad hoc networks whose nodes, equipped with multiple antennas, are capable of performing directional transmissions and receptions. This feature allows multiple communications to be active at the same time even among nodes that are within each other's transmission range (spatial multiplexing), and thus trigger a potential boost in the sustainable throughput. On the other hand, however, it poses novel challenges at the link layer, as directional information transfers worsen the perception nodes have of the network activity going on around them. As a result, for instance, a terminal may incorrectly try to contact a neighbor of its not being aware that it is already involved in other communications, wasting energy and bandwidth as well as increasing the overall interference level. Such uncoordinated accesses can completely neutralize the benefits brought by multi-antenna systems. In Chapter 2 we tackle this issue by introducing a novel cooperative scheme that proposes terminals to help surrounding nodes with an inaccurate knowledge of ongoing activities by sending them short control packets that properly describe the status of the network. Extensive analysis and simulations show how the presented solution significantly improves the overall performance with respect to relevant benchmarks, and provide a first confirmation of the beneficial effects that properly designed collaborative link layers can trigger.

The core of this work (Chapters 3 to 5), conversely, considers ad hoc networks that rely on omnidirectional communications and single-antenna equipped terminals. Such systems are of particular interest in light both of their diffusion and of the suitability to a number of applications offered by their distributed and self-organizing nature. Moreover, in the second part of this thesis we focus on a different paradigm, i.e., *cooperative relaying*, that has recently emerged as a powerful tool to combat channel impairments such as fading and shadowing by smartly exploiting the spread of information that stems from the broadcast nature of the wireless medium. When an omnidirectional packet transmission is performed between a source and a destination, indeed, the sent payload may be overheard by nodes other than the intended addressee that fall within the transmitter's coverage range. Such an effect is usually regarded as an intrinsic inefficiency of the channel, since terminals are forced to listen to a communication they are not involved in, spending resources, e.g., the en-

ergy needed for reception, to retrieve information they do not need and synchronizing to a waveform they would otherwise regard as interference. Cooperative relaying, instead, proposes nodes to cache incoming data even when not addressed to them so as to support the source-destination pair in case of need. In fact, if the addressee does not decode the packet, basic medium access policies require the originator of the data to perform another attempt. However, if the reception has failed due to a bad channel state, successive transmissions are not likely to succeed unless iterated after a period long enough for the wireless medium to decorrelate. This strategy often leads to long and potentially unreliable failure recovery procedures, especially in congested networks. Conversely, when cooperation is implemented, a terminal that has cached the original packet acts as relay by immediately sending redundancy to the destination on behalf of the source. The intended recipient, then, can apply combining techniques on the fragments of information collected over the spatially disjoint and therefore statistically independent source-to-destination and relay-to-destination channels, significantly improving its decoding probability and shortening the average packet delivery time.

While this strategy has been shown to enable important gains from a theoretical angle, e.g., in terms of sustainable rate or outage probability, it explicitly asks terminals to behave selflessly, helping their neighbors without anything in return but the greater good of the network. Observations of this kind may actually deter devices from cooperating and hamper the beneficial effects that could be achieved by having all devices offering relayed support. In view of this, we introduce a novel paradigm that is capable of going beyond such limitation: *hybrid cooperative-network coded ARQ*. With this technique, nodes are offered the possibility of exploiting relayed phases to serve their own traffic as well, by sending a linear combination of the payload originated by a third party source and of a data unit taken from their own queue. By leveraging a novel type of physical layer, our solution allows the destination to retrieve both information contents, encouraging nodes to take part in the collaborative process by offering them a tangible reward while preserving the advantages of basic relaying.

Hybrid cooperative-network coded ARQ is first addressed in Chapter 3, where we present in detail our idea and develop an analytical model that highlights its potential over plain relaying strategies. Moreover, following the approach that characterizes this thesis, we discuss the design of two link layers that are capable of efficiently supporting such and advanced form of cooperation in networks that rely on carrier sensing and time division multiplexing as multiple access control. Extensive simulation campaigns shed light on two impor-

tant aspects. In the first place, they validate the effectiveness of coded retransmissions and quantify the improvements achievable in a variety of networking scenarios. Secondly, our investigation points out how the same collaborative paradigm can lead to radically different performance when implemented obeying to the principles of distinct medium sharing policies.

Such an observation suggests the existence of a strong relation between the benefits that can be reaped from cooperation and the rules of the underlying link layer. We further analyze this aspect in Chapter 4 by explicitly comparing effective implementations of both cooperative relaying and hybrid cooperative-network coded ARQ in CSMA- and TDMA-based networks and by delving into the reasons that make each policy more or less suitable to support collaboration. The main outcome of our investigation is that even a loosely centralized system like TDMA can provide the necessary coordination among nodes to unleash far higher gains.

In spite of this conclusion, carrier sensing stands as one of the most common medium access control policies in present days wireless networks, thanks to standard like the IEEE 802.11 (WLANs) and 802.15.4 (Wireless Sensor Networks), and is thus very often used as reference link layer for the implementation of cooperative protocols. In light of this importance, we specifically devote Chapter 5 to study the impact of CSMA on relaying approaches. Our investigation has two main merits. In the first place, we consider different strategies, ranging from the basic cooperative relaying paradigm discussed earlier to solutions that take advantage of Channel State Information to collaboratively split a data transmission between a source and a destination over a two-hop link involving a helping terminal so as to maximize the sum rate of the communication. This approach, then, allows us to infer conclusions that are of broad applicability, highlighting how the intrinsic nature of the distributed CSMA policy induces detrimental effects on many forms of cooperation. On the other hand, our study combines analytical results obtained developing frameworks that focus on simple topologies and simulation campaigns that deal with large and realistic networks. We believe indeed that both components are fundamental to a thorough understanding of the relations that arise between link layers and cooperation: the analysis of simplified scenarios enables the identification of basic issues and features of the system and allows a proper design of collaborative schemes, whereas a detailed simulation of a large network highlights many practical issues that mathematical models cannot account for.

The final part of the thesis (Appendix A) reports the results of a collaboration carried out during the Ph.D. with the Energy Management and Sensors Systems group at the IBM

Zurich Research Laboratory (Switzerland), whose focus falls out of the main scope of the work presented here. The research activity concentrated on Wireless Sensor Networks, which are emerging as an extremely interesting solution for distributed information gathering and monitoring. Such systems are typically composed by a large number of battery-powered and computationally limited nodes with sensing capabilities that are scattered over a wide area and that have to operate without external intervention for long periods of time. In many applications, the sensors have to form an ad hoc network to deliver data from an information source to one or several destinations in a multihop fashion. Therefore, routing algorithms play a key role in Wireless Sensor Networks, and particular attention has to be devoted to energy consumption, so as to prolong the network lifetime. In this thesis we propose several novel routing solutions, covering issues that range from cross-layer design to support for mobile sensor nodes and robustness to security attacks, always under the paradigm of energy efficiency.

Chapter 2

Cooperative Link Layers in Directional Ad Hoc Networks

Wireless ad hoc networks are attracting an ever increasing deal of research interest, thanks to their capability of providing radio connectivity in a fully distributed fashion and without resorting to any central coordinating unit [1,2]. Indeed, not only is the possibility to exchange traffic in a peer-to-peer fashion anytime-anywhere extremely appealing, but also many critical situations may require data transfer or multimedia streaming in places where no infrastructure is available at all, e.g., battlefields or disaster relief scenarios.

On the other hand, the lack of a coordinating element also poses challenges that may restrain the potential of ad hoc solutions with respect to networks that rely on a centralized structure, e.g., cellular systems or WLANs with an access point. A lower degree of coordination among terminals, in fact, may trigger multiple simultaneous and unorganized accesses to the medium, leading to a level of interference that may be critical for the effectiveness of all the upper layers. From this viewpoint, for instance, the Automatic Repeat Request (ARQ) mechanisms would issue more retransmissions, routes that converge at the same node would contend for access to that bottleneck, and, last but not least, transport protocols like TCP would have to cope with very high packet loss rates. Moreover, all these issues are further exacerbated by drawbacks such as fading and shadowing that characterize the highly time-varying wireless channel. As a result, the overall network performance, e.g., in terms of throughput or delay, can heavily deteriorate [3].

In this context, one of the most promising tools to preserve physical layer (PHY) reliability is the introduction of smart antennas [4]. Such systems exploit multiple antennas and, by means of beamforming, are able to focus the transmitted energy toward a desired di-

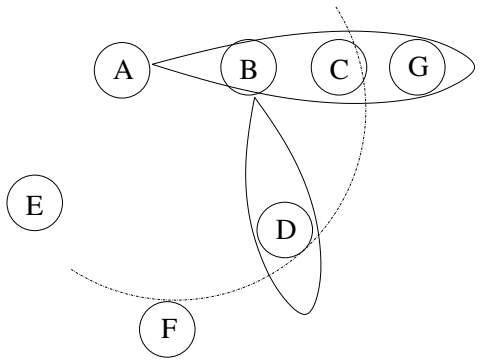


Figure 2.1. *Spatial reuse: a case where conventional protocols would fail. The areas surrounded by the geometrical shapes describe zones where a packet can be correctly decoded.*

rection, adapting the antenna signals so that they may constructively combine in the aimed bearing. In addition, the same technique enables a node to improve the reception of the intended signal by rejecting waveforms coming from unwanted directions, i.e., reducing the perceived interference level. The combination of these effects makes it possible, in principle, to increase the number of simultaneous communications in the network.

However, while the introduction of smart antennas unquestionably enhances the quality of point-to-point links, it also brings up critical issues at the link layer. First of all, simple extensions of protocols designed for omnidirectional antennas, most notably the IEEE 802.11, may be too conservative. Consider, as a reference, the situation depicted in Fig. 2.1, and assume that a communication between $\boldsymbol{A}$ and $\boldsymbol{C}$ has been set up. If the IEEE 802.11 Distributed Coordination Function (DCF) [5] is used as medium access control (MAC), a node, say $\boldsymbol{D}$, that received handshake packets (RTS/CTS) exchanged between source and destination is not allowed to start a new transmission in order to avoid a collision. If directional communications are in place, however, such an approach would inefficiently stymie throughput by preventing sustainable streams, as, for instance, a beamformed transmission between $\boldsymbol{B}$ and $\boldsymbol{D}$ may start without causing excessive interference to the $\boldsymbol{A}$-$\boldsymbol{C}$ data flow.

Another remarkable problem that arises at the link layer when directional communications are in place is *deafness*, that can be defined as the inability of a node to reply to an incoming packet. Such a situation may occur, as in all wireless environments, because of a collision. However, due to the distributed nature of ad hoc networks as well as to the directionality of transmissions and receptions in the scenario under analysis, the addressee of a packet may not respond also because it is already involved in an ongoing communication the sender is not aware of, or simply because it is not listening in the bearing of the transmitter. With reference to the situation reported in Fig. 2.1, this would be the case if $\boldsymbol{A}$

was to contact $\boldsymbol{B}$ while such a node is transferring data to $\boldsymbol{D}$. According to 802.11, however, the sender would always interpret this missing reply as a sign of congestion, reacting in a way that may not be effective. Problems of this kind stem from the lack of coordination among terminals in the network, and, if not properly handled, can completely neutralize the potential advantages offered by smart antennas [6].

Starting from these remarks, in this chapter we propose and discuss a class of novel medium access protocols that are capable of overcoming many of the challenges that beset directional ad hoc networks. The contribution of our work is twofold. In the first place, we develop a thorough analysis of the possibilities and drawbacks offered by smart antennas, shedding light on the proper constraints they impose on the protocol design. This investigation leads to the proposal of a novel, low-complexity packet reception mechanism that can significantly improve performance. Secondly, we introduce the idea of link layer cooperation among nodes. According to the solution that we present, each terminal exchanges with its neighbors control packets describing the ongoing transmissions it is aware of. Such an approach makes it possible to improve the perception a node has of the surrounding network activity, therefore contributing to a better coordination in medium access and bounding the impact of deafness. By means of analysis and extensive simulations, we show that the proposed policy achieves much higher performance than other existing strategies.

The rest of the chapter is organized as follows: Section 2.1 reviews the existing work in the area; Section 2.2 introduces the topic of antenna arrays and the interaction between protocol organization and array design; Section 2.3 analyzes the problem of deafness; Section 2.4 describes the solution proposed by our MAC protocols to deal with this issue; Section 2.5 and Section 2.6 discuss the problem of collisions in a directional scenario and introduce the CMAC alignment algorithm; Section 2.7 presents the simulation results; Section 2.8 quickly introduces an analytical model for predicting the impact of collisions in our system; Section 2.9 outlines the format of some important packets. Finally, Section 2.10 draws the conclusions of this first part of the thesis.

2.1 Related Work

In the past half decade considerable research has been devoted to the creation of cross-layer PHY-MAC protocols that may exploit the benefits of directional antennas while minimizing the new problems introduced by this hardware enhancement. Some of the first attempts studied the IEEE 802.11 DCF (or slightly modified versions of it) on top of this new physical layer [7], [8–11]. These approaches were heavily beset by two main problems:

asymmetry in gain and deafness. The former consists in determining a way to deliver a packet omnidirectionally but at the range achieved by directional transmissions. To this aim the Circular RTS was introduced in [12] and later extended into the CRCM in [13]. This technique solves the asymmetry in gain by successively beamforming a broadcast frame (e.g., a control packet) toward different bearings, until the whole horizon is swept. However, this comes at the price of a higher latency.[1]

As far as deafness is concerned, a further analysis is developed in [15], which suggests the usage of special busy tones to reduce its impact. The performance improvements are remarkable, but this technique needs out-of-band signaling. Another solution to this problem is outlined in [16], where the extent of deafness is limited by a more accurate handling of directional receptions to avoid capture due to unwanted packets.

The issue of whether the simple reception of an RTS/CTS should enforce virtual carrier sense has been addressed in [8], [10] and [17]. The first two independently proposed the concept of Directional NAV (DNAV, used in [18]): the basic idea is to avoid new transmissions only toward the Angle of Arrival (AoA) of an RTS/CTS packet, which represents the bearing of the new sender/destination. Generally speaking, there will be as many countdown timers as forbidden directions. In [17] this concept is further extended, and a node refrains from transmitting only if it is aligned to *both* transmitter and receiver of the RTS/CTS handshake.

Finally, all these protocols are contention-based. We would like to remark that scheduling-based protocols have been designed as well, and one important example is [19]. This distributed scheduling is guaranteed to be collision-free, but comes at the price of a globally time synchronized network (for instance by means of GPS equipment).

2.2 Antenna Array Model

This section briefly outlines the potential of directional radiators. Such systems are usually made up by an array of N equal antennas[2] and employ beamforming techniques [4]. In transmission mode, the array elements are fed with versions of the same signal that are characterized by specific phase lags. In particular, let s be the signal to be delivered and $\mathbf{t}$ be the input vector for the array: thus $\mathbf{t} = s \cdot \mathbf{w}(\varphi)$, where $\mathbf{w}(\varphi) = [w_1(\varphi)\, w_2(\varphi) \cdots w_N(\varphi)]^T$,

[1] The problem has been solved in the neighboring context of MIMO ad hoc networks in [14], but the techniques used in there (Space Time Codes) are not directly applicable in this field.

[2] In most cases, antenna arrays are composed by half-wave dipoles, due to their simplicity and low production cost.

and $w_i(\varphi)$ are complex values lying on a circle of unit radius. A suitable choice of the coefficients,[3] exploiting constructive and destructive interference among transmitted fields, focuses most of the radiated energy in a specific bearing (i.e., φ), and reduces interference generated in other directions. As far as reception is concerned, on the other hand, if the outputs of the N antennas are weighed using the $\mathbf{w}(\varphi)$ vector, it is possible to obtain a directional reception pattern[4] in which the main lobe points toward φ and signals coming from other bearings are filtered out.

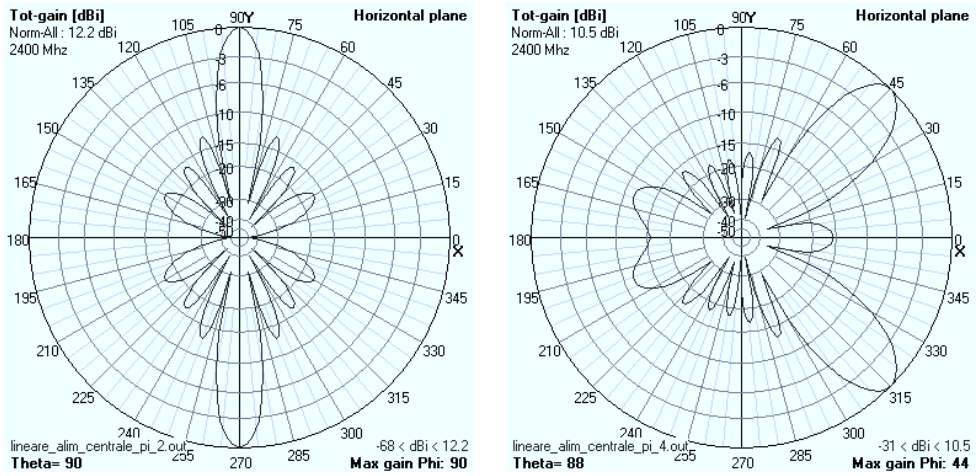
The usage of antenna arrays has many important consequences on the medium access, and, if properly managed, can lead to important performance enhancements. First of all, directional radiators present an extended transmission range (i.e., an increased gain), for a given input power, if compared to an omni-directional solution. Establishing longer links between nodes can lead to a more robust routing, with fewer hops per packet and thus quicker deliveries. Moreover, this feature reduces the problem of network fragmentation, simplifying, for instance, broadcast algorithms as well as communication between far away nodes.

Secondly, if directional transmissions and receptions are employed, it is possible to perform simultaneous communications among nodes that would be neighbors in an omni-directional scenario. Let us consider the situation depicted in Fig. 2.1, where $\boldsymbol{A}$ and $\boldsymbol{B}$ have packets to deliver to $\boldsymbol{C}$ and $\boldsymbol{D}$ respectively. Then, let us suppose node $\boldsymbol{A}$ to be the first to occupy the channel. If an omni-directional transmission were performed, $\boldsymbol{B}$, upon receiving this packet, would delay its communication with $\boldsymbol{D}$ to avoid a collision at node $\boldsymbol{C}$. If, instead, $\boldsymbol{A}$ and $\boldsymbol{B}$ reach their destinations directionally, both communications can be performed at the same time without any collision, as energy radiated in directions other than the desired one is negligible. Beamforming techniques, in conclusion, are able to introduce spatial multiplexing among users.

Multi-antenna systems, in the literature concerning medium access protocols, are typically modeled by the flat top or by the cone and sphere radiation pattern [7, 12, 20]. The former assumes that the antenna gain is constant within an angular region centered on the pointing direction and that no power is radiated elsewhere. The latter, instead, takes into account the interference generated out of the main lobe by means of a low constant gain. These models are rather simplistic and do not faithfully represent the behavior of a real an-

[3] As we shall discuss later, the values of the $\mathbf{w}$ vector as well as the obtained radiation pattern are strongly influenced by the array geometry.

[4] The directional pattern for a phased array is the same in transmission and reception, according to the reciprocity theorem [4].



(a) (b)

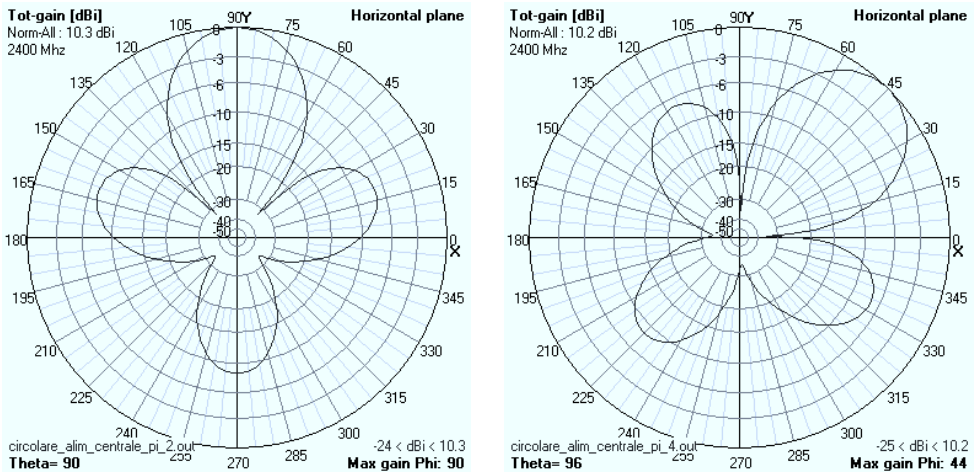
Figure 2.2. *Radiation pattern for an 8 half-wave dipoles linear array pointing toward a) $\pi/2$ and b) $\pi/4$. The antennas are spaced by $\lambda/2$.*

tenna array. As a consequence, their usage in simulations may only result in an approximate evaluation of the protocol performance.

According to these remarks, the approach that we follow starts from an accurate analysis of multi-antenna systems and tries to identify an effective yet realistic solution that can be used to faithfully evaluate the performance of the proposed medium access strategies.

Let us first consider a linear array whose N elements are spaced by $\lambda/2$ in order to limit the effects due to mutual coupling [4]. Fig. 2.2 presents some radiation diagrams obtained with $N = 8$ for different pointing bearings.[5] These patterns are characterized by the presence of two main lobes, because of the axial symmetry. In transmission mode, then, two beams are simultaneously covered, which generates significant interference out of the desired bearing. In reception mode, on the other hand, signals coming from different directions may not be distinguished. Fig. 2.2, furthermore, shows how for a linear array the gain (which is inversely proportional to the width of the main lobe) is strongly influenced by the pointing direction φ. For instance, the gain in Fig. 2.2a is 12.2 dBi, while it drops to 10.5 dBi in Fig. 2.2b and would decrease to as little as 9.4 dBi for an end-fire array. A node would then establish links of different length, depending on the location of its destination, potentially leading to unfairness in routing and medium access strategies. These remarks show how a linear geometry may not always be an effective choice, due to protocol design reasons.

[5]The conclusions of the analysis that we carry on in this section do not depend on the number of antennas in the array.



(a) (b)

Figure 2.3. *Radiation pattern for an 8 half-wave dipoles circular array pointing toward a) $\pi/2$ and b) $\pi/4$. The radius of the structure is equal to $\lambda/2$.*

Let us now, instead, consider an array whose $N = 8$ elements are arranged on a circle of radius $\lambda/2$ [21]. The radiation diagrams obtained with this solution are presented in Fig. 2.3. In this case, even though the peak gain is slightly lower (i.e., 10.2 dB instead of 12.2 dB), the drawbacks of linear arrays are not present, as the pattern is characterized by a single main lobe, whose width and gain are almost independent of the pointing bearing. The circular geometry clearly requires a bidimensional region where to displace the antennas. Nevertheless, if unidimensionality is not a constraint, this solution appears extremely simple yet effective to the aim of protocol design. According to this, throughout our work, we assume the presented circular array to be employed.

We now explain in further detail the reception model. We suppose that each antenna is connected to a dedicated RF chain. In this way we may simultaneously collect N versions of the incoming signal, $\mathbf{r} = [r_1\ r_2\ \dots\ r_N]^T$, characterized by phase lags due to the differences in the path lengths to the array elements. As explained earlier, if these copies are weighed using the $\mathbf{w}(\varphi)$ vector, a directional reception pattern pointed toward φ is obtained. Furthermore, let $\mathbf{W}$ be an $N \times M$ matrix whose columns are given by the sets of weights $\mathbf{w}(\varphi_i)$, $i = 1, \dots, M$. A single DSP can easily perform the product $\mathbf{r}^T \cdot \mathbf{W}$, obtaining a vector of received signals that covers with directional gain all the possible bearings.[6] This system is formally identical to a bank of CDMA-matched filters, although here the match is with respect to a spatial signature rather than to a spreading sequence. For the sake of simplicity, let us refer to a Line of Sight propagation environment (which is the one simulated in

[6] In this way the problem of *asymmetry in gain* at the receiver is solved.

our tests). Given that a packet arrives according to a certain AoA, the inner product of the received signal $\mathbf{r}$ and the matrix $\mathbf{W}$ will be maximum for a certain column $\mathbf{w}$, dependent on the AoA. Let us now consider two packets, whose AoAs are spaced by at least one lobe beamwidth: these frames will be detected by different filters. Each of them impinges in the secondary lobes of the filter that detects the other, causing some mutual interference (the larger the array, the smaller the secondary lobe and thus the interference). This makes it feasible to detect both packets[7] and can be regarded as a simple form of MultiUser Detection (MUD) [22]. As a consequence, it is possible for a node to listen to the channel even during an ongoing data reception. To the best of our knowledge, this is the first time that this usage of antenna arrays is applied in the context of ad hoc networks with directional antennas.

In conclusion, our survey has identified an antenna array which can provide the following advantages with respect to the existing literature:

- the problem of asymmetry in gain at the receiver is not present:[8] the described reception model allows to omni-directionally listen to the channel with directional gain;
- the array radiation pattern is almost constant with respect to the aiming direction: the statement is based on real patterns, while in the literature this is only an assumption based on the idealized flat top or cone and sphere models;
- the multi-packet decoding technique enables to listen to the channel while data reception is in progress;
- unlike the majority of the literature, we do take into account interference generated by secondary lobes, which can have a significant impact on the protocol performance and behavior.

2.3 The Problem of Deafness

Generally speaking, a terminal is said to be deaf if it does not answer an RTS message addressed to it. In an omni-directional scenario, this condition may typically occur because of a collision. For example, two terminals may simultaneously try to access the medium.

[7] In the most favorable case, the system can decode M packets simultaneously, providing that they come from different from different directions (i.e., they are received by different lobes). However, the comparison with CDMA systems suggest that this is possible only if the packets have comparable strengths, otherwise the near-far problem (one strong signal overpowers all the others) would make only one packet detectable.

[8] Asymmetry in gain in transmission is avoided by the protocols of the CMAC family by means of circular delivery of handshake frames, as explained in Section 2.4

The intended destination of the RTS packet, in this case, is not able to decode the frame due to the high level of interference, and cannot reply. The channel contention between the two transmitters, anyway, is solved by the exponential backoff procedure of the 802.11 MAC layer [5], and deafness does not significantly affect the overall network performance.

On the contrary, the impact of this issue becomes more relevant if antenna arrays are employed. Directional transmissions, in fact, enable a spatial reuse of the channel, as simultaneous communications can be established within a set of neighbors [8, 15]. Nevertheless, this result can be only achieved if a terminal is allowed by the MAC layer to access the medium (i.e., to send an RTS frame) even if other neighboring links are active,[9] unless specific alignment conditions are met ([17], Section 2.5). Such an approach, as a drawback, makes it possible for a node to try to establish a connection with a busy neighbor. Let us refer, in particular, to the situation depicted in Fig. 2.1, and suppose that node $\boldsymbol{A}$ sends an RTS message to $\boldsymbol{B}$, which is actually already involved in an ongoing communication with $\boldsymbol{D}$ (i.e., it is sending data). In this case node $\boldsymbol{B}$ is deaf to the RTS (because of its own transmission), and $\boldsymbol{A}$ does not receive any reply. According to the 802.11 policy, node $\boldsymbol{A}$, with the erroneous belief that a contention with another terminal is taking place, starts a backoff procedure, listening to the channel for a random time. At the end of this period, $\boldsymbol{A}$ tries to contact $\boldsymbol{B}$ again, re-transmitting the RTS. Anyway, the length of the backoff window is typically shorter than the time required for $\boldsymbol{B}$ to return idle, and thus $\boldsymbol{A}$'s attempt is unlikely to be successful. As a result, a terminal which contacts a deaf destination may spend a lot of time in useless backoff procedures, delaying other possible parallel data communications and affecting the efficiency of its medium access strategy. Moreover, had the sender ($\boldsymbol{A}$) known that the intended destination ($\boldsymbol{B}$) was busy and had it had a packet buffered for a reachable node (say $\boldsymbol{E}$), $\boldsymbol{A}$ may have tried to contact $\boldsymbol{E}$ instead, thus improving packet delivery latency. Of course this comes at the price of a queueing manager, but the entailed software complexity cost is usually limited. Secondly, the transmission of multiple RTS frames involved by iterated backoff cycles increases the overall interference, potentially damaging other ongoing links. In the worst case, all the attempts to contact the destination may not succeed. In this condition, the initiating node would report a link failure message to the higher layers [5], falsely believing that the addressee has moved out of range. This information would cause the higher layers to start expensive (in terms of overhead and complexity) and pointless route-recovery procedures [23], significantly affecting the network performance.

Finally, we remark that while all the previous causes of deafness are deeply rooted in

[9] We remark that the IEEE 802.11 omnidirectional virtual carrier sense policy would not permit nodes to send any packet if a handshake frame has been recently overheard.

any directional antenna system, a suboptimal usage of these arrays may create additional problems. Let us consider, for instance, a terminal that is performing a backoff procedure. According to the IEEE 802.11 DCF, the node omni-directionally listens to the medium, and is thus able to perceive all the signaling on the channel. Instead, some authors who employ smart antennas, [8,15], require the terminal to only perform a directional reception, steering its lobe toward the addressee of its request. On the contrary, had the pattern been kept omni-directional (e.g., using the model of Section 2.1), such a problem would not have occurred. In this way, though, packets coming from directions out of its reception pattern would not be detected. As a reference, consider again Fig. 2.1. Node $\boldsymbol{A}$, according to [15], during the backoff procedure due to $\boldsymbol{B}$'s unavailability, stays beamformed toward $\boldsymbol{B}$ and actually becomes itself deaf to any request. If in the meantime another terminal, say $\boldsymbol{E}$, tries to contact it, the frame is not detected and no reply from $\boldsymbol{A}$ is obtained. Then $\boldsymbol{E}$ enters a deafness-induced backoff as well. The chain of nodes stuck in useless attempts may be further extended, potentially leading to a severe degradation of the network performance.

2.4 CMAC: a Class of Cooperation-Based Protocols

In the development of our protocols a lot of attention has been focused on the problem of deafness, due to the impact of this issue on network performance. The analysis of Section 2.3 has shown how in a directional scenario the lack of a CTS is mainly due to the erroneous forwarding of packets addressed to destinations which are actually busy. CMAC tries to avoid this situation by letting each node have an accurate perception of the surrounding network activity. To this aim, every terminal has an internal table, called *Communication Register*, or CR, where all the known ongoing communications are reported. For each of them, the nodes involved, the time left before conclusion[10] and the pointing direction needed to reach the data transmitter are stored. The last quantity is estimated from the AoA[11] of the RTS frame, while the other fields are inferred from RTS, CTS or ACK packets (see Section 2.9): every time any of these frames is received, a new entry is created in this table, if necessary. When a frame has to be delivered, the terminal checks its CR in order to determine whether the desired destination is already involved in another activity, preventing useless attempts. The more updated the CR, the more effective this strategy. According to this principle, the

[10]The *time left* field is suitably decreased by the node. When the counter has reached zero, the registered communication has come to an end, and the CR entry is removed.

[11]The usage of smart antennas allows to perform algorithms that determine the Angle of Arrival for one or more impinging signals, [21,24,25].

CMAC protocol employs three solutions to let the nodes have a fair image of their neighbors' state: A) circular delivery of RTS and CTS, as in [12, 13]; B) use of the reception model described in Section 2.1 that, although well known in the physical layer community, has not been incorporated in MAC protocols so far; C) cooperation among nodes, which is actually a new proposal.

A) Circular Handshaking. RTS and CTS packets in CMAC are sent in a circular fashion, resembling the strategy proposed in [13]. These frames are broadcast omni-directionally by means of a sequence of directional transmissions, suitably spaced in order to cover the whole horizon. In particular, if θ_{BW} is the main lobe beamwidth, this can be obtained by beamforming toward φ_1, $\varphi_2 = \varphi_1 + \theta_{BW}$, ..., $\varphi_M = \varphi_1 + (M-1) \cdot \theta_{BW}$, where $M = \lceil 2\pi/\theta_{BW} \rceil$. This approach introduces some overhead with respect to protocols that perform simple directional handshaking [7, 8], but nevertheless it turns out to be fundamental in reducing deafness, as all the neighboring nodes that are listening to the channel are informed about the link being established. Section 2.7 will show how the advantages offered by an omni-directional delivery of handshake frames outweigh the drawbacks and contribute to the overall CMAC performance improvements.

B) Multiple Receptions. The reception model described in Section 2.2 allows a node to simultaneously decode up to M packets, provided that they come from different angular sectors. In our protocol, then, unlike the solutions proposed in the literature [8, 12, 13, 15], a terminal is able to listen to the channel and to update its CR even when involved in the reception of a frame. The joint use of circular handshaking and multiple receptions leads to a significant reduction of deafness (see Section 2.7): when two terminals, say $\boldsymbol{A}$ and $\boldsymbol{B}$, start a channel negotiation, all the neighbors which are idle or in reception mode become aware of the link being established. Hence, transmissions toward $\boldsymbol{A}$ and $\boldsymbol{B}$ are avoided for the period of their communication.

C) Cooperation among nodes. Circular handshaking and multiple receptions, as explained, protect nodes that are listening to the medium at the moment of a channel negotiation, but are not able to completely solve the problem of deafness. If frequency duplexing is not employed, in fact, simultaneous transmissions and receptions are not possible, and a terminal that is sending a frame actually becomes deaf. Suppose that, referring to the topology of Fig. 2.1, $\boldsymbol{A}$ successfully negotiates the channel and starts transmitting a data packet to $\boldsymbol{E}$. Moreover, assume that during this communication a link between $\boldsymbol{B}$ and $\boldsymbol{D}$ is established. In this case, $\boldsymbol{F}$, supposed idle, and $\boldsymbol{E}$ (i.e., nodes that are listening to the channel) correctly update their CR and become aware of $\boldsymbol{B} \rightarrow \boldsymbol{D}$ activity. On the other hand, terminal $\boldsymbol{A}$

cannot decode RTS or CTS for this transmission. It is possible, then, that at the end of its communication with $\boldsymbol{E}$, $\boldsymbol{A}$ erroneously tries to contact either $\boldsymbol{B}$ or $\boldsymbol{D}$ but fails because of deafness.

To address this issue, CMAC proposes a cooperation scheme among nodes, trying to inform terminals that go back to idle mode after a transmission about communications that have started in the meantime. Let us refer again to Fig. 2.1. Nodes informed about both $\boldsymbol{A} \rightarrow \boldsymbol{E}$ and $\boldsymbol{B} \rightarrow \boldsymbol{D}$ links are aware that the second transmission has been established while $\boldsymbol{A}$ was deaf. Therefore, they know that $\boldsymbol{A}$ has an outdated image of the surrounding activity. In our situation this is the case for terminals $\boldsymbol{E}$, $\boldsymbol{F}$ and $\boldsymbol{C}$. At the end of $\boldsymbol{A}$'s activity (i.e., after the potential reception of an ACK packet from $\boldsymbol{D}$), these nodes send to $\boldsymbol{A}$ a directional cooperation frame containing information about the ongoing link involving $\boldsymbol{B}$ and $\boldsymbol{D}$. In this way, $\boldsymbol{A}$ is able to update its CR and to prevent communication attempts with busy neighbors. A detailed description of the collaboration algorithm will be given at the end of this section.

The proposed cooperation method is robust to mobility. Let us refer to Fig.2.1 and let us focus on $\boldsymbol{A} \rightarrow \boldsymbol{E}$ data exchange. Node $\boldsymbol{E}$ can beamform the cooperation packet very accurately because it has perfect knowledge of $\boldsymbol{A}$'s location (it has just received a data packet from it). The other idle nodes in the network that decide to cooperate, estimated $\boldsymbol{A}$'s bearing when they received $\boldsymbol{A}$'s RTS, which was sent shortly before the transmission of the data packet. Given our data rates, a data packet and its handshake last for about 11ms. Even at relatively high vehicular speeds (120 km/h = 75 mi/h) the maximum displacement is moderate (80 cm = 2ft 8in), and unlikely to significantly affect the nodes relative position. Therefore the cached AoA can be considered as reliable and the cooperation packet will be properly delivered.

The additional signaling introduced by cooperation frames increases the overall interference. Nonetheless, our simulations show that there is a net gain in network performance (see Section 2.7). These benefits are due to two main reasons: firstly, the cooperation packets are short and directionally sent (unlike RTS/CTS), thus limiting interference due to this mechanism. Secondly, a more updated CR may prevent useless and harmful handshake frames sent to nodes otherwise reputed as deaf.

Thus we remark that due to the finite size of a collaboration packet, only a limited number of unknown communications can be reported. This implies that some new transmissions might still be ignored by the node that receives these frames.

From the above discussion, we may conclude that cooperation has three main limits: a)

the amount of information that can be delivered is finite, b) some interference is generated and c) some overhead is introduced in order to deliver collaboration frames.

Let us consider the class of CMAC protocols whose cooperation packet size is no larger than x bits ($x = 160$ bits in our simulations). Once the overhead is bounded, the efficiency of the protocol is proportional to the number of new communications that can be conveyed in a collaboration frame. This quantity may vary between 1 and k, where k is the number of links the packet recipient is unaware of.

In order to evaluate the efficiency of this family of CMAC schemes, we have simulated two systems, called L-CMAC and B-CMAC, that we compare for a fixed value of the overhead (i.e., drawback c) has the same impact on both the protocols). The former represents an ideal solution that fully exploits the potential of the collaboration mechanism, and may be regarded as an upper bound for the CMAC protocol family. According to B-CMAC, at the end of each communication the transmitter is informed (at no cost in time or energy) of all the k other links that have been established in the meantime. This simulates the presence of an idle node that transmits a cooperation packet carrying data about all these new communications but generating no interference, i.e., drawbacks a) and b) are not taken into account.

On the other hand, the L-CMAC protocol provides a basic implementation of CMAC structure. In this case, idle nodes actually send collaboration frames and thus generate interference (drawback b)). Moreover, these packets are very short and contain information concerning just one out of k new ongoing communications. In this way, the impact of drawback a) has the maximum impact. It is thus clear, according to the previous discussion, that L-CMAC represents a lower bound to the performance of the CMAC protocols.

Next, we give a brief description of the cooperation procedure proposed in L-CMAC.

IDLE NODE

If *Time Left* field of a CR entry (entry #1) expires and the RTS for that communication has been received (i.e., the bearing of the transmitter has been cached at the beginning of the data exchange):

A) Check if there is another entry (entry #2) registered after entry #1. If so:

   A1) Create a cooperation packet containing ids of source and destination and time left of entry #2.

   A2) Get the bearing of the data transmitter of the frame registered in entry #1. Beamform towards this node and send the cooperation packet.

RECEIVER NODE[12]

If the data reception is successfully completed:

A) Beamform toward the transmitter and send an ACK packet.

B) Check the CR to determine whether any entry has been scheduled during the data reception. If so:

B1) Create a cooperation packet containing ids of source and destination and time left of the entry.

B2) Beamform toward the transmitter of the received data frame and send the cooperation packet.

TRANSMITTER NODE

After the reception of the ACK frame:

A) Listen to the channel waiting for cooperation packets. If any is received:

A1) Update the CR exploiting information contained in the cooperation frame.

B) Go back to idle mode.

2.5 Collision Avoidance in CMAC

As discussed in the introduction of this chapter, in order to exploit the potential of smart antennas and to obtain spatial multiplexing, it is extremely important to both enhance coordination among nodes (limiting deafness) and to cope with the problem of collisions. Generally speaking, a collision occurs each time a terminal tries to access a channel that is already in use for a communication, generating an amount of interference that prevents a correct reception for the data packet under exchange. Protocols designed for omnidirectional antennas (such as the IEEE 802.11) often rely on the virtual carrier sense mechanism. While this approach abates the impact of collisions, it is also exceedingly conservative for multi antenna systems, and it would stymie much of the available parallelism among communications. As explained in Section 2.2, a different behavior is required in order to achieve spatial multiplexing. Refer to the scenario of Fig. 2.1 and let us focus our attention on node $\boldsymbol{D}$. This terminal is not in line with the $\boldsymbol{A}$-$\boldsymbol{C}$ bearing and thus cannot generate a collision, as it is not able to access the portion of the channel used for the ongoing data exchange. Any directional transmission performed by $\boldsymbol{D}$, in fact, does not interfere with the main lobe

[12]With *Receiver Node* we refer to a node that is receiving a data packet, and with *Transmitter Node* to a terminal that is transmitting a data frame.

used in reception[13] by $\boldsymbol{A}$ and $\boldsymbol{C}$, but instead it would hit one of their secondary lobes, thus generating a small amount of interference. Let us furthermore consider an aligned terminal ($\boldsymbol{G}$). Once again, it is not necessary for $\boldsymbol{G}$ to delay potential communications, unless the main lobes used to deliver the messages cover the common direction under which $\boldsymbol{A}$ and $\boldsymbol{C}$ are seen.

In a directional scenario, then, a node can determine its medium access policy as soon as it understands whether it is aligned with a pair of terminals that are communicating. If this happens, the node has to delay any transmission that involves potentially harmful bearings. This procedure can be easily accomplished once the angular positions of all the terminals involved in the channel contention are known. If the higher layers do not provide these data by means of location tables or GPS, they have to be determined every time a potential alignment condition has to be evaluated. To this aim, the protocols of the CMAC family rely on the exchange of handshake frames: when a node decodes an RTS or a CTS, it performs an AoA estimation, computing the bearing under which the transmitter is seen. Referring to Fig. 2.1, suppose that terminal $\boldsymbol{G}$ overhears both the channel negotiation frames for the $\boldsymbol{A} \rightarrow \boldsymbol{C}$ link. Recalling that the asymmetry in gain has not to be taken into account, $\boldsymbol{G}$ is aware of being inside the coverage range of $\boldsymbol{A}$ and $\boldsymbol{C}$ and thus that it could potentially collide with their communication. In this case, the algorithm described in detail in Section 2.6, exploiting the angular positions obtained by means of AoA estimation, determines whether $\boldsymbol{G}$ is in line with $\boldsymbol{A}$ and $\boldsymbol{C}$. If this happens, the node updates its DNAV, and it avoids any transmission that overlaps the alignment bearing for the period of time reserved for the ongoing communication.

In the end, if both negotiation messages are received, a fair medium access is assured in CMAC, but it is important to consider the possibility of a single handshake packet reception. Let us focus, in particular, on the case in which node $\boldsymbol{G}$ only decodes the CTS frame for the $\boldsymbol{A} \rightarrow \boldsymbol{C}$ transmission. This may occur due to three main reasons: 1) $\boldsymbol{G}$ could be outside $\boldsymbol{A}$'s transmission range, 2) the RTS could not be sent toward $\boldsymbol{G}$, and 3) a collision at node $\boldsymbol{G}$ may prevent a correct reception. Let us analyze the first two conditions in further detail.

1) *$\boldsymbol{G}$ is outside $\boldsymbol{A}$'s coverage range.* Observing that handshake frames are processed with maximum transmission power (as they are supposed to inform as many neighbors as possible), it is not possible for $\boldsymbol{G}$ to collide at node $\boldsymbol{A}$. More precisely, if $\boldsymbol{G}$ started a transmission covering the $\boldsymbol{A}$-$\boldsymbol{C}$ bearing, $\boldsymbol{A}$ could not receive the message, but would still perceive a certain amount of interference that nevertheless is likely to be tolerable due to the possibly

[13] During the exchange of a data frame, the transmitter and the receiver steer their patterns toward each other.

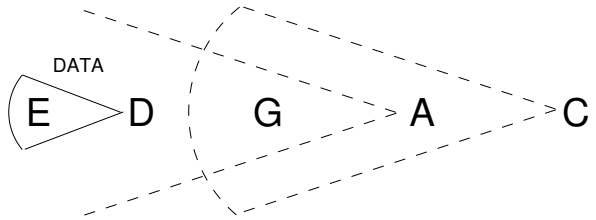


Figure 2.4. *Critical topology for the alignment algorithm.*

high path loss. The alignment detection procedure and the discussed case in particular assume, for the sake of simplicity, that the communication and the interference ranges are equal. Nonetheless, this simplifying assumption is removed in our simulations, where we employ a realistic radiation pattern and carefully evaluate the actual interference at each receiver, thereby testing the effectiveness of our collision avoidance algorithm. under general conditions.

2) *RTS is not transmitted toward* $\boldsymbol{G}$. Let us refer to the topology of Fig. 2.4 and suppose that node $\boldsymbol{D}$ negotiates the channel to communicate with $\boldsymbol{E}$. $\boldsymbol{A}$, upon the reception of RTS and CTS, updates its DNAV, and avoids transmissions involving $\boldsymbol{D}$-$\boldsymbol{E}$ bearing. Suppose furthermore that, while $\boldsymbol{D} \rightarrow \boldsymbol{E}$ is still active, $\boldsymbol{A}$ has a packet to deliver to $\boldsymbol{B}$. The node starts sending a circular RTS, but does not cover the critical direction, according to DNAV constraints. In this way, though, $\boldsymbol{G}$ does not receive the handshake packet, and is able neither to estimate $\boldsymbol{A}$'s location nor to perform the alignment algorithm and to update its DNAV. If a transmission that overlaps $\boldsymbol{A}$-$\boldsymbol{B}$ bearing is started, a collision may take place.

This brief analysis shows that some critical conditions may occur if a node receives a single negotiation frame, but it is important to stress that in these cases the terminal is by no means able to determine whether it is aligned with an ongoing communication. The protocols of the CMAC family follow an aggressive strategy, and potentially update the DNAV only if both RTS and CTS are successfully decoded, taking as a reference condition 1). There are two main reasons lying at the basis of this choice. First of all, scenario 1) is more likely to happen than 2) and 3). This may be explained by observing that all the cases require alignment of three or more nodes, yet 2) and 3) also need some form of synchronization among communications. Secondly, a more conservative approach could lead to some inefficiencies. Refer to Fig. 2.4 and suppose that terminal $\boldsymbol{G}$, having received only a CTS coming from $\boldsymbol{C}$, prevents any transmission toward φ_C bearing (i.e., the angle under which $\boldsymbol{G}$ sees $\boldsymbol{C}$). This choice would guarantee collision avoidance for the event sequence of scenario 2), but would also worsen the network performance for all the topologies in which $\boldsymbol{G}$ is not actually aligned with $\boldsymbol{A}$ and $\boldsymbol{C}$ (condition that is more likely to be met). An example is depicted

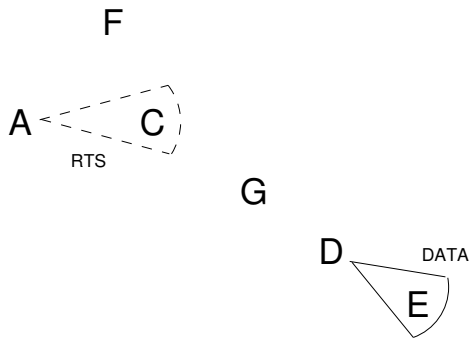


Figure 2.5. *Critical topology for a conservative alignment approach.*

in Fig. 2.5. Here, as in Fig. 2.4, node $\boldsymbol{A}$ and $\boldsymbol{C}$ start their circular handshake. However, $\boldsymbol{G}$ receives only $\boldsymbol{C}$'s CTS but not $\boldsymbol{A}$'s RTS since nodes $\boldsymbol{A}$-$\boldsymbol{G}$-$\boldsymbol{D}$-$\boldsymbol{E}$ are in line. Nonetheless, $\boldsymbol{G}$ has no reason to delay potential transmissions with any node $\boldsymbol{X}$ (e.g., in $\boldsymbol{F}$ in our topology) such that $\varphi_X \simeq \varphi_C$.

2.6 CMAC Alignment Algorithm

Let us consider the topology depicted in Fig. 2.1, and suppose that the channel negotiation for a $\boldsymbol{A} \rightarrow \boldsymbol{C}$ communication has been performed. Nodes $\boldsymbol{B}$ and $\boldsymbol{G}$, upon the reception of both RTS and CTS for this link has to decide whether to update their DNAV in order to avoid collisions.

Let $\theta_x(y)$ be the angle under which node x sees node y with respect to a common reference direction shared by all the terminals in the network.[14] Let, furthermore, θ_{BW} be the beamwidth of an angular sector for directional transmissions and receptions. Following this notation, node $\boldsymbol{X}$ (with $\boldsymbol{X}$ being a node that receives these packets, such as $\boldsymbol{B}$ or $\boldsymbol{G}$) estimates $\theta_X(\boldsymbol{A})$ and $\theta_X(\boldsymbol{C})$ from the reception of RTS and CTS. Moreover, $\boldsymbol{X}$ is aware of the value of $\theta_A(\boldsymbol{C})$, as it is included in the CTS packet (see Section 2.9). According to this, it is easy to prove that transmissions covering $\theta_X(\boldsymbol{A})$ and $\theta_X(\boldsymbol{C})$ have to be delayed according to the following algorithm:

$$\Big|\theta_C(\boldsymbol{A}) - \theta_X(\boldsymbol{A})\Big| \leq \frac{\theta_{BW}}{2} \;\Rightarrow \text{ delay tx toward } \boldsymbol{A}$$
$$\Big|\pi + \theta_C(\boldsymbol{A}) - \theta_X(\boldsymbol{C})\Big| \leq \frac{\theta_{BW}}{2} \;\Rightarrow \text{ delay tx toward } \boldsymbol{C}$$

[14] The assumption of a common reference bearing could be removed at the cost of a complexity increase for the alignment algorithm. For further details, refer to [26].

The proposed collision avoidance strategy is strongly influenced by the value of the beamwidth of an angular sector: the wider θ_{BW}, the larger the number of transmissions delayed due to DNAV constraints. A conservative approach, for instance, could take as a reference the angle between two consecutive null directions, limiting (or potentially avoiding) collisions yet reducing the number of communication attempts. Nevertheless, it should be noticed that the maximum transmission and reception gains are available during a data exchange, and they provides a certain amount of interference margin. Then, a more aggressive strategy could lead to a performance enhancement, allowing transmissions that may generate a small level of interference. The protocols of the CMAC family follow this principle, using a 3-dB definition for the beamwidth of an angular sector. The next Section will show that such a policy leads to higher performance.

2.7 Simulation Results

We have compared the CRCM protocol with the bounds on the cooperation performance described in Section 2.4, L-CMAC and B-CMAC respectively, by means of extensive simulations. In order to evaluate the impact of collaboration over multiple receptions on the network behavior, we have also tested a protocol called MUD: this is a simplified version of L-CMAC, which employs the antenna model described in Section 2.2 but does not rely on the idea of collaboration; that is, it only employs strategies A and B described in Section 2.4. All these systems have been evaluated by means of the OPNET modeler (version 11.5) [27]. The network was made up by 10 or 15-nodes, uniformly and independently located inside a 1500 m $\times$ 1000 m rectangle. The results have been averaged over 30 independent simulations for each network size and packet generation rate. Each simulation lasts 30 s. This duration has been chosen long enough to stabilize the results.

Every node produces Poisson traffic at nominal rates of 8, 16, 25, 50, 100, 150 pk/s. The maximum arrival intensity has been chosen high enough to reach saturation in all the protocols. There are five types of packets (Data, ACK/NACK, RTS, CTS and collaboration), whose sizes are equal to 10240, 120, 240, 250 and 160 bits, respectively, in agreement with the IEEE 802.11 specifications [5]. In addition, the bandwidth is equal to 1 MHz, the modulation is BPSK and the final data rate is 1 Mbit/s.

Each terminal is equipped with a circular array composed by $N = 8$ antennas. Since this array presents a 3-dB beamdwidth of 44.1^o, we need $M = 9$ sectors to cover the whole horizon. The antenna patterns used for our protocols emulate the realistic behavior depicted

in Fig. 2.3. On the other hand, CRCM has been simulated employing a flat top model, as proposed in [12].

The propagation environment is purely Line of Sight and we have calibrated the transceiver parameters with a 10% target packet error rate. The maximum transmission power is equal to 1 mW, which enables two nodes located at the opposite corners of the deployment area to communicate with each other provided that their main lobes are steered to maximize the gain. The simulated network is thus single hop and no routing is needed. This choice is meant to highlight the efficiency of the analyzed protocols. The medium access coordination, together with the capability of guaranteeing spatial reuse, in fact, are deeply tested only if all the nodes may simultaneously contend for the channel. Moreover, single hop scenarios are more frequent in directional antenna networks because of the increased communication range and higher connectivity. Finally, the single hop nature of our networks implies that there are no hidden nodes. Therefore, all the terminals are in visibility and may heavily interfere with each other. This highlights how the number of nodes in the network (up to 15) is large enough to stress the protocols. Nonetheless, the impact of the solutions that we propose for multihop environments and routing is an interesting problem and is part of our future work.

All the results for the 10-node network are reported in this chapter, whereas only the most significant graphs for a 15-node configuration are shown due to space constraints and similarity with the 10 terminal case.

The first two metrics under study are the number of link failures and the number of packets addressed to deaf nodes. These have been chosen because they are directly related to the concept of deafness. Let us consider Fig. 2.1: given that $\boldsymbol{A}$ tries to contact $\boldsymbol{B}$ (and $\boldsymbol{B}$ is deaf), the latter metric is increased in either of these two situations: 1) all backoff cycles are unsuccessfully undertaken, or 2) $\boldsymbol{A}$ understands that $\boldsymbol{B}$ is actually busy and consequently aborts its attempt. The link failure metric on the contrary accounts only for condition 1). Case 2) happens if the source, while waiting for a CTS reply, overhears a packet (i.e., an RTS, CTS or ACK) coming from or addressed to its destination. Since this is a sign that $\boldsymbol{B}$ is actually busy as it is engaged in some communication, $\boldsymbol{A}$ delays the packet delivery. Therefore $\boldsymbol{A}$ is aware that the link is still active and will not report a link failure to the higher layer. The successful transmission of the packet is probably just postponed, not impossible as the link failure would suggest. In both cases, the impairments are due to deafness.

The results for the percentage of packets sent to deaf nodes for a 10 user network are reported in Fig. 2.6. This metric is depicted against the nominal load (generation rate of the

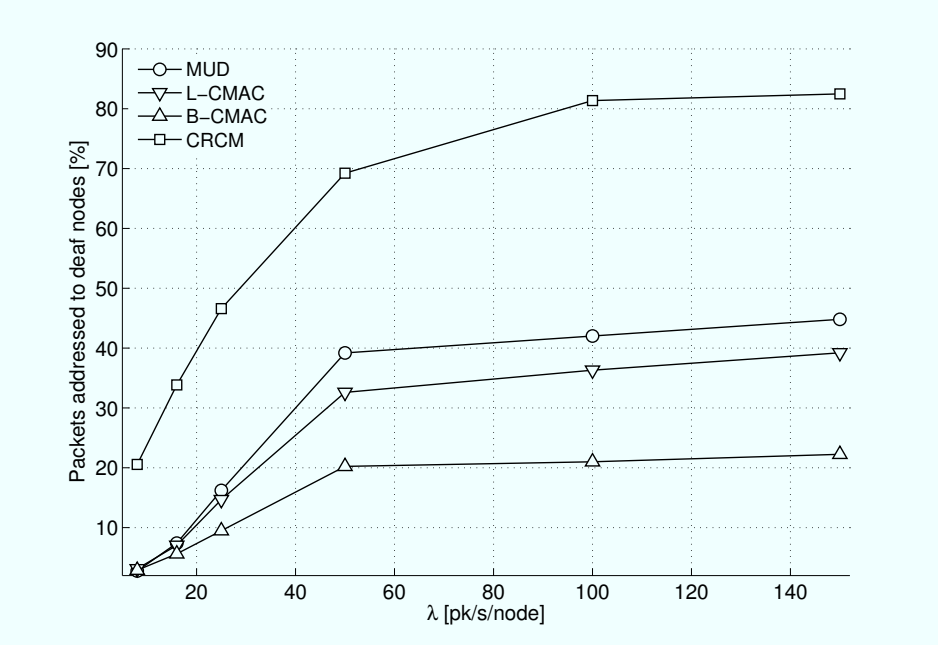


Figure 2.6. *Deafness vs nominal load for a 10-node network*

Poisson process). First of all, we note that the CRCM protocol addresses the majority of its packets to nodes that are unable to reply. This can be explained observing that the registers do not faithfully represent the neighbors' status because they cannot be updated during the exchange of data packets (i.e., for [13] and [12], a node cannot listen to the channel while transmitting or receiving data). Since these registers do not provide reliable information on the surrounding network activity, the virtual carrier sense no longer works. Instead, all the proposed systems significantly abate this phenomenon by a factor of at least two. Firstly, the simple MUD scheme is able to halve the extent of the problem, since the receiving nodes can continue decoding other packets. The possibility to receive control packets (such as RTS/CTS) is particularly beneficial as they convey the status of the network. Moreover, an effective usage of the cooperation mechanism (see B-CMAC curve) may further halve this metric, reducing it to almost 20%. This bound proves that the proposed technique may significantly lessen the impact of deafness. In addition, even in the worst case (L-CMAC) cooperation is able to achieve some improvement over MUD, and the presence of deafness has been strongly diminished with respect to CRCM. Our research is currently looking for a mechanism that may attain a bigger share of the untapped performance improvements.

Fig. 2.7 shows the behavior of this metric in a 15-node network. The trends are similar, giving further foundation to our previous comments. In addition, all the protocols show some performance loss because of the higher interference generated by the larger network.

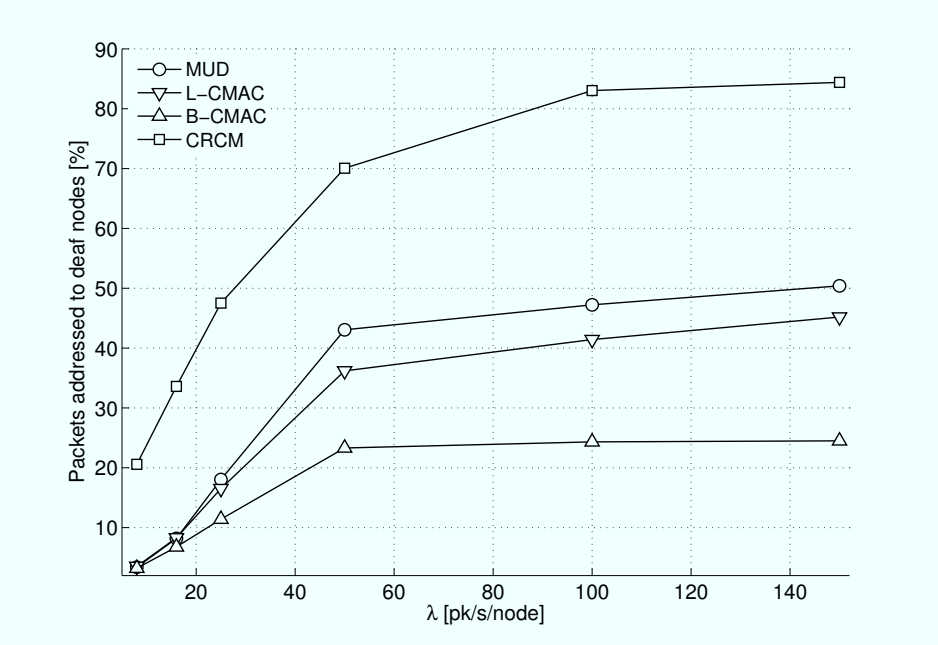


Figure 2.7. *Deafness vs nominal load for a 15-node network*

The more numerous collisions beset the proposed solutions, because the increased level of interference affects the signaling packets (i.e., RTS, CTS). These frames are then corrupted more frequently, and the impact of deafness and link failures is raised. Incidentally, we notice that B-CMAC loses only 2-3% with respect to Fig. 2.6: this stresses that a good implementation of cooperation schemes is more robust to interference and collisions.

Tightly correlated with the number of packets addressed to deaf nodes is the number of link failures. As explained, such an event happens when a node tries to contact another terminal but all the backoff cycles fail. Therefore link failures are actually special (and detrimental) cases of packets for deaf destinations. Let us remark that our nodes are stationary and the propagation environment is not affected by shadowing. Link failures, then, cannot be ascribed to mobility, but only to protocol inefficiencies. The main reason for this phenomenon is deafness: it is highly likely that a destination is unreachable in all these attempts because it is engaged in one or more data exchanges, hence this node should have been regarded as unavailable by its neighbors. The simulation outcome for a 10 user scenario is reported in Fig. 2.8, where the percentage of transmission attempts that lead to a link failure are shown against the nominal load per node. The MUD system already brings a remarkable improvement over CRCM, since link failures are cut by a factor of nearly three. In addition, if cooperation is exploited to the utmost (B-CMAC), link failures are reduced to as little as 15%, one sixth of the CRCM value. This fact highlights again the potential of this

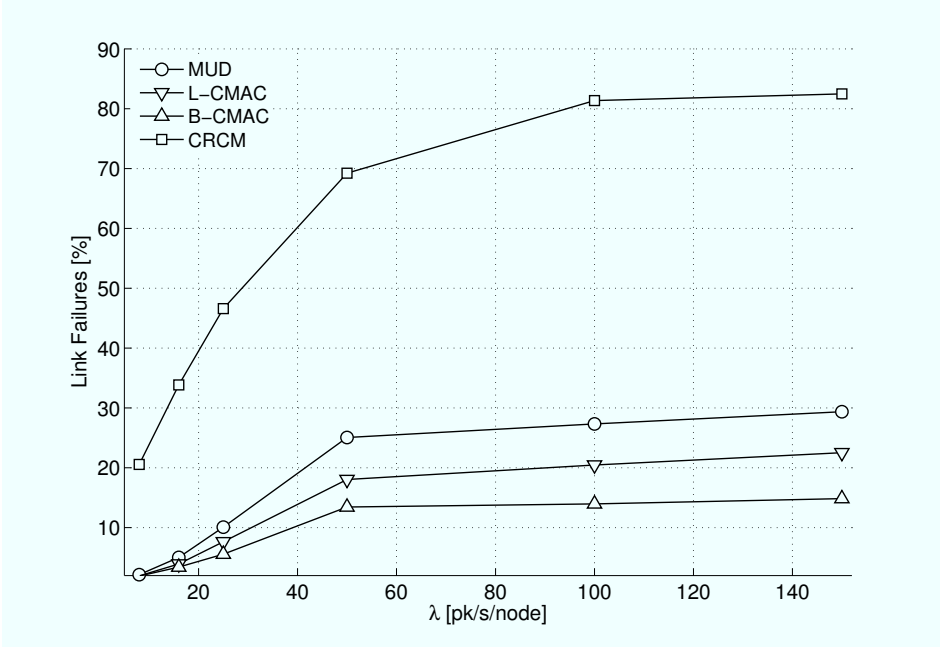


Figure 2.8. *Link failures vs nominal load for a 10-node network*

technique, and even in the worst case (L-CMAC) cooperation achieves a link failure rate of 25%, which is better than MUD alone. We want to point out that the improvement of L-CMAC over MUD in this context is the same as in Fig. 2.6 (which analyzed the number of packets addressed to deaf nodes). That is to say, these two metrics are reduced by the same quantity when cooperation is included. This could have been expected, and the reason is the following: by the above discussion and bearing in mind that our networks are static, it turns out that a link failure may happen only when a node tries to contact a terminal that is deaf throughout the duration of the backoff cycles. Thus, every link failure always stems from a communication with a deaf terminal and any method (like cooperation) that prevents to send packets to deaf nodes also avoids potential link failures. However, collaboration packets are sent only at the end of a data exchange and if afterward an RTS for a deaf node is transmitted, no further help can be provided by this solution. Therefore, in this condition, the behavior of MUD and L-CMAC protocols is the same (i.e., the node may either succeed in its backoff procedure, recover from it or eventually incur a link failure). This confirms our view that cooperation (at least as we have implemented it) is a proactive rather than reactive mechanism, because it is aimed to prevent these problems, but not to solve them after their appearance.

Finally, note that all the curves change slope between 50 and 100 pk/s. At the latter packet generation rate, the offered load is at least $100 \cdot 10240$ bit/node/s. Since the data rate

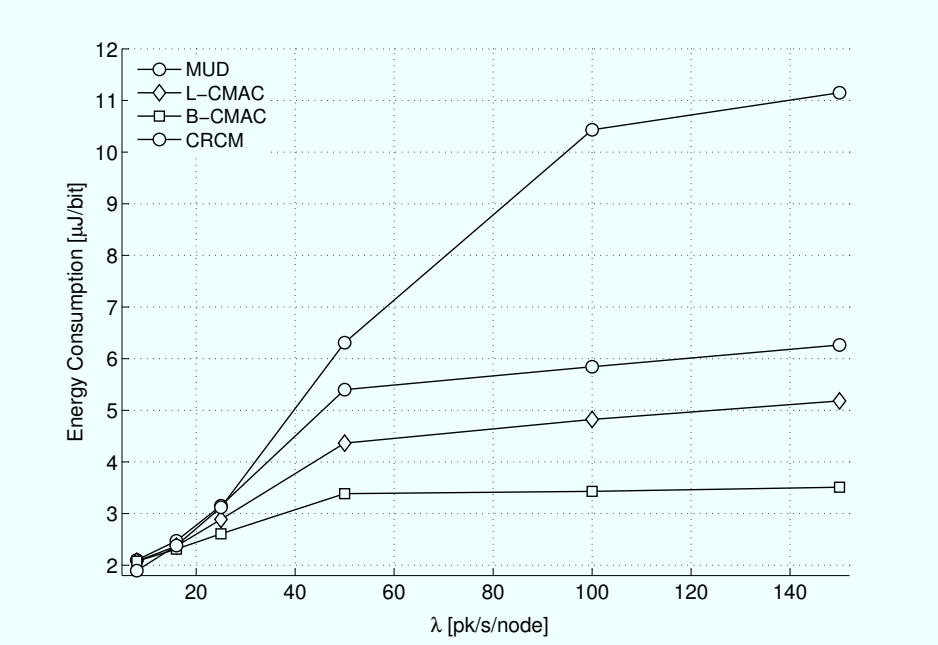


Figure 2.9. *Energy consumption vs nominal load for a 10-node network*

is 1 Mbit/s, the bandwidth is too small to support the traffic and congestion is unavoidable. Our protocols degrade more gracefully than CRCM in these situations.

The third metric under study is transmit energy consumption, plotted against the nominal load per node, which is a vital issue for ad hoc networks (that are often battery-constrained). It has been computed as the total energy used by the nodes in transmission divided by the number of bits successfully sent across the channel. Congestion is one of the factors that affect energy consumption, and we expect that our enhancements will provide a significant boost, since they are able to coordinate the transmissions more efficiently and to reduce packets sent to deaf nodes. The simulation results in the case of a 10 user network (Fig. 2.9) prove that savings of up to 52% are achievable with the simple MUD. The inclusion of cooperation packets may reduce the energy consumption between 58% (L-CMAC) and 67% (B-CMAC). This last figure shows that an efficient usage of cooperation can spare two thirds of the transmit energy needed by [13]. We remark that the better network activity perception implied by the collaboration approach offsets the power drawn by the increased signaling. This is due to the fact that an updated CR reduces the number of packets sent to deaf nodes and therefore useless backoff cycles with the subsequent RTS packets, leaving more time to send data packets or just reducing interference and power consumption. These facts show that the proposed improvements can provide extremely interesting savings at a limited cost. It is acknowledged that these simulations do not include processing consumption, but the

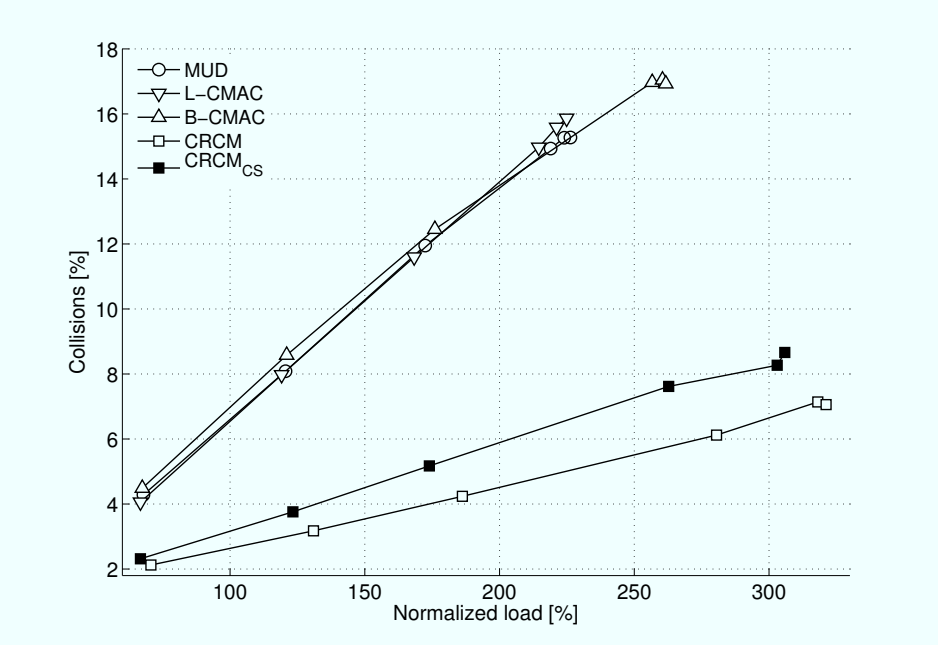


Figure 2.10. *Percentage of collisions vs normalized load for a 10-node network*

MUD scheme under use is extremely simple (it is actually a bank of matched filters), and the collaboration power requires only to examine the communication registers, which are lookup tables. Therefore, we do not expect that the benefits would be significantly eroded if the processing power were explicitly taken into account. Finally, while cooperation packets in B-CMAC are sent at no energy cost, we believe that this bound is not exceedingly loose because these packets are short and thus the amount of required energy is limited.

We have also investigated the impact of collisions on the network performance. This study was needed because the effectiveness of the alignment algorithm had to be tested, and we also wanted to check how adversely the interference due to cooperation may affect this metric. This quantity is computed as the ratio of the data packets that are not positively acknowledged over all the data frames sent, because once the channel is reserved and no fading is present, only excessive interference (i.e., a collision) can destroy the packet. We have also simulated a version of CRCM (called CRCM$_{CS}$) which includes a slightly more realistic antenna model. In this case, rather than the flat top pattern, a cone and sphere approximation was used, with a ratio between main and secondary lobe of 11.0 dB. The results are reported (only for a 10-node network) in Fig. 2.10 against the normalized load, defined as the total number of transmitted bits normalized by the product of simulation duration and bit-rate.

As far as our protocols are concerned, we remark that collisions are not affected by the in-

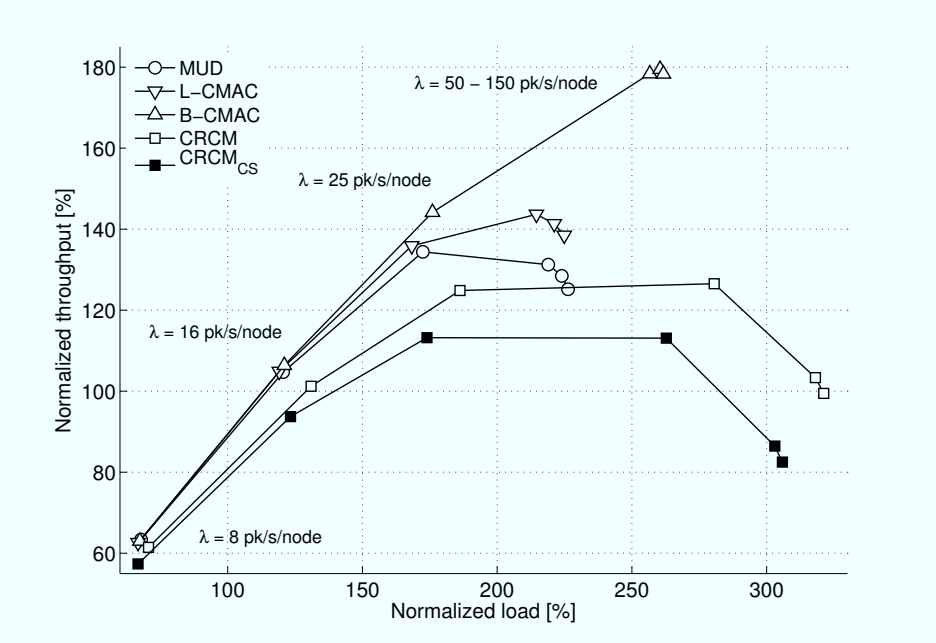


Figure 2.11. *Normalized throughput vs normalized load for a 10-node network and 10240 bit payload. The numbers close to the symbols are the nominal load in pk/s/node*

clusion of the cooperation. As a matter of fact, the curves for L-CMAC (where collaboration packets do create interference) basically overlap with the result for MUD. This proves how the choice of a short length for cooperation packets and a directional transmission strategy does not compromise spatial reuse. In addition, the curve for B-CMAC shows a higher percentage of collisions at saturation load, but this is justified because of the larger throughput achieved.

On the other hand, CRCM generates fewer collisions than the CMAC protocols. This fact is due to the more conservative collision avoidance algorithm, which averts a bigger share of communications in order to prevent this problem. The tradeoff between transmission attempts and collision avoidance is well known, and we have opted for a more aggressive strategy so as to exploit the possible spatial reuse of antenna arrays (see Section 2.6). In addition, in CRCM$_{CS}$ the increased interference due to the secondary lobes leads to a predictably higher percentage of collisions, which negatively impacts the protocol behavior, as discussed in the following.

The last metric studied is the normalized throughput as a function of the normalized load. The former is measured as the total amount of data bits successfully delivered and acknowledged normalized by the product of simulation duration and bit-rate, while the latter is defined as in Fig. 2.10. A percentage above 100% means that on average more than

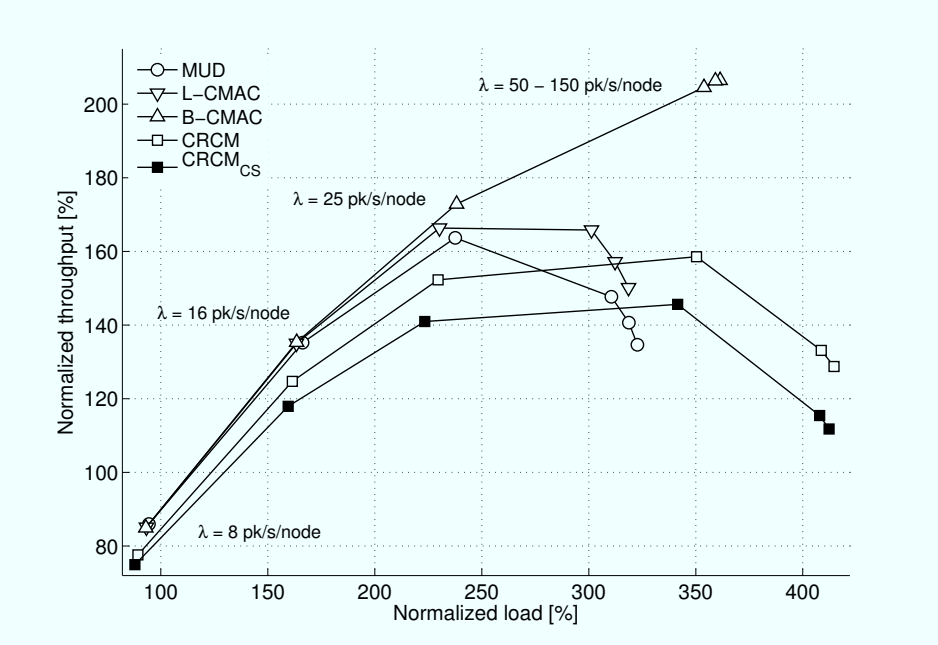


Figure 2.12. *Normalized throughput vs normalized load for a 15-node network and 10240 bit payload. The numbers close to the symbols are the nominal load in pk/s/node*

one link was active in any given instant of time, and thus it is a proof of spatial re-use.

We have studied this metric for three different data packet sizes (4800, 10240 and 36000 bits), to assess the impact of overhead (like RTS/CTS) on the system performance. The first length is almost equal to the RTS/CTS/ACK sequence and the last is representative of an Ethernet jumbo packet.

Let us first consider Figs. 2.11 and 2.12 which compare the results for 10 and 15-node networks, respectively, with payload size fixed to 10240 bits. In both cases, our protocols achieve significant improvements over CRCM and a more graceful degradation with respect to the load. L-CMAC gains between 20% and 10% in throughput close to saturation conditions, while B-CMAC can deliver about 50% more data frames than CRCM. This is a consequence of the increased coordination brought by the enhancements: the time wasted because of collisions, transmissions to deaf nodes and other protocol inefficiencies is extremely limited. Therefore more time is left to carry data packets and the interference due to unsuccessful packets (like the RTS sent in backoff state) is reduced. The performance gap is extremely significant and proves that cooperation, along with the other mechanisms, is a very promising road to abate deafness and its performance drawbacks. In the 15-node network throughput gains by our protocols are slightly lower due to the harsher interference. We stress that also a weak use of collaboration (L-CMAC) is able to enhance the overall

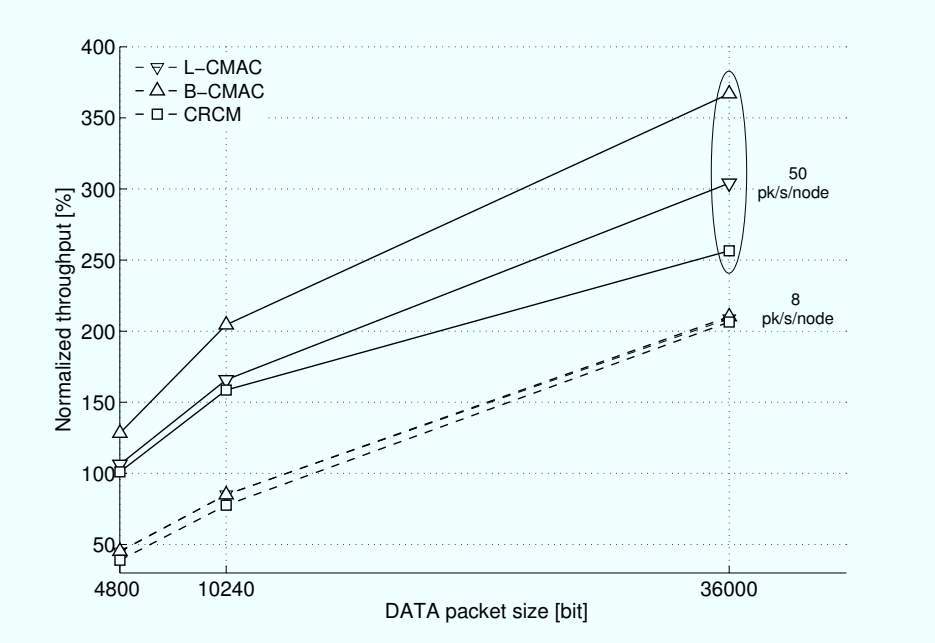


Figure 2.13. *Normalized throughput vs payload length in a 15-node network. The numbers close to the symbols are the nominal load in pk/s/node*

performance with respect to the CRCM solution. This testifies once again how the deafness reduction achieved by cooperative protocols leads to more efficient and better coordinated medium access. In addition, the presence of some gain also at low loads actually highlights the value of the protocol, because in such conditions (very low congestion) we would expect all the protocols to deliver very comparable performances. Instead, the increased coordination enables to get close to optimal performance (throughput/load $> 90\%$). This is still an effect of the improved network coordination and collision avoidance due to deafness reduction.

In addition, we have also included (still in Figs. 2.11 and 2.12) the performance for CRCM$_{CS}$. The presence of the secondary lobes creates additional interference. Therefore more collisions occur (see the discussion for Fig. 2.10) and the throughput of CRCM diminishes by 10% or 15%. This shows the importance of using realistic antenna patterns for a reliable estimation of the protocol performance. We stress that the simple upgrade to the cone and sphere model (which is still far from an actual antenna diagram) cuts the metric by 15%. Thus, it reasonable to expect that more realistic models could reduce the effective performance even further.

We have also studied the dependence of throughput on the packet size in a 15 user network. Fig. 2.13 reports this metric as a function of the frame size. The curves are divided

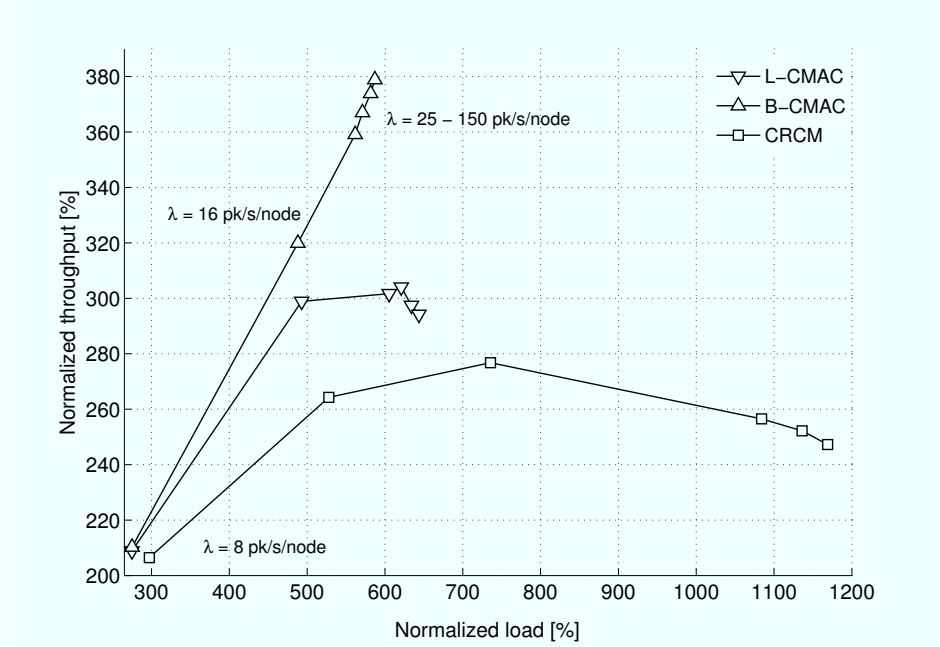


Figure 2.14. *Normalized throughput vs normalized load for a 15-node network and 36000 bit payload. The numbers close to the symbols are the nominal load in pk/s/node*

into two triplets: the lower one is related to a packet generation rate of 8 pk/s/node (very low load), while the higher group stands for a 50 pk/s/node arrival intensity (saturation). All protocols generally show important improvements, and this could have been foreseen. It can be observed that the longer the packet, the larger the throughput on an absolute scale, and this is due to two main reasons. First of all, the relative overhead of the handshake decreases with respect to the payload size. Thus, more time is spent sending data packets, rather than control frames. Secondly, protocols that are more affected by deafness can enjoy a more limited improvement, because the interference due to this problem increases the packet error rate. Thus the benefits of a larger payload are counterbalanced by the heavy losses inherent in the protocol itself. Now it is more evident that cooperation uses more efficiently the available resources with an increasing payload size. Especially in saturation, CMAC has a higher throughput than CRCM and the depicted curves are steeper. This graph proves that the behavior of the CMAC protocols scales better with respect to the packet size.

The last figure about this metric (Fig. 2.14) reports normalized throughput against normalized load for the large payload length of 36000 bits and a 15-node network. The same observations made for Fig. 2.13 can be repeated, but Fig. 2.14 with Fig. 2.11 illustrate that the curves are steeper for smaller packets. This can be expected, too, and the reason is that longer packets are more vulnerable to interference and protocol inefficiencies (like deafness).

In order to support this statement, we have computed the curve slope for B-CMAC at loads in the 8-25 pk/s range (where the curve fits well a linear approximation). In the short packet case (whose figure has not been presented for limits of space and similarity of trends with Fig. 2.12) the slope is very close to 1, while for the longest packet format it is just half that value. Moreover, CRCM shows an inclination of about 0.9 in the former case, but it plummets to just 0.2 in the latter case, thus confirming that deafness is generating a high level of interference also at low loads.

Moreover, we notice that for the long packets, the CRCM and L-CMAC curves change their slope already at rates between 16 and 25 pk/s, the reason being that for the second generation rate the nodes try to inject into the system more bits than the network can actually support. Each successful packet needs $36000 + 9 \cdot 240 + 9 \cdot 250 + 240 + 160 = 40770$ bits (respectively for payload, RTS, CTS, cooperation and ACK). At a rate of 25 pk/s each node creates about 1020 kbit/s, which is already more than the bandwidth capacity.

2.8 An Analytical Model for the Impact of Collisions

Since the purpose of the alignment algorithm (Sections 2.5 and 2.6) is to reduce the impact of collisions, and because of the relevance of such a procedure for the overall protocol efficiency, it was important to get some insight on the performance of our system. For instance, it may be interesting to have some qualitative dependence of the probability of a collision on the node density or on the number of antenna sectors. To this aim, we have developed an analytical model that computes the chance of this event on a data packet. However, some simplifying assumptions have been adopted, in order to make the problem tractable:

- the analyzed protocol does not start backoff cycles after the lack of a CTS reception;
- we do not explicitly take into account cooperation or Communication Register (because collisions are not affected by this mechanism, as explained in Section 2.7);
- the antenna patterns are of the pie-slice type. They are described by circular sectors of unit radius and angular width $2\pi/M$, M being the node beamwidth;
- packets have exponential duration. While frame lengths are not exponentially distributed, our model does not distinguish between the different types of packet (e.g.: data or handshake). Thus, the exponential assumption emulates the presence of very short frames like acknowledgments or RTS/CTS;

- the only reason why a node does not update its DNAV is that it was transmitting a packet while a channel negotiation was taking place. We point out that only a transmitting node may miss a DNAV update because the antenna model described in Section 2.2 would enable it to decode RTS/CTS even during a data reception.

- nodes transmit packets in an asynchronous and independent fashion.

The full description and development of the model is a tedious process that we shall not pursue because of the limited space. The interested readers are referred to [26] for the complete details. Instead, we shall describe the key points and report the important results.

Let us assign the value 1 to the maximum communication distance between two nodes (i.e., all distances are normalized to this value), and let us regard the network as a circle of radius R (normalized to the communication range) which hosts N terminals uniformly distributed in this area. Each of them generates Poisson traffic with intensity ρ.

A data exchange between two terminals, $\boldsymbol{A}$ and $\boldsymbol{B}$, fails because of a collision if a third aligned node, which is inside the region where the receiver can correctly decode an incoming packet, erroneously starts a new transmission. Then, the probability p_c that one of the $N-2$ nodes not involved in the $\boldsymbol{A}$-$\boldsymbol{B}$ communication may cause a collision is equal to the product of two terms:

$$p_c = p_{geom} \cdot p_{NAV} \,. \tag{2.1}$$

where p_{geom} is the probability that the interfering node is in line with the $\boldsymbol{A}$-$\boldsymbol{B}$ direction and is equal to $1/(MR^2)$, whereas p_{NAV} is the probability that this node has not been able to check its alignment (see Section 2.6) and that it starts a new frame before the other communication is over. This causes a collision, because the interferer would send a circular packet (either RTS or CTS), thus covering also the critical $\boldsymbol{A}$-$\boldsymbol{B}$ bearing. p_{NAV} is the sum of three independent terms:

- p_1: a node was sending a data packet and starts a new frame immediately after that as its queue is not empty;

- p_2: the node ends a packet transmission and becomes idle again, but a new frame arrives from the higher layers and its negotiation begins;

- p_3: the node completes a packet delivery, returns idle, receives an RTS and sends the corresponding CTS.

The expression for these probabilities are the following:

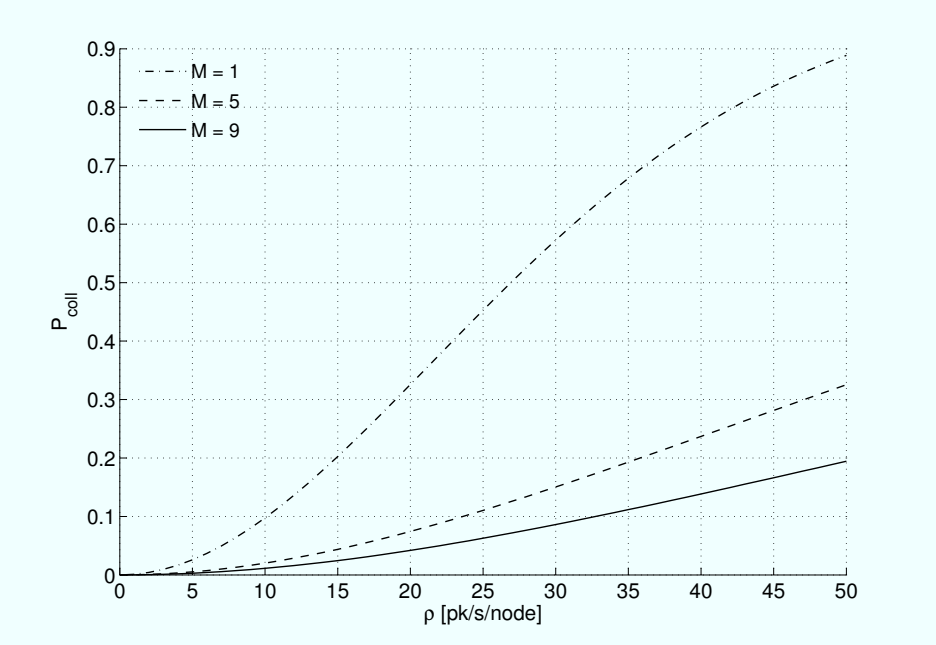


Figure 2.15. *Collision probability vs nominal load ρ [pk/s/node] for different values of the number M of sectors required to cover the whole horizon.*

$$p_1 = \left[\left(1 - e^{-\mu T_{AB}}\right) p_q\right] \tag{2.2}$$

$$p_2 = \pi_1 \left(1 - e^{-\rho T_{AB}}\right) \left(1 - e^{-\mu T_{AB}}\right) \tag{2.3}$$

$$p_3 = \pi_1 \int_0^{T_{AB}} (1 - \mu v) f(T_{AB} - v) dv \tag{2.4}$$

where $T_{AB} = 1/\mu$ is the average duration of the communication between $\boldsymbol{A}$ and $\boldsymbol{B}$, p_q is the probability that the packet queue is non empty, π_1 is the probability that a node is transmitting but has an empty queue, and $f(x)$ is the probability of receiving a packet from another node in an interval of time x seconds long.

Each node may cause a collision with probability p_c. By our assumptions, we can suppose that each of the $N - 2$ nodes may collide independently of the others. Then the total probability of a collision may be computed as the chance that at least one node generates this event, which is:

$$p = 1 - (1 - p_c)^{(N-2)} . \tag{2.5}$$

Let us consider the dependency of this metric against some important parameters. The first is the beamwidth, which is expressed as $2\pi/M$. The result is reported in Fig. 2.15, which proves first of all that networks equipped with omnidirectional antennas would be heavily

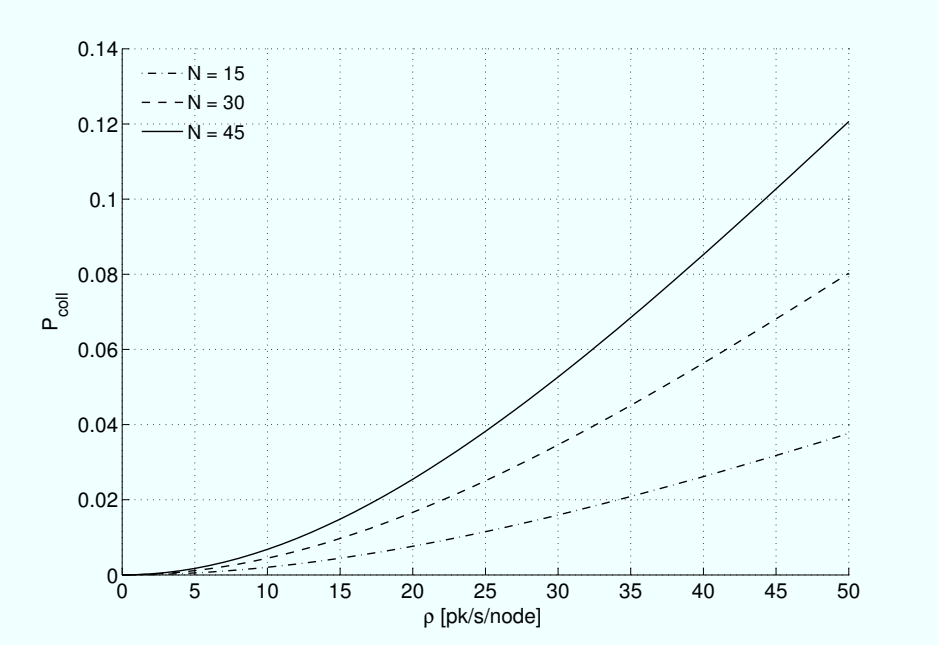


Figure 2.16. *Collision probability vs nominal load ρ [pk/s/node] for different node densities. N represents the number of nodes in a network of radius $R = 3$, while the number M of sectors used to span the whole horizon has been set to 9.*

beset by this problem. Instead, a moderate number of antennas (4-8, corresponding to $M = 5$, $M = 9$) can reduce this problem by a factor of 3 at high loads. Our computations were purely based on the interference suppression capabilities of smart antennas, while other mechanisms (like backoffs or cooperation) have not been included. This testifies once again that even in this simple setting the achievable capacity improvements are very important.

Moreover, the proposed algorithm should properly degrade with the node density. In Fig. 2.16 the collision probability is depicted vs the network size N, given a constant radius $R = 3$ and number of sectors $M = 9$. We observe that for moderate densities (1/2 a node/unit area) the collision probability is indeed small (4%). This number scales linearly with the node density (up to 12% in our figure) at a density of 1.5 nodes/unit area. This implies that even at a high level of interference, this protocol can manage to have an acceptable loss rate (just above 10%).

Finally, let us examine Fig. 2.17, which provides the experimental validation for our model. The simulated protocol has been the CMAC lower bound without backoff mechanism (called Modified CMAC or M-CMAC). Let us remark that, as for all the figures of this section, the metric is not depicted against the nominal Poisson arrival intensity, but against the effective arrival rate (i.e., ρ in our theoretical model). In the simulations discussed in

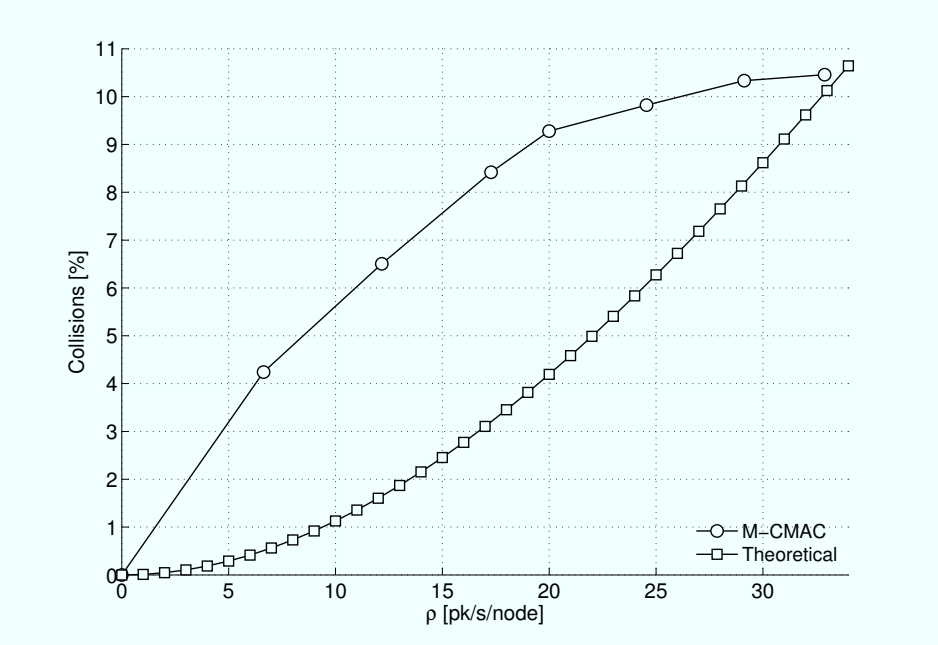


Figure 2.17. *Experimental validation for the proposed analysis: comparison between M-CMAC and theoretical model.*

Section 2.7 this can be computed as the average number of packets that are effectively sent over the air, divided by the simulation duration, neglecting frames provided by the higher layers that are not transmitted because of deafness or network congestion.

It can be observed that under saturation conditions there is fair agreement between model and simulation. Since the heavy traffic setting generates the largest number of collisions, the model is able to correctly predict this metric in the most critical scenario. In addition, the experimental curve is above the theoretical one because the model neglects some causes of possible packet corruptions. For instance, in our assumptions we have taken into account only one cause of DNAV update failure, while other situations may contribute to this problem (as discussed in Section 2.5).

2.9 Important Packet Formats

In this section we briefly describe the format of the cooperation and CTS packets as implemented in L-CMAC.

The cooperation frame includes the following fields:

- Frame control (2 bytes): the conventional MAC header, which includes the packet type

(collaboration, in this case), the protocol version and the usual 802.11 flags (encryption, whether the packet is a new frame or a retransmission, etc.) [5];

- Duration (2 bytes): estimated remaining time for the communication;
- Receiver Address (6 bytes): address of the receiver
- Transmitter Address (6 bytes): address of the transmitter
- CRC (4 bytes)

The CTS packet contains two additional fields with respect to the conventional version: it incorporates the CTS sender address and it also includes the bearing of the RTS transmitter with respect to the node it is trying to contact. This field is needed for the alignment algorithm, and it is one byte long, because an accuracy of $360^o/(2^8) = 1.4^o$ is sufficient in most situations.

2.10 Discussion and Conclusions

This chapter has designed and analyzed a new family of MAC protocols for directional antenna ad hoc networks. A deep study of antenna arrays has given us the necessary insight to define three new methods aimed at reducing the impact of two important problems for the considered communication systems: deafness and collisions. We have proposed a low complexity multiuser detector, a new collision avoidance algorithm and a cooperative mechanism. We have explored the gains achievable by these techniques, and we have proved that not only can the extent of deafness be extremely reduced, but also that overall performance metrics like throughput and power consumption can significantly benefit from our proposed solutions.

We have tested our protocols in single hop networks in order to verify their ability of performing spatial reuse. Nonetheless, there are still important areas of research that we are exploring. In particular, we are focusing our attention on multihop scenarios and on a mobility-aware version of our systems, in order to deal with the nomadic nature of many ad hoc networks.

In conclusion, the study carried out in the first part of this thesis has provided a first remarkable example of how cooperation, intended in this case as the will of sharing information to improve coordination, can be beneficial, as even if a node has to spend part of its resources to help other terminals, the overall network performance tends to improve.

Moreover, the discussion of the results presented here has shed light on the importance of defining MACs that are capable of effectively supporting cooperative behaviors. Such considerations will be further extended in the next chapters, considering other and more advanced forms of cooperation in ad hoc networks.

Chapter 3

Hybrid Cooperative-Network Coded ARQ in Wireless Networks

In the initial part of this thesis we have introduced a first instance of cooperation in wireless ad hoc networks, given by nodes that share information on ongoing communications to better coordinate medium access at the link layer. While the presented solution has the merit of shedding light on some of the beneficial effects that can stem from properly designed cooperative schemes, several other approaches can be thought of to implement collaboration among nodes. From this viewpoint, *cooperative relaying* has recently emerged as a powerful tool to enhance the performance of wireless networks, and is attracting an ever growing attention in the research community. The basic principle of this paradigm is to let terminals act as *relays* by sending some form of redundancy on a data unit on behalf of the original source, so as to provide the intended addressee with pieces of information transmitted over the statistically independent source-to-destination and relay-to-destination channels. By having the recipient apply combining techniques on the collected fragments, such an approach enables a boost for the average quality of wireless links in terms of either sustainable rate or outage probability thanks to the spatial diversity it entails. The advantages offered for node-to-node communications, in turn, have the potential of inducing beneficial effects at a broader level. Indeed, the reduced number, or alternatively the shorter duration, of the transmissions required to deliver data units leads to both a lower interference level and a higher parallelism in the network.

It is important to remark that this form of collaboration differs from the one introduced in Chapter 2 in some key aspects. In the first place, cooperative relaying proposes a terminal to help its neighbors by directly transmitting data, rather than by simply exchanging with

them information later to be used by each node to access the channel. Secondly, this strategy does not only focus on improving coordination at the link layer, but rather aims at enhancing the performance of a wireless link by means of a cross-layer PHY-MAC approach, as both coding techniques and medium access control are involved. Finally, while the collaborative approach of Chapter 2 is restricted to the specific case of multi-antenna systems that rely on directional communications, relaying can be implemented in any kind of wireless network, and thus applies to several different scenarios. [1]

Cooperative relaying has been thoroughly shown in the literature to trigger important gains over non-collaborative solutions both from a theoretical and from a practical angle. However, the basic paradigm described before exhibits some intrinsic limitations which may hamper its potential. In this perspective, for instance, it is interesting to notice that relay nodes are asked to behave selflessly, spending their own resources to help other terminals without pursuing a goal of their interest, e.g., delivering their own traffic, and without having an immediate benefit but the greater good of the network. Such an objection may actually deter nodes from taking part in relayed phases, thus bounding the gains achievable in real-world implementations.

Starting from this remark, we propose a way to mitigate the aforementioned drawback by introducing the concept of *hybrid cooperative-network coded ARQ*. The key idea that underpins our solution is to allow relays transmit a linear combination of a third-party packet and of a data unit taken from their own queue, instead of simply sending redundancy on the former. The advantage of this approach is twofold. In the first place, involvement in cooperation can be boosted, since collaborating nodes are offered the opportunity to achieve a benefit of their own while helping other terminals. On the other hand, encoded retransmissions enable more traffic to be served at no additional cost in terms of energy and bandwidth with respect to the plain cooperative paradigm, possibly inducing positive effects on the overall network performance.

These benefits are obtained by smartly merging the concept of cooperative relaying with that of *network coding*, which has recently been proposed as a way to improve network throughput by coding multiple packets together rather than by transmitting them separately. Furthermore, the form of collaboration that we introduce leverages a novel type of physical layer, called MIMO_NC, which makes it possible to retrieve both the original information content sent by the source and an additional data unit from a cooperator starting from a corrupted version of the former while keeping the decoding gain of plain relaying

[1] Throughout the rest of this thesis we consider wireless networks with single antenna-equipped nodes.

and entailing a low additional cost in terms of complexity.

The goal of this chapter is to present and investigate the idea of hybrid cooperative-network coded ARQ, discussing its implementation issues and proposing efficient link layers that support the scheme, as well as to study its performance by means of analysis and simulations. With an eye on this, we start in Section 3.1 by dealing with the concepts of cooperative relaying and network coding and by reviewing some of the key contributions that have been proposed in the literature in these fields. Then, in Section 3.2, we describe in detail the principles of the solution that we introduce, highlighting its potential as well as the requirements it entails both at the physical and at the link layer. With Section 3.3 we provide an overview of the MIMO_NC PHY that supports our collaborative paradigm, while Section 3.4 describes and discusses an analytical model for cooperative-network coded ARQ in simple network topologies, prompting an initial glance on the performance improvements that can be achieved and offering some hints on the proper design of link layers that rely on such a mechanism. The second part of the chapter is then dedicated to the design of medium access protocols that implement the proposed approach in several different scenarios. As a first step, we focus on carrier sense based environments in Section 3.5. The performance of the outlined scheme is thoroughly investigated and compared to that of reference benchmarks by means of extensive simulations in a variety of settings, including single- and multi-hop ad hoc networks. Subsequently, in Section 3.6, we concentrate on wireless systems that employ time division multiple access (TDMA) as channel control policy, proposing a cooperative-network coded protocol and analyzing its behavior. Finally, Section 3.7 draws the conclusions of the chapter, stressing two key remarks that can be inferred from our line of study. On the one hand, the improvements offered in both CSMA and TDMA environments over basic schemes will confirm the effectiveness of the idea that underpins the proposed cooperative approach. On the other hand, an accurate comparison of the results achieved in the two scenarios will highlight how different medium access policies can have a dramatic impact on the performance of cooperation, laying the foundations for the analyses carried out in the following chapters.

3.1 Related Work

The collaborative paradigm that we propose in this chapter is based on two communication approaches that have recently emerged in the literature as promising means to improve the performance of wireless networks: *cooperative relaying* and *network coding*.

Cooperative relaying takes advantage of spatial diversity to combat fading and channel

impairments, allowing faster and more reliable links. As discussed in the introduction of this chapter, the basic idea that underpins these schemes is to let nodes in the proximity of an active source-destination pair, referred to as *relays*, help data delivery by forwarding to the addressee redundancy on a message originally sent by the source. The receiver, in turn, collects fragments of information conveyed through the disjoint and ideally independent source-to-destination and relay-to-destination channels and can apply combining techniques so that the average performance of the link, in terms of either achievable rate or outage probability, improves.

Pioneering studies [28–30] proposed relaying approaches where, in case of a decoding failure at the destination, cooperating terminals retransmit a copy of the packet they received from the originator of the data. In particular, in the Amplify&Forward variation such nodes are supposed to transmit an amplified version of the analog signal (plus noise) received from the source, whereas with Decode&Forward relays have to decode, re-encode, and send a copy of the original data unit. Among the significant works that can be found in the literature, a detailed and insightful theoretical investigation of these schemes was presented in [31].

A different line of study was first introduced in [32], starting from the observation that having relays simply repeat information bits does not represent an efficient behavior from a coding perspective. Therefore, cooperating nodes were proposed to retrieve the payload sent by the source and to re-encode it so as to transmit to the destination a different part of the original codeword, following a distributed Hybrid Automatic Repeat Request (HARQ) error control rationale. These techniques generally fall under the name of *coded cooperation*, and have been shown to achieve full diversity, i.e., the maximum gain obtainable by exploiting spatially disjoint channels [33, 34].

While the aforementioned works mostly focus on relaying at the physical layer, more recent studies have also started to propose medium access strategies that try to efficiently support it, e.g., [35–39]. In order to reap the potential gains of cooperation, indeed, some additional coordination at the link layer is required in order to detect when relaying is convenient as well as to choose the terminals that should be involved in a collaborative phase among a set of candidates. From this point of view, remarkable examples are [38] and [39]. In the former case, an IEEE 802.11-style protocol is proposed, with each node selecting whether to perform a direct transmission towards its destination or to split the link taking advantage of a helper so as to minimize the total delivery time based on a knowledge of the involved channels gathered from past communications. In [39], instead, the authors introduce a dis-

tributed procedure for selecting the cooperator that requires the exchange of some control packets among relay candidates when the source-destination link has to be established. Let us remark, however, that, despite these research efforts, the definition of link layers tailored to cooperative relaying and the study of their behavior in large and realistic networks is still an open issue. These aspects will be thoroughly discussed in Chapters 4 and 5.

Alongside cooperative relaying, another strategy that has proven to strongly improve communication efficiency is Network Coding (NC), whose aim is reducing the number of transmissions required to deliver a set of data units in a wireless scenario by implementing the *store, decode and forward* paradigm. Along this line of reasoning, each node stores all the packets it receives, even those not addressed to it, and, when granted access to the medium, sends out linear combinations of data units of its own and of part (or all) of the stored information, in place just of the former elements. When all the terminals implement such a policy, it is possible for each of them to retrieve P data units as soon as P linearly independent coded packets are received, by simply solving a system of linear equations [40, 41]. This requires the knowledge of the coefficients used when combining together different payloads, which, in general, can be included in the header of the transmitted coded packet [42]. Extensive studies have shown that significant gains are attainable for both throughput and delay when NC is employed, e.g. [42, 43].

It is interesting to observe that cooperative relaying and NC have different but complementary goals. Indeed, the former tries to make a smart use of redundancy in order to leverage spatial diversity, whereas the latter increases the efficiency of each transmission by sending multiple data units combined together. Moreover, while cooperative relaying proposes several nodes to send copies of the same packet to the same destination, with NC several terminals send different packets to different terminals. Finally, cooperation and NC also suffer from opposite weaknesses, with the former offering protection against channel impairments at the cost of a less efficient bandwidth usage, and the latter improving channel reuse by means of a clever reduction of redundancy at the expense of an increased vulnerability to high error rates.

Such points of contact have led in the past years to several attempts of combining the best of both techniques [44–53]. However, the vast majority of the papers in this field focuses on physical layer performance such as bit error rate and diversity order [46–48] or on information theoretic metrics like capacity regions [50–52]. In all cases, very simple MAC protocols are considered, and some centralized scheduler is often assumed, so as to study the aforementioned performance criteria without undesired cross-layer interactions. Moreover, the

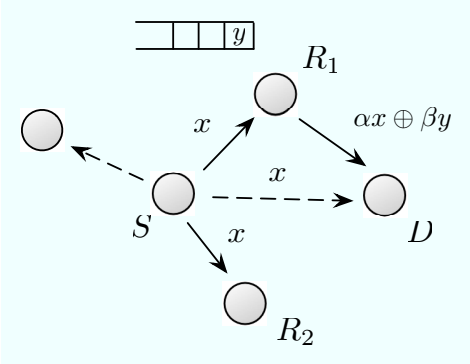


Figure 3.1. *Reference topology for the description of hybrid cooperative-network coded ARQ.*

protocols are tested in toy topologies, which again avoid coupling between PHY and higher layers but leave open the question of the actual performance in more complex and realistic networks involving large and random node deployments. The work presented in this thesis, instead, is one of the few aimed at creating practical medium access policies based on a hybrid cooperative-NC physical layer. To the best of our knowledge, the only other research effort that has developed an actual protocol based on these ideas is [53], which implements in a real-world testbed the concept of Analog Network Coding [51]. Nonetheless, [53] only analyzes topologies composed by up to 5 nodes, while the solution that we propose is tested by simulation in a variety of network settings. In addition, while it is not clear how the concept in [53] can scale with the network size or node arrangement, we prove that our solution can work on systems of arbitrary dimension and analyze it in scenarios with up to 100 nodes.

3.2 Hybrid Cooperative-Network Coded ARQ

With reference to the topology depicted in Fig. 3.1, let us assume that node $\boldsymbol{S}$ has packet x to deliver to $\boldsymbol{D}$. In the event of a communication failure over the direct link, basic retransmission policies, i.e., Automatic Repeat reQuest (ARQ), suggest the source to perform another try after a time interval whose duration depends on the underlying medium access control. However, if the decoding at $\boldsymbol{D}$ has not succeeded due to a bad state of the propagation channel, such attempts are not likely to succeed unless performed after a period of time long enough for it to decorrelate. It follows that the time-correlation property of the wireless medium may require several transmissions for the payload to be eventually delivered, with detrimental effects on latency, throughput and energy consumption. On the other hand, thanks to the broadcast nature of the wireless channel, another node, say $\boldsymbol{R}_1$

or $\boldsymbol{R}_2$, may have decoded the packet sent by the source even if it was not intended for it. In this case, the decode-and-forward cooperative paradigm proposes one such terminal to immediately perform a retransmission on behalf of $\boldsymbol{S}$, re-encoding the data unit x and sending it to the addressee.[2] The destination, then, has at its disposal two copies of the same packet received over statistically independent channels and can perform Chase combining, strongly improving the decoding probability.[3] In a nutshell, cooperative relaying substitutes the temporal diversity that characterizes plain ARQ with spatial diversity, enhancing the reliability and reducing the latency of recovery phases. All these properties, thus, make it particularly appealing, especially in harsh wireless environments. Incidentally, we also observe that the described strategy implements an hybrid ARQ (HARQ) policy of type II. Therefore, in the remainder of this chapter we will use the expressions *cooperative relaying* and *cooperative HARQ* interchangeably.

The gains achievable with this approach, however, come at the expense of having relays spend their own resources, e.g., energy and bandwidth, to help other terminals, without any personal payoff but the contribution they offer to network efficiency and the hope that somebody else would behave similarly towards them when in need. Clearly, such a selfless behavior is not compelling in general and may actually discourage nodes from cooperating. From this point of view, indeed, when supporting a surrounding terminal, a relay accelerates the exit from an otherwise longer failure recovery phase of a competitor of its. Not only then does a collaborative behavior entail for the relay a resources expense to serve third-party traffic, but also it reduces the average bandwidth available to those who cooperate by throwing back in additional channel contenders. This remark may especially apply to large and congested networks, which, instead, would represent an ideal scenario to reap cooperative gains at the utmost due to their high interference level and subsequent lower average link quality.

In light of this discussion, this thesis introduces a novel approach that aims at preserving the diversity gain offered by cooperative HARQ while reducing the drawbacks it entails. In order to describe the key principles of our approach, let us refer again to the topology of Fig. 3.1, and suppose that a relay candidate for the $\boldsymbol{S}$-$\boldsymbol{D}$ link, say $\boldsymbol{R}_1$, has a packet y in its

[2]Throughout this chapter we focus only on decode-and-forward cooperation. Different forms of cooperative relaying, e.g., coded cooperation, will be considered in Chapter 5.

[3]In particular, cooperative relaying increases the diversity order d that characterizes data exchanges over the $\boldsymbol{S}$-$\boldsymbol{D}$ link, defined as the asymptotic slope of the error probability P_e vs SINR curve in its logarithmic form, i.e., $d = \lim_{\mathrm{SINR}\to\infty} \frac{\log P_e}{\log \mathrm{SINR}}$ [54]. While d is equal to 1 for plain ARQ, spatial diversity with a single relay has the potential of moving it up to 2.

queue addressed to $\boldsymbol{D}$. In this condition, we propose the cooperator to transmit a linear combination $\alpha x \oplus \beta y$ of the payload originally sent by $\boldsymbol{S}$ and of a data unit of its own, rather than simply delivering a copy of the former. If such an encoded retransmission is successful, i.e., if the destination can actually retrieve the information content of both x and y, two significant improvements can be achieved with respect to the plain decode and forward strategy. On the one hand, nodes may be encouraged to take part in cooperative phases, since they are offered the possibility of opportunistically serving some traffic of their own without the medium negotiation they would have had to undergo otherwise. Secondly, the solution that we propose allows to deliver an additional packet at no additional cost in terms of energy and bandwidth over cooperative HARQ, as y is simply coded together with a data unit that would have been transmitted anyways. From a broader perspective, then, network coded relayed phases can lead to an overall performance boost, thanks to the smarter usage of the available bandwidth they enable.

In order to implement our idea, it is necessary to somehow combine different frames. From this viewpoint, network coding appears as the ideal solution, since, as discussed in Section 3.1, it allows nodes to code packets together so as to improve the overall network efficiency. However, standard NC can retrieve information only from correctly delivered Protocol Data Units (PDUs), while our scheme inherently has to deal with retransmissions and thus with corrupted packets. This issue can be overcome by using a special type of physical layer named MIMO_NC [48] and in particular the so called Super MIMO_NC [49]. Such a PHY is based on the multiple-input multiple-output nature that characterizes both MIMO and network coded systems. In particular, MIMO_NC leverages the fact that the packet mixing introduced by NC resembles the propagation through a wireless MIMO channel, applying network coding to physical layer PDUs and then decoding them by MIMO signal processing, treating the encoded data units as multiple inputs and regarding the received frames as multiple outputs. While a more in-depth discussion of this physical layer will be provided in Section 3.3, for the moment we remark that, with reference to Fig. 3.1, not only does MIMO_NC allow $\boldsymbol{D}$ to decode both x and y when a coded retransmission is performed, but also it preserves the maximum diversity order (equal to 2) for the corrupted frame.[4]

A physical layer capable of supporting coded retransmissions is not the only necessary ingredient to implement the solution that we propose. Indeed, in order to take advantage of hybrid cooperative-NC ARQ, additional coordination among nodes has to be provided. In

[4]Incidentally, we also stress that, with reference to Fig. 3.1, the encoded packet $\alpha x \oplus \beta y$ is neither a simple bitwise XOR nor a sum of modulated waveforms, but instead it is a linear combination of vectors in a Galois field according to network coding principles (see Section 3.3)

this perspective, for instance, it is essential to determine who has to act as relay if multiple candidates are available, as well as to design a policy to decide whether or not a coded retransmission shall be performed. These details can be implemented in several different ways, according to the nature of the underlying medium access control, e.g., distributed or centralized. Nonetheless, defining link layers that effectively support hybrid cooperative-network coded ARQ is a task of paramount importance so as to make the most out the potential gains it offers. In this chapter, in particular, we propose two different medium access policies that achieve this goal in carrier sense-based (Section 3.5) and time division-based (Section 3.6) wireless networks.

3.3 The MIMO_NC Physical Layer

As discussed in Section 3.2, hybrid cooperative-network coded ARQ asks nodes to retrieve two data units exploiting only a corrupted version of the former and a linear combination of both of them. From this perspective, standard NC techniques do not appear as a viable solution, since they have been designed to work only with properly decoded packets. Instead, such a task can be actually accomplished by resorting to a novel type of physical layer called MIMO_NC, which was first introduced in [48]. This work has shown how to get the diversity gain of cooperation with the throughput efficiency of NC by means of signal processing techniques borrowed from Multiple Input Multiple Output (MIMO) systems. The basic property of MIMO_NC is that N nodes may send combinations, called coded packets, of the same P original information units, IUs, and simultaneously achieve a diversity order of at least $N - \lfloor (P-1)/2 \rfloor$, as well as a coding gain with respect to conventional NC. For instance, if a relay sends a combination of a retransmitted data unit and a packet of its own ($N = 2, P = 2$), the diversity order is still two as in a conventional cooperative protocol, but the relay has been able to both collaborate in the classical sense and serve its own traffic.

An accurate description of such a PHY goes beyond the scope of this thesis. In this section, instead, we sketch its basic working principles in networks with single antenna equipped terminals,[5] referring the interested reader to [48, 49].

The idea of combining MIMO and NC stems from the observation that these techniques are based on a similar description of the system. In both cases, indeed, data are retrieved as the solution of a linear system in the form $Ax = b$, where b contains the received data

[5] We remark that even though MIMO_NC employs signal processing algorithms that are typical of MIMO systems, multiple antennas are not necessary for it to properly work.

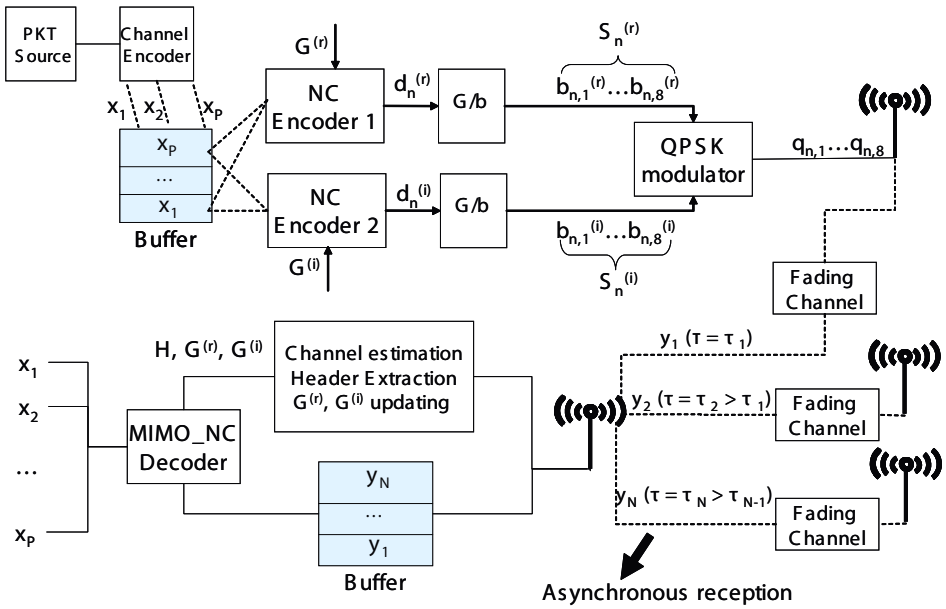


Figure 3.2. *The MIMO_NC encoding/decoding procedure [48].*

(MIMO) or samples (NC) and A is the channel matrix H (MIMO) or the coding matrix G (NC). This similarity is useful to develop an integrated system where MIMO and NC coexist at the same layer. To achieve this goal, MIMO_NC introduces two key novelties: i) all NC operations are moved towards the physical layer, and ii) a different decoding phase based on soft-decoding rather than on the inversion of linear systems is designed. The complete flow of the MIMO_NC scheme is represented in Fig. 3.2 and can be divided in two phases: encoding and decoding. During the former phase, each node combines the IUs it has stored in its buffer according to the basic network coding approach, thus producing coded packets. However, in contrast to classic NC procedures, the network encoding phase is moved after channel encoding. Each coded packet is composed by two Linear Combinations (LCs) of the IUs, and each LC is a sequence of Galois symbols in $GF(2^8)$. The transmitter modulates QPSK symbols whose real part is shaped by the bits in the first LC in the CP and the imaginary part is due to the second LC.

As far as the decoding phase is concerned, the considered scheme performs it at the physical layer in order to employ MIMO decoding algorithms (see the bottom part of Fig. 3.2). The received samples are stored into a buffer before being processed by means of a soft decoding scheme based on sphere decoding [55] that reconstructs the IUs. From this point of view, MIMO_NC differs from the classic NC in that all received packets, including corrupted and redundant ones, can contribute to the decoding phase. Also notice that, in this way, all the transmitted energy is employed in the decoding process.

In conclusion, MIMO_NC allows coded retransmissions to take place, enabling a throughput boost while preserving the diversity order of basic decode-and-forward techniques. Furthermore, let us remark that the computational complexity entailed by the scheme is rather low, as it employs the low-complexity sphere decoding algorithms. This aspect is further eased in the case under analysis, as MIMO_NC has to solve at most a 2×2 system (2 inputs, the information units to retrieve, and 2 outputs, the received packets). Hence, the additional cost of implementing our solution on top of an already existing cooperative scheme is indeed very limited.

3.4 An Analytical Model for Cooperative-Network Coded ARQ

In order to gain insight on the behavior of hybrid cooperative-network coded ARQ as well as to get a first quantification of the gains achievable over simpler solutions, we have set up an analytical model to compare them. The scenario is the uplink of a three-node network composed by two terminals, called $\boldsymbol{A}$ and $\boldsymbol{B}$, and an access point. These are surrounded by an infinite population of similar three-node subnetworks, whose positions follow a two dimensional Poisson point process with density σ_i. Both $\boldsymbol{A}$ and $\boldsymbol{B}$ always have a packet to transmit, i.e., the network is at saturation.[6] Time is slotted, and during each slot one data packet and the corresponding error-free feedback are sent. At any given slot, each subnetwork transmits with probability p_s, so that the effective interferers density is $\sigma_i p_s$. It stems that the generated interference follows a b-stable distribution [56], where $b = 2/\alpha$ and α is the path loss exponent.

Terminals are $p(X)$-persistent, i.e., each time a node is given the chance to transmit in a slot, it will do so with probability $p(X)$, where $p(X)$ depends on the number X of retransmissions the current packet has undergone. In other words, a geometric backoff is assumed. Three different protocols are analyzed, named *Basic*, *D&F Coop* and *Coop-NC* for coherence with the rest of the chapter and implementing plain ARQ, decode-and-forward cooperation and hybrid cooperative-NC ARQ respectively. All the schemes are modeled by a Markov chain, which keeps track of the protocol status (which node will perform the next transmission and the value X for the packet being served) as well as of the channel conditions. Hence, the state can be represented by four variables $(X, \mathrm{N},\ H_A,\ H_B)$, where X is the number of times the packet has already been transmitted and varies between 0 and SRL-1, SRL

[6] Saturation conditions are of particular interest in the study of cooperative solutions, since the harsher contention and interference level tend to trigger a larger number of retransmission phases.

being the Short Retry Limit,[7] N is the node that will transmit in the next frame (N is hence either $\boldsymbol{A}$ or $\boldsymbol{B}$) and H_A, H_B represent $\boldsymbol{A}$'s and $\boldsymbol{B}$'s channel Signal to Interference Ratio (SIR), respectively. The SRL value has been set to 3 for Basic, while the cooperative protocols only rely on two independent retransmissions.[8] The channel is subject to correlated Rayleigh fading, and the correlation is modeled according to [57], with carrier frequency of 2.4 GHz and Doppler spread of 40 Hz, in agreement with the network simulations that will be discussed in Sections 3.5 and 3.6. Incidentally, let us remark that while the model which we have set up relies on time division multiple access as medium access control, its memoryless backoff also accurately reproduces the behavior of wireless LANs [58]. Therefore, the results that will be discussed in the remainder of this section apply well to the networking scenarios considered in both Sections 3.5 and 3.6.

In any state, the chain may perform a transmission with probability $p(X)$. Hence, the average steady state transmission probability p_s can be computed as:

$$p_s = \sum_{\forall \text{states}} p(X)\pi(X, \text{N}, H_A, H_B) \ , \tag{3.1}$$

where $\pi(X, \text{N}, H_A, H_B)$ is the steady state probability of being in state (X, N, H_A, H_B). Of course, the $\pi(X, \text{N}, H_A, H_B)$ depend on the packet success probability and hence on p_s. Given an estimate of p_s, the $\pi(X, \text{N}, H_A, H_B)$ are computed and then a new estimate for p_s is evaluated according to (3.1). The process is iterated until convergence is achieved. One of the virtues of this model is to analyze a simplified interference network based on the above protocols. Fig. 3.3 shows how the states are organized and outlines the possible transitions for D&F Coop and Coop-NC.

In Basic, a node that sends a new packet (say $\boldsymbol{A}$), retransmits it until it is correctly delivered or the maximum number of attempts has been reached. When either condition is met, the other node ($\boldsymbol{B}$ in this case) will send its own data. In order to gain a deeper understanding on how the chain works, let us track its evolution in the following situation. Node $\boldsymbol{A}$ has to deliver a new packet for the first time, so the chain starts from state $(0, \boldsymbol{A}, H_A, H_B)$. Let us call $P_{\text{succ}}(H_A)$ the probability to correctly deliver a PDU given that $\boldsymbol{A}$ transmits and its channel state is H_A. If $\boldsymbol{A}$ does not get its frame across (which happens with probability $p(0) \cdot (1 - P_{\text{succ}}(H_A))$), the chain transitions into state $(1, \boldsymbol{A}, H'_A, H'_B)$; if it transmits suc-

[7] The Short Retry Limit represents the maximum number of attempts to deliver a packet performed at the MAC layer before dropping it [5].

[8] Cooperative solutions in general require fewer retransmissions than plain ARQ to achieve a given reliability level, thanks to the spatial diversity they exploit. This aspect, as well as the SRL values chosen for the model, will be further discussed and justified in Section 3.5.

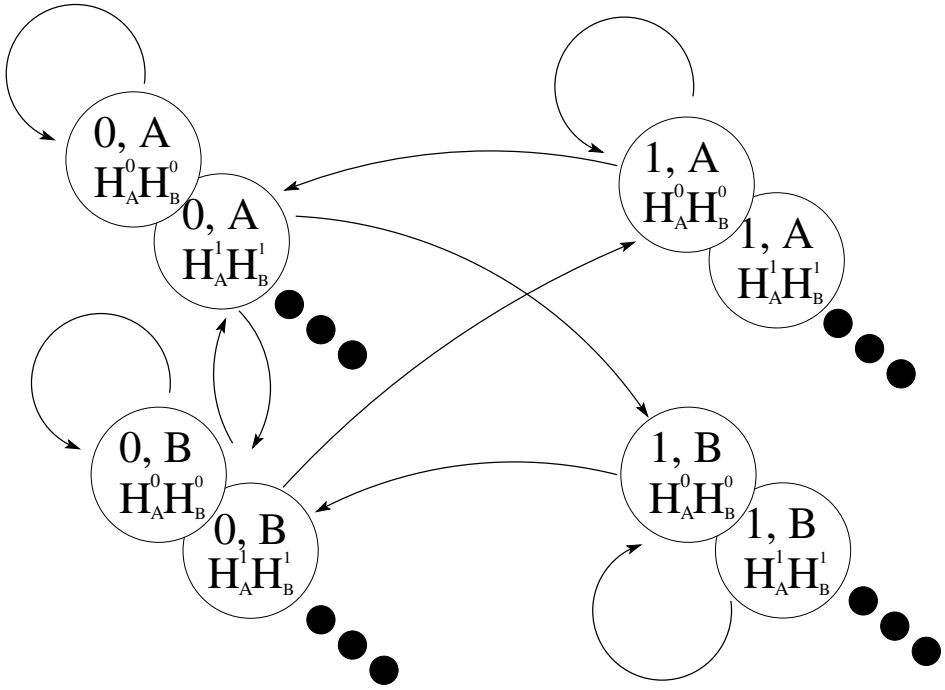


Figure 3.3. *Representation of the Markov chain for* D&F Coop *and* Coop-NC.

cessfully (probability $p(0) \cdot P_{\text{succ}}(H_A)$), it moves into $(0,\ \boldsymbol{B},\ H'_A,\ H'_B)$, otherwise (probability 1-$p(0)$) state $(0,\ \boldsymbol{A},\ H'_A,\ H'_B)$ is the destination. If the system is in state $(1, \text{A},\ H_A,\ H_B)$, it may not transmit and transition into $(1, \text{A},\ H'_A,\ H'_B)$ (probability 1-$p(1)$); it may deliver the packet and move into $(0,\ \boldsymbol{B},\ H'_A,\ H'_B)$ (probability $p(1) \cdot P_{\text{succ}}(H_A)$), or finally it may fail and go into state $(2,\ \boldsymbol{A},\ H'_A,\ H'_B)$ (probability $p(1) \cdot (1 - P_{\text{succ}}(H_A))$). From state $(2,\ \boldsymbol{A},\ H_A,\ H_B)$, either there is a transmission and hence with probability $p(2)$ the chain transitions into $(0,\ \boldsymbol{B},\ H'_A,\ H'_B)$ because the SRL has been reached, or the node backs off, going into state $(2,\ \boldsymbol{A},\ H'_A,\ H'_B)$.

In D&F Coop, if node $\boldsymbol{A}$ fails to deliver its packet to the access point, $\boldsymbol{B}$ is assumed to have correctly received it and will transmit the payload on behalf of the original source. The access point will decode $\boldsymbol{A}$'s frame by performing Chase combining on the two received packets, which convey information on the same data. The transitions for the chain of this protocol have a few subtle but important differences with respect to Basic. From state $(0,\ \boldsymbol{A},\ H_A,\ H_B)$, node $\boldsymbol{A}$ may unsuccessfully send a packet with probability $p(0) \cdot (1 - P_{\text{succ}}(H_A))$ and move into state $(1,\ \boldsymbol{B},\ H_A,\ H'_B)$. Note that H_A has not been updated, because the system needs to keep track of the SIR of the first transmitted packet to compute the success probability of the Chase combining decoding. When node $\boldsymbol{B}$ resends $\boldsymbol{A}$'s frame, the chain transitions into state $(1,\ \boldsymbol{B},\ H''_A,\ H''_B)$ and H_A is updated, because $\boldsymbol{B}$ will then transmit

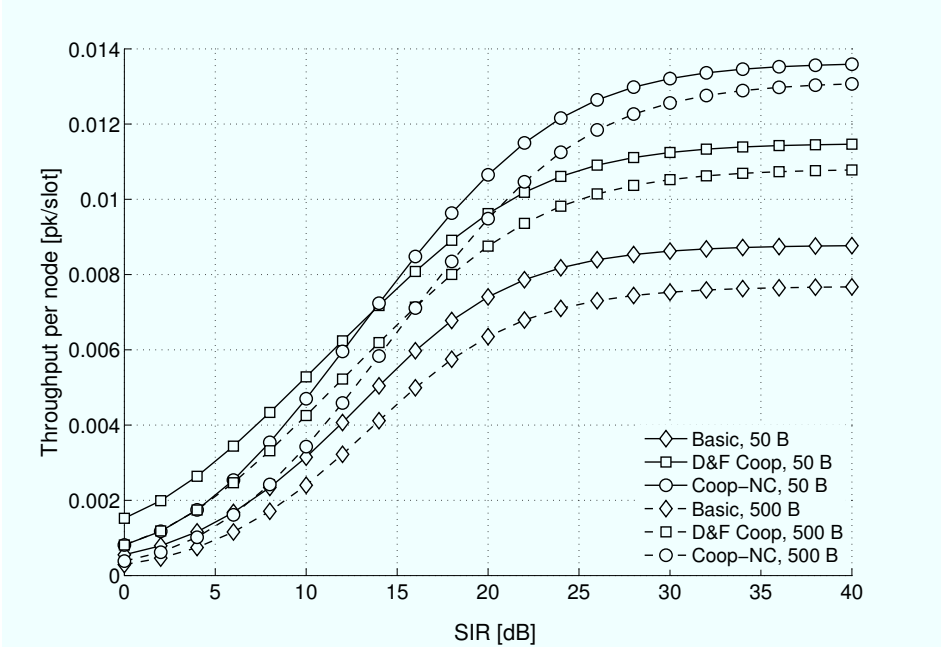


Figure 3.4. *Throughput per slot per node as a function of the average SIR and the packet size in bytes. The average initial window size is 64 slots*

a new packet of its own and there is no need to retain the old value of H_A, since it no longer affects the decoding process. However, the chain has not kept memory of how many slots have passed between $\boldsymbol{A}$'s and $\boldsymbol{B}$'s transmissions. Therefore, it cannot exactly compute the transition probabilities as far as $\boldsymbol{A}$'s channel is concerned. Hence, the channel evolution is approximated by assuming that the number of elapsed slots is equal the average backoff length, which is $(1 - p(0))/p(0)$.

Finally, Coop-NC's chain is identical to the one of D&F Coop, in that the allowed transitions are the same. However, the success probabilities do change, since during a retransmission MIMO_NC has to decode two packets rather than only one. These probabilities have been computed by simulating MIMO_NC decoding.

The final goal of our model is to evaluate the throughput per slot per node. This can be computed by associating to each transition a reward, equal to the number of correctly delivered packets [59]. In Basic and D&F Coop, every time a frame is successfully delivered, it brings a reward of one packet. In Coop-NC, a successful retransmission yields two data units (the corrupted one plus what the relay coded into the frame). Fig. 3.4 reports the analytical results on the throughput per slot per node of the described systems as a function of the average SIR. In order to make the scenario homogeneous with those analyzed in the simulation campaigns that will be discussed later for carrier sense-based medium access,

we have set the value $p(X)$ (one of the model parameters) such that the average backoff lengths in the model and in the simulations are the same. In the latters, see Section 3.5, the backoff is uniformly chosen in a window 128 slot long, thus the average duration is 64 slots. This implies that in the model $p(X) = 2^{-(6+X)}$, $0 \leq X \leq 2$ for Basic and $0 \leq X \leq 1$ for D&F Coop and Coop-NC. Moreover, the results for two packet sizes (500 and 50 bytes) are reported, so as to get insight on the influence of the payload dimension. Incidentally, we also point out that the depicted throughput is small in absolute value because it is normalized to the bandwidth, in order to keep the discussion general and not to be bound to a specific data rate.

A first observation that stems from Fig. 3.4 is that the magnitudes of the relative gain of Coop-NC over D&F Coop (about 25 %) and of D&F Coop over Basic are roughly the same. This suggests that our solution is able of doubling the gains of cooperative relaying with respect to plain ARQ. Such a qualitative observation will be supported by our simulation results both in carrier sense-based (Section 3.5) and in time division-based (Section 3.6) scenarios. It is also interesting to note that a rough upper bound on the throughput is given as follows: an isolated node with errorless channel would transmit on average one packet every $2/CW = 2/128 \simeq 0.0156$ slots, thus the throughput would be roughly 0.0156 pk/slot/node. The plot confirms this result, and further highlights that while all protocols are below this bound, Coop-NC is not far from it, giving merit to our proposal.

Finally, it can be pointed out that there exists a critical SIR value Λ^* such that Coop-NC outperforms D&F Coop and Basic for all SIRs larger than Λ^*. This fact suggests that network coded retransmissions may only be convenient when the destination has cached a version of the corrupted packet originally sent by the source with a minimum quality level. Such a result is rather intuitive, as in the event of very poor channels conditions over the direct link the chances of retrieving two data units with a single retransmission are low, while a more conservative approach may be able to guarantee the decoding of at least one of them. In other words, the intersection of the throughput curves stems from the fact that even though MIMO_NC preserves the diversity order of cooperative relaying, its bit error rate is slightly worse, especially in the low SINR region [48]. This discussion provides an important insight for the design of link layers that support hybrid cooperative-NC ARQ, suggesting to switch between traditional or coded retransmissions based on the SINR that affects the corrupted packets. In addition, the dependence of Λ^* on the frame size is rather weak. This underpins the choice of a switching threshold that does not change with the PDU length, and shows that the proposed mechanism is suitable for a variety of traffic types.

3.5 Cooperative-Network Coded ARQ in Carrier Sense Based Wireless Networks

In the first part of this chapter we have presented the key principles of hybrid cooperative-network coded ARQ, highlighting the advantages it can offer with respect to plain ARQ and decode-and-forward cooperation as well as studying its performance in simple interference-limited networks by means of an analytical model. In order to fully understand the potential of the proposed strategy, however, it is of paramount importance to both define efficient link layers capable of supporting this advanced form of cooperation and to investigate the behavior of such schemes in large and more realistic networking environments.

We start pursuing these goals by proposing in this section an implementation of network-coded relaying in carrier sense-based ad hoc wireless networks. Such scenarios are of interest for several reasons. In the first place, carrier sense is greatly widespread as a link layer policy, and it almost represents the paradigm for ad hoc networking, thanks to standards like the IEEE 802.11 (WLANs) [5] and the IEEE 802.15.4 (Low Rate PANs such as Wireless Sensor Networks) [60]. Moreover, the distributed and self-organizing nature that characterizes ad hoc environments poses several challenges for the design of link layers capable of effectively taking advantage of cooperative techniques, and thus represents a stressful and significant test for the evaluation of advanced ARQ solutions.

We initiate our study in Section 3.5.1.1 by reviewing the basics of carrier sense multiple access and by describing an extended version of it that supports decode-and-forward cooperation (Section 3.5.1.2). Then, in Section 3.5.1.3, we introduce and discuss a novel MAC, named Phoenix, that leverages both cooperative relaying and network coding techniques. Section 3.5.2, conversely, focuses on some extensive simulation campaigns that we have performed, considering both completely distributed networks with single hop (Section 3.5.2.1) and multi-hop flows (Section 3.5.2.2) as well as tree networks (Section 3.5.2.3).

3.5.1 Protocols Description

3.5.1.1 Carrier Sense Multiple Access

Throughout this chapter we consider as a reference for carrier sense medium access policies (CSMA) the IEEE 802.11 Distributed Coordination Function (DCF) without channel negotiation [5], which is based on Binary Exponential Backoff (BEB). According to this strategy, a node, before transmitting a packet, has to sense the channel to determine whether it is occupied by other communications already in place. To this aim, the terminal listens to the

medium for k slots[9], with k randomly chosen in $[1, 2^{CW_i}]$. If a power level higher than a carrier sense threshold is perceived during this operation, the node freezes its counter until the medium is sense idle again for at least a Distributed coordination function InterFrame Space (DIFS) period. When this condition is met, the terminal continues its countdown and once the backoff timer eventually expires, the packet is sent. When the transmission is accomplished, the source waits for a feedback from the destination of the frame. If the packet is successfully received, an acknowledgment message (ACK) is sent and the communication comes to an end. Otherwise, no response is sent, and the sender of the data unit performs another attempt by choosing a new contention window in $[1, 2^{CW'_i}]$, with $CW'_i = 2CW_i$ according to the BEB mechanism. This procedure implements a plain ARQ mechanism, and is iterated until either the message is successfully acknowledged or a maximum number of attempts (Short Retry Limit) is reached.

3.5.1.2 Cooperative Carrier Sense Multiple Access

In order to take advantage of decode-and-forward relaying, we have designed a protocol named Cooperative CSMA (CCSMA) that suitably extends CSMA. Let us refer again to the topology depicted in Fig. 3.1, and assume that a communication between $\boldsymbol{S}$ and $\boldsymbol{D}$ performed as per the rules described in Section 3.5.1.1 has failed. Any terminal other than the intended addressee that successfully decodes packet x (e.g., $\boldsymbol{R}_1$ or $\boldsymbol{R}_2$), caches it. At the destination side, two conditions may have occurred if the reception fails: i) the node is not able to decode the header of the packet; or ii) the header is correctly received but the payload is corrupted. The former case may be induced by harsh channel conditions or by the fact that the node was synchronized to another ongoing transmission. In this situation, since the intended destination did not gather any information on the packet being sent, no feedback can be provided and CCSMA resorts to a basic ARQ procedure. On the other hand, if ii) occurs,[10] node $\boldsymbol{D}$ becomes aware of the attempt performed by $\boldsymbol{S}$ and triggers a cooperative phase by caching the corrupted version of x and by transmitting a Not ACKnowledgement (NACK) frame asking for a relayed version of the payload. Terminals that receive the NACK packet and that have a cached version of the data, say $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, enter a distributed contention phase based on carrier sensing to elect the relay. Each candidate starts a backoff, whose duration n, in slots, is uniformly drawn in the set $\{0, 1, \ldots, CW_{rel} - 1\}$, where CW_{rel} is the length of the contention window for the relay election. When the count-

[9] For details on slot duration and contention window indexes please refer to Tab. 3.1

[10] We assume the packet header to be protected by a stronger FEC than the payload and to have a separate CRC.

down has reached the last slot, the node senses the medium and compares the aggregate received power P_{agg} with a reference value $\bar{P}$, possibly equal to the carrier sense threshold. If $P_{agg} \geq \bar{P}$, the relay candidate assumes that another terminal is performing the retransmission and therefore gives up the procedure, going back to its own activity. On the contrary, if the medium is sensed free (i.e., $P_{agg} < \bar{P}$), the node re-encodes x, sends a copy of the packet to the original addressee and then returns to its own activity. When a cooperative transmission takes place, the destination performs Chase combining and provides a feedback to the source node: if the reception succeeds, an ACK is sent, otherwise, a NACK is transmitted. In the latter case, no further cooperative phase is triggered, and $\boldsymbol{S}$ chooses whether to perform another attempt resorting to the BEB mechanism. The whole procedure (transmission by the source and potential cooperative phase) is iterated until the packet is correctly decoded at the destination or the SRL is reached.

Let us point out a few remarks on the distributed relaying scheme that we propose. First of all, in CCSMA cooperation is triggered by the destination node, introducing additional overhead with respect to plain CSMA (i.e., NACK packets). Nevertheless, this approach offers important advantages, as it prevents unnecessary relayed transmissions, thus saving resources and reducing the overall interference. Moreover, the described strategy increases the probability of success for cooperative phases, as the constraint of decoding the NACK packet allows only nodes that have sufficiently good channel conditions to the destination to be relay candidates.

Secondly, we observe that a relaying phase could fail for three main reasons: i) no cooperators may be available, ii) some collisions among relayed packets may occur, and iii) a potential relay could sense the medium busy and give up the contention because of aggregate network interference rather than because of an actual cooperative transmission. The first factor is related to both network density (i.e., topological availability of relays) and overall network activity (i.e., potential cooperators may not decode data coming from $\boldsymbol{S}$ or a NACK, due to interference). This issue is not specific to our strategy but rather affects all the decode-and-forward relaying procedures. As far as collisions among relayed packets are concerned, they are a drawback of the completely distributed approach that we employ. It is clear that the shorter CW_{rel} the higher the probability of having two or more cooperators choosing the same value for n. On the other hand, a long contention window would prolong the duration of the procedure, thereby reducing the advantages of fast failure recovery of cooperative ARQ. In our protocol we have chosen to use a maximum contention period that is much shorter than the backoff window employed by a pure CSMA strategy and that at

the same time keeps sufficiently low the probability of collision for the considered network densities (see Section 3.5.2.1). Moreover, we try to enhance the probability of success for a relaying phase by favoring nodes that experience good channel conditions towards the destination. In fact, the threshold $\bar{P}$ represents the average power of all data packets received from $\boldsymbol{D}$. Therefore, if a NACK is decoded with a power higher than $\bar{P}$, the fading between the potential cooperator and $\boldsymbol{D}$ is favorable.

Let us now focus on the criterion chosen to give up the contention. The described carrier sense mechanism aims at informing all the contending terminals as soon as a relay accesses the medium. Nevertheless, as pointed out earlier (iii), a node always overhears some amount of overall interference due to current transmissions in the network, and there is no way to distinguish a specific message out of it. Therefore, if the noise and interference level exceeds the threshold, the contention may be abandoned even if no other node is acting as a relay. In order to reduce the impact of this factor, we ask a node to sense the medium only in the last slot of its backoff window.[11] Suppose, in fact, that a neighboring communication is in place when the potential cooperator receives the NACK frame but ends before the countdown is over. With the usual carrier sensing scheme, the terminal would unnecessarily abandon the contention, while with our approach cooperation may still take place. On the other hand, we would like to observe that giving up cooperation if the overall interference level is still high when the contention window actually expires may be a good solution. This stems from the spatial correlation of interference. In fact, when a potential relay decides to abort its transmission, the destination, which is typically one of its neighbors, is likely to be affected by unfavorable conditions as well. Should the node cooperate anyway, not only would its transmission experience a very low SINR at the receiver but also the other ongoing communications might experience a collision, with detrimental effects on the network performance.

In conclusion, it is worth noticing that although the protocol could easily be extended to more cooperative retransmissions per each phase, we have decided not to follow this approach. This choice stems from the observation that a failure of a cooperative delivery is typically due to harsh interference conditions at the destination. Therefore, other attempts are unlikely to succeed unless performed after a sufficiently long interval, and in such a condition the BEB mechanism turns out to be more effective.

[11] This solution does not worsen the probability of collision among relays. In fact, if a terminal accesses the medium, its transmission lasts much longer than the relay backoff window chosen by any other contending node.

3.5.1.3 The Phoenix Protocol

The medium access control of CCSMA can be further developed so as to implement the hybrid cooperative-network coding ARQ paradigm described in Section 3.2. The outcome of this refined design is a protocol called Phoenix. Let us consider again the situation of Fig. 3.1, and suppose that node $\boldsymbol{D}$ has received the header of packet x without being able to decode the payload. If the quality of the cached frame is extremely poor, the decoding probability for a network coded phase might become too low with respect to pure decode-and-forward relaying. [12] Therefore, before sending a request for cooperation, $\boldsymbol{D}$ checks the average SINR that characterized the corrupted frame. If this value is below a given threshold Λ_{Th}, the terminal sets a one-bit field *NACK_flag* of the NACK packet to 0. Otherwise, the flag is set to 1. Note that this strategy resembles a rate adaptation policy, since it links the number of data units coded together with the SINRs of the channels. Nodes that receive the NACK frame and that have correctly decoded packet x start the relay election phase of CCSMA. The terminal that wins the contention (i.e., that senses the medium free at the end of the backoff), say $\boldsymbol{R}_1$, determines which type of retransmission is to be performed. If the *NACK_flag* is set to 0 or the relay has no packets in its queue, the CCSMA cooperative procedure takes place. On the contrary, if the destination allowed a coded transmission and $\boldsymbol{R}_1$ itself has traffic to serve, a hybrid cooperative-NC phase is initiated, distinguishing two conditions as follows.

A) *The relay has at least one packet y for $\boldsymbol{D}$ in its queue.* In this case, the cooperator generates a linear combination of x and y following the network coding principles and transmits the obtained PDU. At the end of the reception, the destination sends two feedback frames, either ACK or NACK: the first one is addressed to the original source $\boldsymbol{S}$ and regards x, while the second one informs $\boldsymbol{R}_1$ about the decoding of y. [13]

B) *The relay has no packets for $\boldsymbol{D}$ in its queue, but it has a packet y addressed to another node.* In such a condition, the cooperator tries to determine whether a hybrid retransmission is appropriate. Indeed, should the addressee of y, say $\boldsymbol{R}_2$, have no cached version of x (e.g., because it did not synchronize to the transmission by $\boldsymbol{S}$), it would not be able to extract the payload y from a linear combination of x and y. In this case, not only would a cooperative-NC phase prevent the relay from successfully serving its traffic, but also it would worsen the performance at $\boldsymbol{D}$'s side with respect to Chase combining.

[12] Let us recall again that with MIMO_NC, the error probability for a hybrid retransmission (i.e., the receiver jointly decodes x and y) is slightly higher than the one of Chase combining (i.e., the receiver combines two copies of x), yet the diversity order is preserved [48].

[13] Notice that if y is not successfully decoded, the relay keeps the packet in its queue for later transmission.

Therefore, the cooperator transmits a Request To Send (RTS) packet addressed to $\boldsymbol{R}_2$, containing a field that uniquely identifies x. $\boldsymbol{R}_2$ replies with a Clear To Send packet (CTS) only if the RTS is decoded and the node has in its cache either a correct copy of x or a corrupted version with SINR above Λ_{Th}.[14] If the CTS is not received, the relay falls back on a pure cooperative transmission, as in CCSMA. On the contrary, if the handshake succeeds, the cooperator sends a linear combination of x and y. This message is followed first by an ACK/NACK by $\boldsymbol{D}$ addressed to $\boldsymbol{S}$, according to the outcome of the decoding of x. Then, $\boldsymbol{R}_2$ does the same, informing $\boldsymbol{R}_1$ of the reception of y. In case of a NACK, node $\boldsymbol{S}$ decides whether to perform another attempt after a suitable backoff interval, while $\boldsymbol{R}_1$ puts y back in its queue for a later time.

Phoenix gives priority to type A) retransmissions since they generate less overhead and interference than type B).[15] Moreover, we remark that the aim of the RTS/CTS procedure is twofold. First, as discussed, it prevents hybrid retransmissions that would have no chance to succeed. Secondly, in a CSMA-based MAC, the exchanged negotiation messages forestall concurrent communications in the neighborhood of both the cooperator and its destination.[16] The consequent reduction of interference is likely to significantly enhance the success probability for the cooperative phase. These reasons justify the additional handshake overhead.

3.5.2 Simulation Results

To evaluate the benefits offered by hybrid cooperative-NC ARQ in complex and realistic scenarios, we have tested the performance of Phoenix by means of extensive simulations using the Omnet++ modeler [61]. Our scheme has been compared to two benchmarks, namely plain CSMA and CCSMA. In order to keep a uniform notation throughout the chapter, the three schemes are referred to in the plots that will be presented as *Basic* (CSMA), *D&F Coop* (CCSMA) and *Coop-NC* (Phoenix), respectively.

All the protocols have been studied in a wireless environment subject to correlated Rayleigh fading with Doppler frequency equal to 40 Hz, corresponding to a speed of 5 m/s at 2.4 GHz and path loss with exponent $\alpha = 3.5$. The values of the Short Retry Limit (SRL) have been chosen so that all the MAC schemes offer a similar reliability for single

[14] In Phoenix, unlike CCSMA, a terminal caches the latest received packet whose header has been correctly decoded (be the packet addressed to it or not), regardless of the outcome of the payload decoding.

[15] In type A) retransmissions, since traffic for $\boldsymbol{D}$ is preferred over flows for other nodes, unfairness may arise. However, this effect is confined to strongly asymmetrical topologies.

[16] Notice that terminals in such regions may not be forced to silence by the transmission between $\boldsymbol{S}$ and $\boldsymbol{D}$.

Table 3.1. *Parameters used in our simulations of hybrid cooperative-NC ARQ in CSMA-based networks*

Transmission power	10 dBm
Noise Floor	-102 dBm
Carrier sense threshold	-100 dBm
Carrier sense threshold for relay contention, $\bar{P}$	-100 dBm
Detection threshold	-96 dBm
Path loss exponent, α	3.5
Carrier Frequency	2.4 GHz
Maximum Doppler shift	40 Hz (5 m/s)
Data Rate, B	6 Mbit/s
Slot, DIFS, SIFS duration	20, 128, 28 μs
Initial maximum contention window	128 slots
Number of slots used for relay contention, CW_{rel}	32
Short Retry Limit - *D&F Coop* and *Coop-NC*	2
Short Retry Limit - *Basic*	3
Queue size at the MAC layer, Q	16
Minimum SINR to trigger a cooperative-NC phase, Λ_{Th}	3 dB
DATA header Basic - *D&F Coop*	272 bit
DATA header *Coop-NC*	280 bit
Payload, L	2000 bit
ACK/NACK/CTS	112 bit
RTS	160 bit
Simulation Time	30 s
Simulation Transient (metrics not collected)	5 s

hop flows (see Tab. 3.2). In particular, the SRL for plain CSMA has been set to 3, while 2 independent transmission attempts are sufficient for CCSMA and Phoenix, thanks to the advantages offered by cooperative relaying techniques. The initial maximum contention window 2^{CW_i} is 128, because if a smaller value was selected, the collision probability in the large cardinality networks considered here would become so high that all the protocols would show very poor performance, and basic inefficiencies would simply hide all the phenomena we are interested in. As far as the relay election is concerned, the maximum length of the contention window, CW_{rel}, has been identified considering two opposite trends, as discussed in Section 3.5.1.2. On the one hand, the larger the CW_{rel} value, the lower the collision probability among cooperators. On the other hand, a lengthy backoff interval reduces the gains of cooperative relaying because of the longer failure recoveries. Starting from these remarks, we have performed some preliminary simulations, and we have determined that a reasonable tradeoff between the two factors for the topologies under study was represented by $CW_{rel} = 32$. Finally, in our simulations we allow a hybrid cooperative-NC phase only if the node asking for a retransmission has in its cache the corrupted frame with SINR $\geq \Lambda_{Th} = 3$ dB. This value stems from the BER/SINR tables for MIMO_NC, as with SINRs lower than 3 dB for the cached packet, the decoding probability of a coded retransmission falls below 2/3. A complete list of the parameters used in our simulation campaigns can be found in Tab. 3.1.

All the results reported in this section have been averaged over multiple simulations, so that the 95% confidence interval never exceeds 3% of the estimated value. The duration of each simulation was chosen long enough to stabilize the results.

3.5.2.1 Single Hop Networks

We start our investigation of the effectiveness of hybrid cooperative-NC ARQ by focusing on ad hoc scenarios with single hop traffic flows. To this aim, we consider networks composed by 35 nodes spread over a a $260 \times 260\, m^2$ area, each of which generating packets addressed to its neighbors according to a Poisson traffic model with intensity λ (pk/s/node). This configuration allows multiple simultaneous communications in the network, and tests the protocols when hidden terminals and external interference are present. Such a setting is therefore meaningful to highlight the performance of MAC schemes in harsh medium contention conditions.

The first metric that we analyze is the Packet Delivery Ratio (PDR), depicted against the nominal load in Fig. 3.5. As a general remark, we notice that all the schemes offer a similar

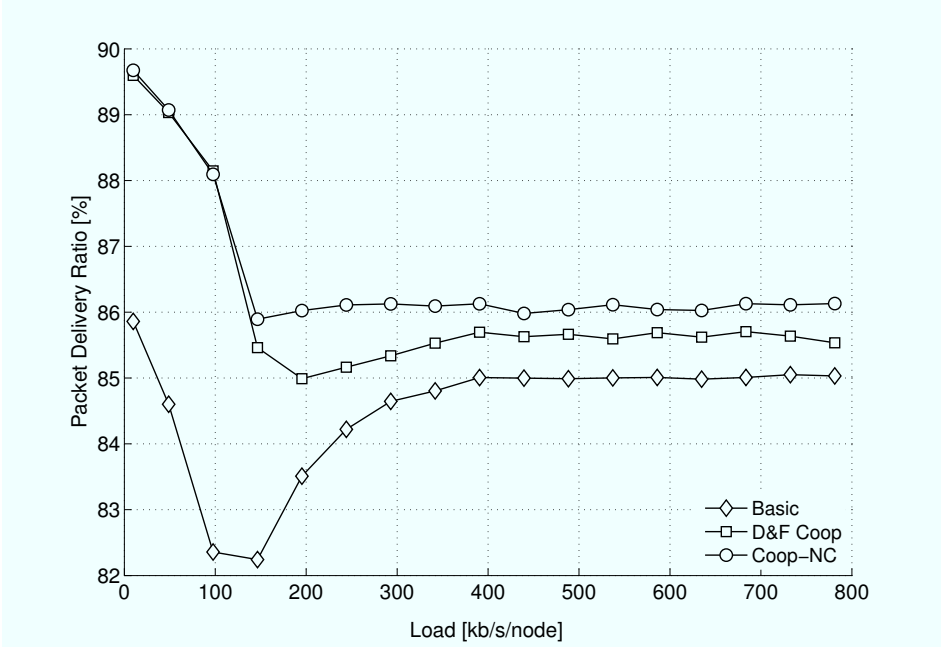


Figure 3.5. *Single hop flows: packet delivery ratio vs nominal load [kb/s/node].*

reliability level at saturation, thus justifying the choice of the SRL values discussed earlier. Furthermore, the figure shows that Phoenix performs better than CCSMA. The reason is that the coded retransmission system that we propose enables to safely deliver more packets by means of reliable relays. Indeed, when a cooperative phase succeeds, CCSMA can increase the numerator of the PDR (i.e., the number of acknowledged packets) by one unit. Instead, Phoenix increases it by two units and the denominator (i.e., the number of frames which are transmitted at least once) by one. Since the contention system often leads to a good choice of the relay, the retransmission is successful with high probability, and in the end this leads to a higher PDR. It is also interesting to observe that all the curves reach a minimum value for rather low loads and then tend to rise, before stabilizing at saturation. This stems from the well known unfairness that characterizes carrier sense based networks that rely on the Binary Exponential Backoff mechanism. Nodes that are located in more favorable positions, i.e., closer to their destinations, experience on average a very high success rate. Thus, as the load increases, these terminals quickly deliver their traffic and tend to access the channel more and more often, while nodes that suffer from worse conditions get stuck in longer backoff cycles and therefore contribute less to the overall amount of transferred packets. As a result, a larger fraction of the acknowledged traffic comes from those terminals that exhibit a better delivery ratio, biasing the metric as reported in the plot.[17] While such an effect

[17]This interpretation has been confirmed by other simulation results not reported here due to space con-

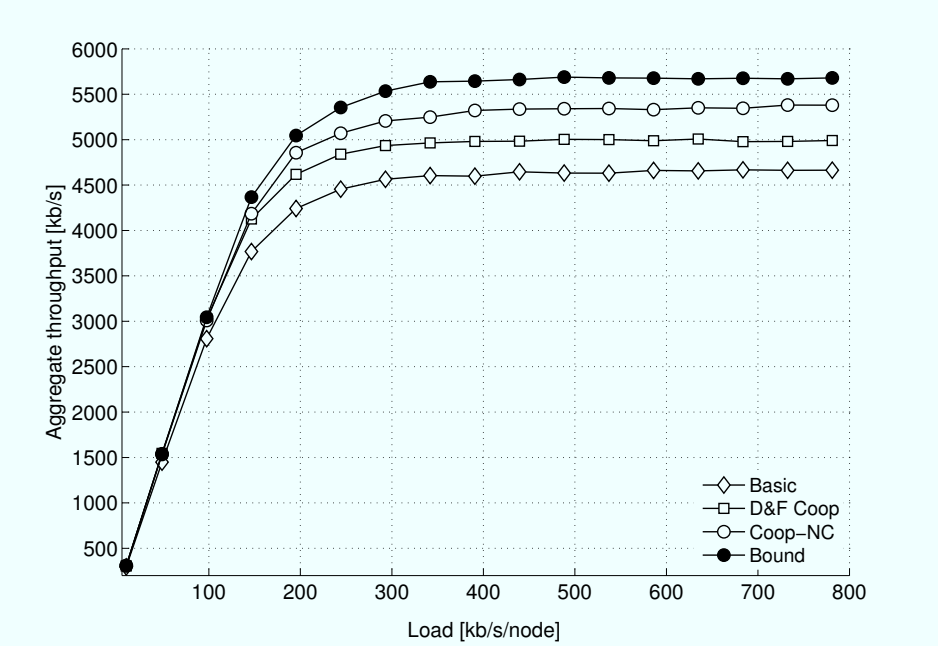


Figure 3.6. *Single hop flows: aggregate throughput vs nominal load [kb/s/node]*

affects all the considered policies, its impact is less pronounced for cooperative solutions. As a matter of fact, relaying mechanisms help exactly those terminals that find it difficult to deliver their own traffic, favoring a fairer distribution of the resources in the network.

Fig. 3.6 reports the aggregate throughput against the nominal load per node. Cooperative relaying techniques show their beneficial effect already at relatively low loads. Their faster and more reliable failure recovery procedures boost the performance of both CCSMA and Phoenix with respect to plain CSMA, with at least a 10% gain at sufficiently high loads. On the other hand, the impact of hybrid cooperative-NC phases becomes more evident as traffic increases, since the more the enqueued packets, the higher the number of coded retransmissions. The capability to serve additional traffic during relaying phases turns out to be profitable in terms of aggregate throughput: at saturation Phoenix outperforms CSMA by 18% and CCSMA by almost 10%. Fig. 3.6 can also be analyzed in light of the results of Section 3.4. Our analytical model predicted similar throughput percentage gains for Phoenix over CCSMA and for CCSMA over plain CSMA. Such trend is clearly confirmed by the curves of Fig. 3.6. Incidentally, we notice that the gap between the schemes as computed in Section 3.4 is higher than the one achieved in our simulation campaigns. This difference stems from medium access issues such as header losses and cooperator unavailability

straints, which showed how in perfectly symmetrical topologies (i.e., circular networks) such trend completely disappears.

that could not be considered in the analytical model due to complexity but that significantly affect the performance of the protocols in a CSMA environment, and that will be discussed in detail in Chapter 5. To investigate the effectiveness of Phoenix's approach to hybrid cooperative-NC phases, we have also tested a reference protocol, named *Bound*, that implements the proposed medium access policy taking advantage of some ideal assumptions. First, all the cooperative retransmissions performed using Bound succeed irrespective of the SINR (i.e., x is decoded in case of Chase combining or both x and y are retrieved in case of hybrid retransmission). Thus, the protocol allows each cooperative phase to be network-coded by always setting the *NACK_flag* to 1. Secondly, a relay candidate is assumed to be aware of which neighbors have cached a copy of x. Exploiting this information, cooperative-NC retransmissions involving secondary destinations can take place without the RTS-CTS handshake procedure implemented in Phoenix. In view of these two properties, the considered scheme, while obviously impossible to realize in practice, represents an upper bound for the class of protocols that implement hybrid cooperative-NC ARQ relying on a distributed CSMA-based contention for the choice of the relay. We report the results for Bound only for the aggregate throughput, since the other metrics would exhibit similar trends. Fig. 3.6 shows that the bound (black markers line) only offers a limited improvement with respect to Phoenix, approximately 5% at saturation. This highlights how our solution indeed represents a good realization of hybrid cooperative-NC ARQ techniques in a CSMA environment, since it achieves most of the available gains at low cost in terms of protocol complexity.

While Fig. 3.6 dealt with the overall data rate the whole network can support, it is insightful to understand how the throughput is distributed among the nodes. In order to evaluate this aspect, we introduce a reference throughput $\bar{\tau}$ and we divide the users in classes with respect to this target. Let n_c be the number of terminals that are inhibited due to the carrier sense mechanism by the transmission of a given node. It stems that these users cannot serve their traffic during the current communication, and they must defer their access to the channel. In such a system, we define $\bar{\tau}$ as the saturation nominal load per node $\bar{\lambda}$ multiplied by the payload length L. In turn, $\bar{\lambda}$ is the minimum value of λ that fully utilizes the data rate B, i.e., $\bar{\lambda}\, n_c L_e \triangleq B$, where L_e is the shortest time required to perform a complete data transmission multiplied by the data rate. With the parameters used in our simulations, $\bar{\tau} = 150\,\text{kb/s}$. This model does not take into account imperfections of the MAC protocol, like collisions. Therefore, in a real implementation, nodes will achieve only a fraction of $\bar{\tau}$. Fig. 3.7 depicts the number of users that achieve a certain share S_τ of the target throughput,

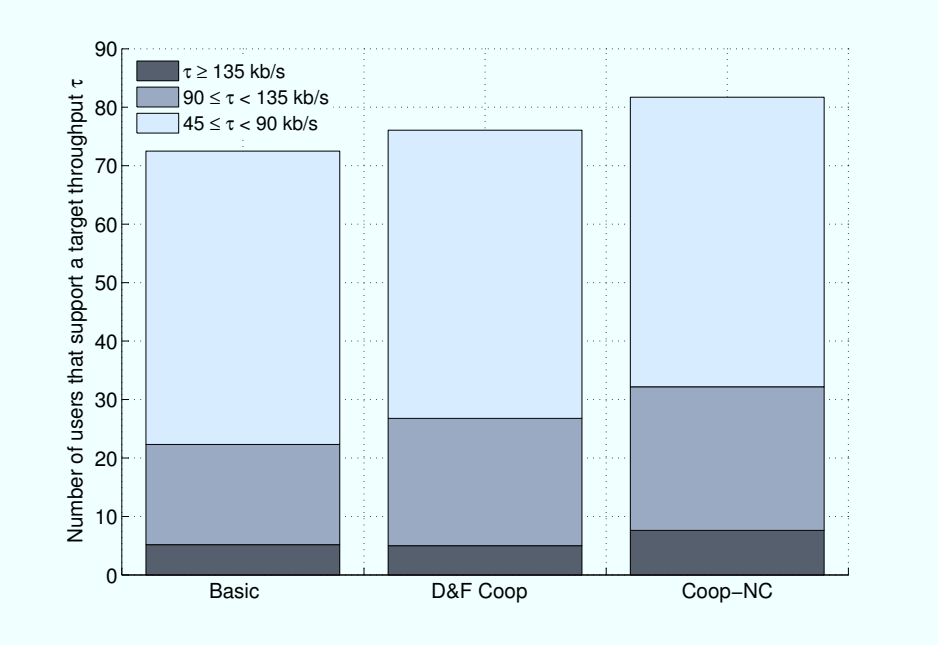


Figure 3.7. *Single hop flows: number of users that enjoy a given QoS.*

namely $S_\tau = 90,\ 60,\ 30\%$. For example, the darkest bars represent the number of nodes that enjoy at least 90% of $\bar{\tau}$. The improvement offered by Phoenix over its competitors is twofold. On the one hand, the share of terminals that support the minimum reference throughput increases by 8% with respect to CCSMA and by 14% with respect to plain CSMA. On the other hand, our protocol boosts the number of nodes with medium and high QoS by as much as 21% and 45% respectively. These results show that the combination of cooperation and NC can guarantee a minimum service level to a larger population. Let us focus on the trends for CCSMA and CSMA. Two remarks can be made: i) CCSMA increases the number of nodes that support the minimum throughput; ii) the cardinality of the highest QoS class for CCSMA is slightly smaller than the one for CSMA. This offers an interesting insight on the impact of cooperation. With a decode-and-forward approach, relay nodes spend some of their resources in order to help other terminals. Hence, not only do cooperators reduce their performance, but also terminals that benefit from this help become more aggressive, as their higher success rate leads them to contend for the channel more often. Both these factors are detrimental for relays. We can then infer that cooperation redistributes the resources in the network at the expense of users with high QoS. This effect, on the contrary, has no impact on Phoenix: MIMO_NC does not disadvantage relays, since they can deliver their own traffic as well.

Another metric of interest is the average delivery latency at the link layer, reported in

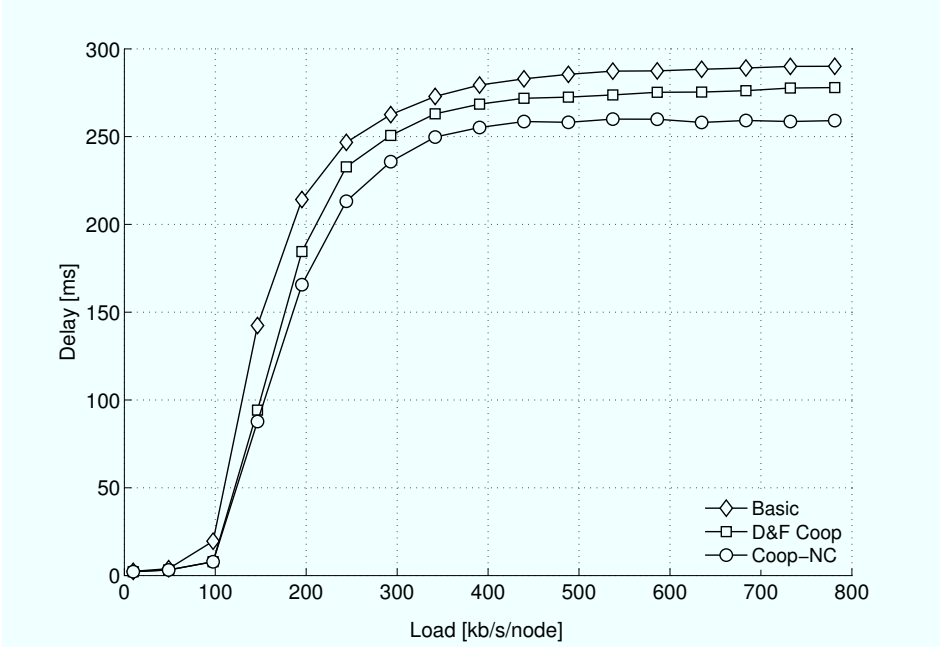


Figure 3.8. *Single hop flows: average delivery delay for successfully acknowledged packets vs nominal load [kb/s/node].*

Fig. 3.8. Phoenix reduces delay at high loads by about 6% over CCSMA and 13% over CSMA. CSMA undergoes a higher delay because of the three retransmission phases, which, in turn, are required in order to have a comparable reliability with the other protocols. The key reason that enables a significant delay reduction by Phoenix over CCSMA and CSMA is the transmission of additional data packets by means of NC. Indeed, these frames sent in an opportunistic fashion have much lower delay than the others, as they basically have no latency due to MAC contention procedures and are therefore subject only to queuing delay. Since the latency due to MAC contention delay tends to be the predominant factor, even a few such packets can significantly improve the mean delay.

Fig. 3.9 shows the average transmission energy consumption per successfully acknowledged information bit against the nominal load. First of all, we notice that the CSMA curve decreases as the traffic increases beyond a certain point. This effect is due to the unfairness that characterizes medium access schemes based on carrier sensing. As discussed, with such protocols nodes that experience poor average link conditions, e.g., because they are far from their destinations, incur longer backoff cycles and therefore tend to access the medium with lower frequency. On the contrary, terminals that enjoy favorable positions are inclined to transmit with shorter contention phases and with higher success probability, grabbing a larger share of the bandwidth. Such a behavior becomes more evident as traffic increases.

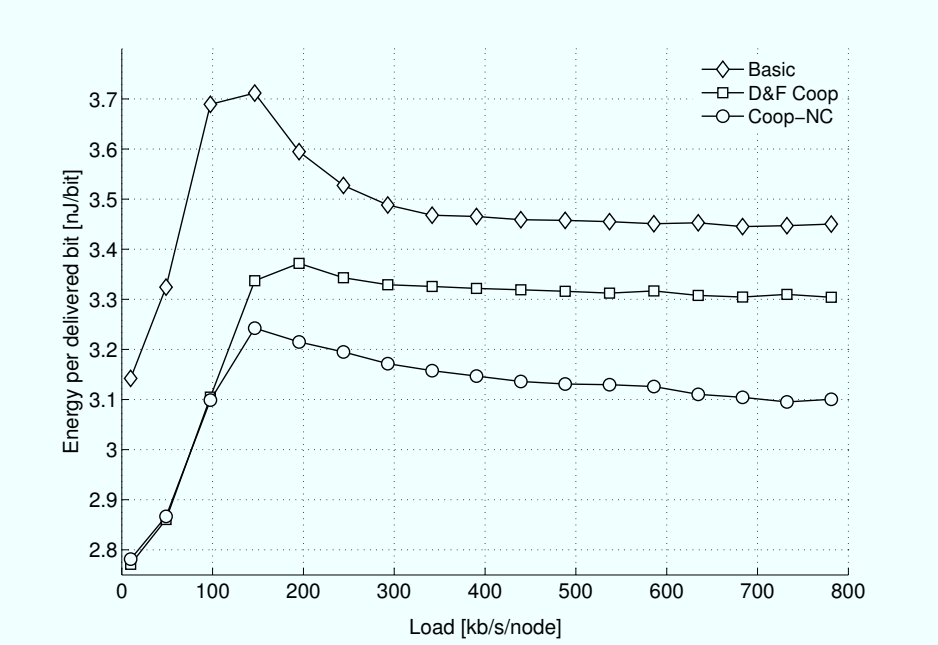


Figure 3.9. *Single hop flows: transmission energy consumption per successfully delivered data bit vs nominal load [kb/s/node].*

Therefore, at high nominal loads most of the transmissions are performed by nodes that belong to the latter class, resulting in a lower average cost in terms of energy. The impact of unfairness in CSMA leads to an almost 10% drop in energy consumption from low to high loads. On the other hand, this effect is far less pronounced for CCSMA. This stems from the beneficial influence of cooperation, that shortens ARQ phases also for terminals that experience bad channel conditions, preventing them from being stuck in energy-costly backoff cycles. As far as Phoenix is concerned, a drop in energy consumption at high loads can be noted. However, unlike for CCSMA, this trend is not due to unfairness, but to the higher number of coded retransmissions that take place in such conditions, which make it possible to deliver packets at no additional cost in terms of energy. Two further observations can be made on Fig. 3.9. First, at low traffic rates, the cooperative protocols outperform CSMA by as much as 20%. In these conditions, all the terminals manage to access the medium, and the higher number of retransmissions required to successfully deliver a packet, due both to the lack of spatial diversity and to the higher SRL, highlights the energy inefficiency of plain CSMA. Secondly, at high loads Phoenix provides some improvement over CCSMA, which stems from the capability of our protocol to deliver additional information bits during cooperative phases at no cost in terms of energy and bandwidth.

Finally, Fig. 3.10 shows the percentage (computed over the number of failed transmis-

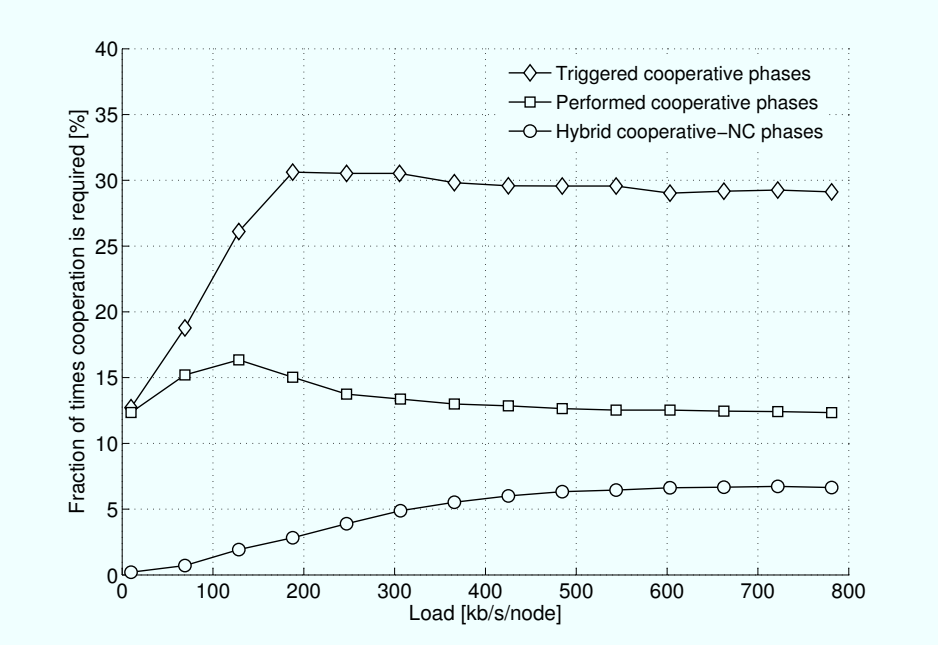


Figure 3.10. *Single hop flows: percentage of times in which cooperative behavior is shown vs nominal load [kb/s/node].*

sions) of data packets that may support a cooperative retransmission (*triggered cooperative phases*), i.e., those for which the destination sends a NACK; of contention phases that are actually carried out (*performed cooperative phases*), i.e., a relay has retransmitted the corrupted data packet; and of times in which network coded retransmissions take place (*hybrid cooperative-NC phases*). Several conclusions can be inferred. Firstly, in 44% of the requested cooperative phases, some node actually performs the retransmission. The trace files have shown that in a large number of cases (40% of all requested retransmissions) there are no relays. This is partly due to topological reasons (there may be no relay close to source and destination) but in most cases the data packet or the NACK have not been received, because the relays may be decoding another packet in the meantime, or since fading caused the packet loss. In the remaining 16%, there were relays, but carrier sense prevented them from transmitting. We also point out that the contention system itself is rather efficient, since in 73% (=44%/60%) of all contention phases where some nodes have engaged in the relay election at least one cooperator actually transmits. This discussion suggests that the highlighted issues are not related to the MAC that we have implemented, as confirmed by the performance of the Bound reported in Fig. 3.6, but rather stem from the intrinsic nature of CSMA. Nonetheless, our scheme has proven able to double the gains with respect to decode-and-forward cooperation over plain ARQ. It is remarkable that the improvements shown so far

have been achieved by acting on only the cooperative packets, that is to say 16% of all the traffic. Hence, even if MIMO_NC is actually used only a few times, it delivers an interesting performance bonus. A more in depth discussion of these issues will be carried out in the following chapters.

In our work, we have also studied the dependence of the different schemes on some protocol parameters, namely SRL and payload length. The results of these analyses are reported at saturation in Tab. 3.2. Let us first focus on the impact of the former parameter. As expected, higher values of SRL raise the reliability, i.e., the Packet Delivery Ratio (PDR), and the average latency, due to the increased number of retransmission attempts. Incidentally, we notice that Phoenix slightly improves the PDR over CCSMA for all values of the SRL. This effect again stems from the high reliability of cooperative-NC phases induced by a suitable choice of Λ_{Th}. Moreover, the results of additional studies that we have performed, not reported here due to space constraints, have pointed out that fairness decreases as the value of SRL increases. This can be explained observing that the larger the number of retransmissions, the longer the backoff cycles that nodes are likely to enter, and therefore the less homogeneous the bandwidth distribution in the network. The impact of unfairness can be seen in the decreasing trends for energy consumption, outage delay and outage throughput. [18] We remark that Phoenix outperforms its competitors for all the metrics in all the considered scenarios, but these advantages mildly shrink as the SRL increases. This is an effect of the longer backoffs that nodes experience for higher SRL, which improve the temporal diversity at the receiver. As a consequence, the impact of spatial diversity is reduced. Incidentally, we stress that Tab. 3.2 confirms how the reference values of SRL used for CSMA and for the cooperative protocols in our simulations have been chosen in order to provide a comparable PDR.

Let us now consider the performance dependence of the MAC schemes on the payload length. As is reasonable to expect, some metrics (like PDR or outage delay) worsen as the packet length increases. In addition, it is important to notice that the gains of Phoenix over the other protocols are roughly invariant with respect to the frame size. Therefore, our scheme works well for different packet dimensions, as the analytical model of Section 3.4 predicted.

Finally, the protocols behavior has been also studied varying the network density. In particular, besides the reference case (35 nodes), two more settings have been analyzed, with 25 or 56 nodes, in the same area. The average number of neighbors (AD) is 3.5, 5.5 and 7.5

[18] The outage throughput (delay) is defined here as the 20-th (80-th) percentile of the throughput (delay) experienced by all the nodes in the network.

Table 3.2. *Parametric Studies on SRL (with payload set to 2000 bits) and payload length (with $SRL_{Basic} = 3$, $SRL_{D\&FCoop} = 2$, $SRL_{Coop-NC} = 2$) for the single hop scenario in carrier sense based networks. All the results are reported at saturation load.*

Protocol	Parameter	PDR [%]	Delay [ms]	Energy [nJ/bit]	Outage Delay [ms]	Outage Thr. [kb/s]
Basic	SRL 3	85.0	289	3.45	505	72.2
	SRL 4	90.9	323	3.30	597	65.7
	SRL 5	95.0	348	3.09	730	58.0
	SRL 6	97.5	369	2.90	818	52.6
D&F Coop	SRL 2	85.5	277	3.32	458	80.1
	SRL 3	93.6	316	3.27	558	75.5
	SRL 4	98.1	333	3.12	616	73.7
Coop-NC	SRL 2	86.1	260	3.10	416	91.8
	SRL 3	93.6	298	3.07	531	79.9
	SRL 4	97.5	322	2.99	601	77.2
Basic	1000 bits	88.7	211	1.79	337	112
	2000 bits	85.0	289	3.45	505	72.2
	2500 bits	83.5	328	4.37	570	60.6
D&F Coop	1000 bits	89.7	204	1.75	312	128
	2000 bits	85.5	277	3.32	458	80.1
	2500 bits	83.5	311	4.14	520	69.2
Coop-NC	1000 bits	90.1	193	1.65	280	148
	2000 bits	86.1	260	3.10	416	91.8
	2500 bits	84.8	296	3.88	492	72.9

respectively. We do not report the detailed results of these investigations due to space constraints. However, some conclusions can be inferred. First of all, the relative performance is approximately the same for all configurations, showing that the system is robust to node density. This is a non trivial result, because node density affects two competing factors in Phoenix. More nodes imply higher interference and more collisions, which beset any protocol. On the other hand, a denser network offers more potential relays. By and large, it turns out that these factors balance themselves out, and thus Phoenix is a suitable choice for a wide range of scenarios.

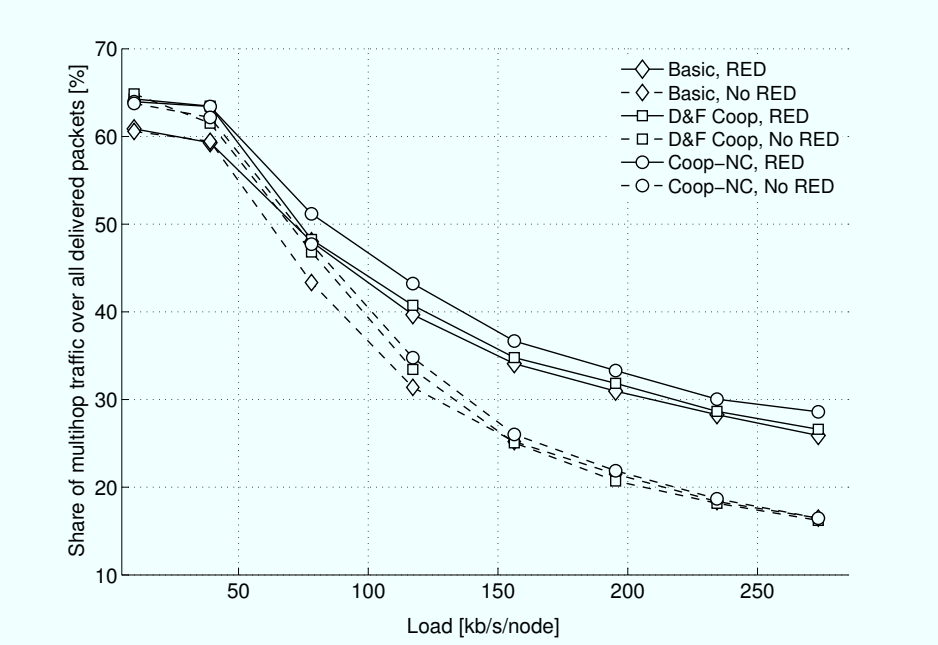


Figure 3.11. *Comparison of the performance of the protocols with and without Random Early Depletion (RED): share of multihop traffic successfully delivered over total delivered traffic.*

3.5.2.2 Multihop Networks

In order to get a further insight on the performance of Phoenix and its competitors, we have studied them in a multihop environment. In this case, 25 nodes are deployed in a $200 \times 200\, m^2$ square. The considered topologies are connected, and all-to-all Poisson traffic with intensity λ (pk/s/node) is generated. We assume the routing tables to be known a priori at each terminal. In such a scenario, data packets may not be correctly delivered because of PHY related impairments (e.g., fading and interference) or MAC level problems like buffer overflow. It is well known that the latter issue tends to reduce the share of successfully delivered multihop traffic in favor of single hop. However, the aim of our analysis is to specifically evaluate the advantages brought by hybrid cooperative-NC retransmissions on longer routes.

Therefore, to stress the impact of multihop flows, we have implemented a Random Early Detection (RED) policy [62]. We divide packets processed at terminal $\boldsymbol{A}$ in three classes: $\mathcal{C}_1$ identifies traffic that is generated at the current node; $\mathcal{C}_2$ contains frames that have undergone one or two hops to reach $\boldsymbol{A}$; while frames that have traveled more than two hops belong to $\mathcal{C}_3$. Furthermore, letting Q be the size of the buffer at the MAC layer, we associate a threshold value $T_i = \alpha_i \cdot Q$ to each class. When a packet of class $\mathcal{C}_i$ is received from the network layer, our MAC schemes check the number n of frames currently in the buffer. If

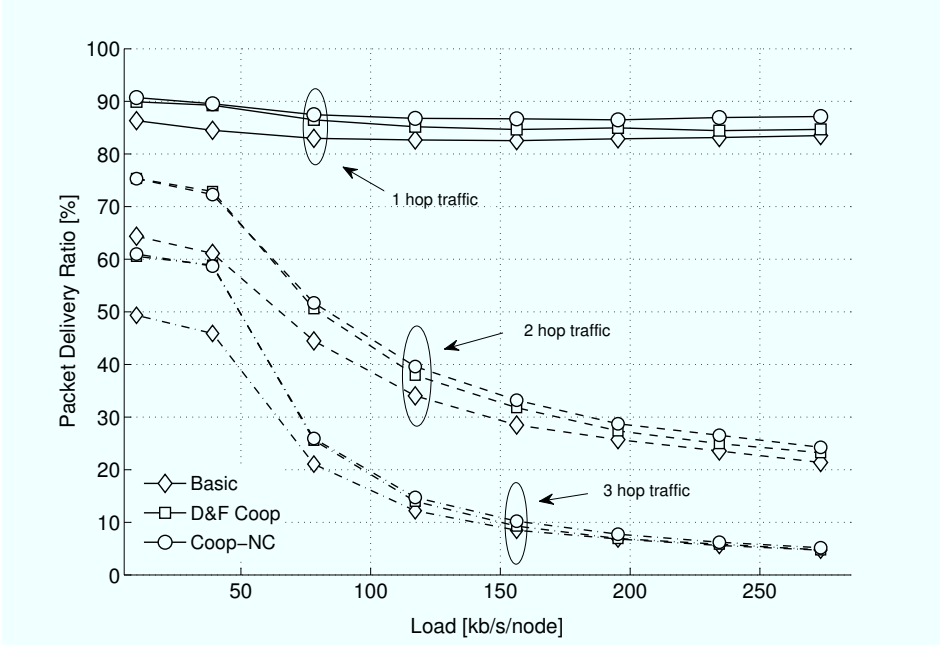


Figure 3.12. *Multihop networks: packet delivery ratio vs nominal load [kb/s/node]. Each set of curves shows the behavior for a specific route length in hops.*

$n \leq T_i$, the packet is inserted in the queue for later transmission. Otherwise, the packet is accepted with probability $1-p_d$ and discarded with probability p_d, where $p_d = (n-T_i)/(Q-T_i)$. Clearly, the larger α_i (and hence T_i), the smaller the rejection probability p_d. In our studies, $\alpha_1 = 1/2$, $\alpha_2 = 3/4$ and $\alpha_3 = 1$. The beneficial effect of RED is apparent in Fig. 3.11, where the share of multihop traffic increases for Phoenix from 16% to 30% at high loads.

Fig. 3.12 presents the PDR achieved by the considered MAC policies for different route lengths, depicted against the nominal load per node. Single hop traffic enjoys a high reliability with all the protocols, whereas multihop paths, as expected, incur more losses. The impact of cooperation is apparent: CCSMA and Phoenix improve the PDR with respect to CSMA by 10% for two hop routes and by up to 15% for paths composed by three hops. The benefit stems from the better failure recovery capabilities of relaying techniques with respect to plain ARQ, that help to keep alive multihop flows. This effect is even magnified by hybrid cooperative-NC procedures, as shown by the fact that Phoenix outperforms CCSMA regardless of the load. The ability to exploit relaying phases to serve additional traffic has two main consequences: not only do coded packets proceed closer to their destination, but also the saturation of MAC queues is slowed down, positively affecting the whole network. Moreover, the high reliability offered to multihop traffic is particularly beneficial to the overall system, since it repays the network for the efforts made to route packets through

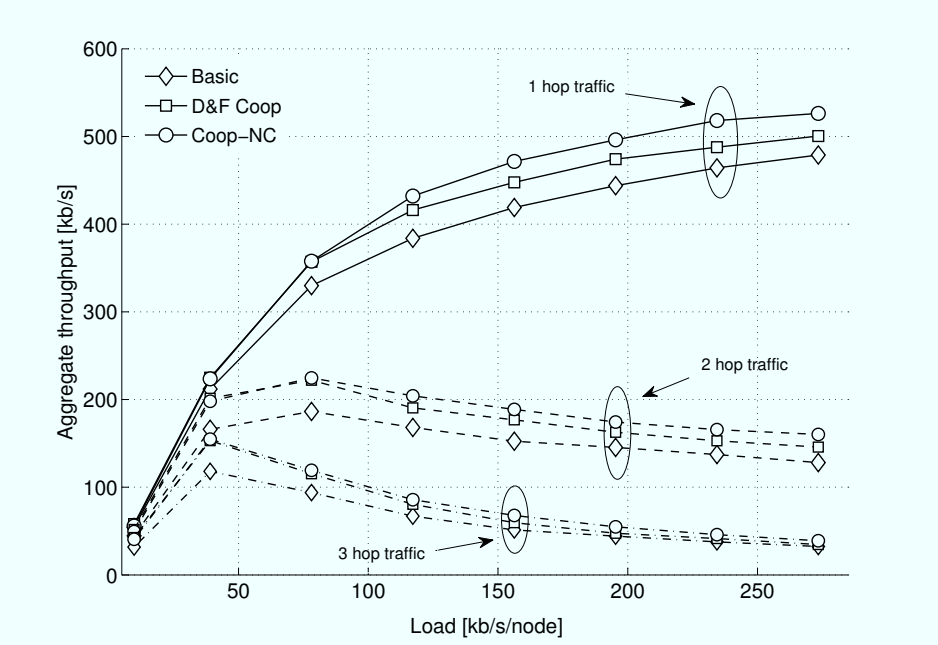


Figure 3.13. *Multihop networks: aggregate throughput vs nominal load [kb/s/node]. Each set of curves shows the behavior for a specific route length in hops.*

several hops.

Another metric of interest is the aggregate throughput, reported in Fig. 3.13 according to the distance in hops. First of all, we notice that the enhancements offered by Phoenix for single hop flows, almost 10% over CCSMA, confirm the trends discussed in Section 3.5.2.1. On the other hand, the gains of our protocol are magnified when multihop paths are considered: for two hop routes Phoenix beats CCSMA by more than 10% and plain CSMA by more than 25%, while in three hop paths the improvements are up to 18% and almost 30% respectively. These boosts stem once again from the higher fairness that Phoenix provides to frames that travel longer distances. If we compare the aggregate throughput due to multihop traffic as a whole,[19] our scheme outperforms its competitors by 40% (CCSMA) and 60% (CSMA).

In conclusion, we consider the average end-to-end delay, depicted against the nominal load in Fig. 3.14. The plot clearly shows two main trends. On the one hand, cooperative relaying helps in containing the latency with respect to plain CSMA thanks to the shorter failure recovery procedures. On the other hand, coded retransmissions make it possible to both reduce the time packets have to spend in the buffer and to avoid delays due to medium

[19] In the simulated topologies, routes up to six hops were possible. We do not depict paths longer that three hops in our figures for the sake of clarity.

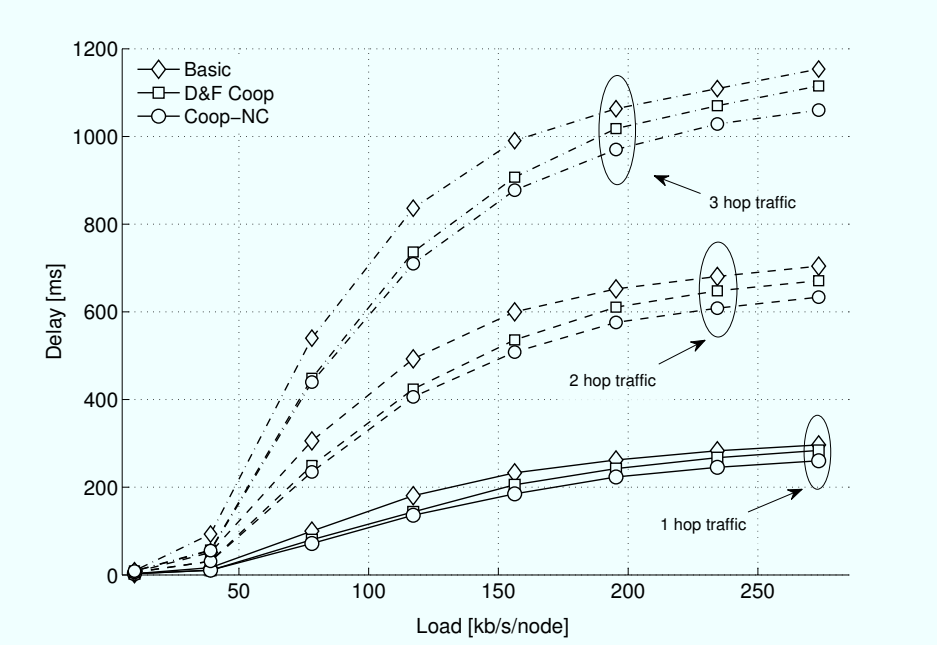


Figure 3.14. *Multihop networks: average end to end delay vs nominal load [kb/s/node]. Each set of curves shows the behavior for a specific route length in hops.*

access contention. The impact of these beneficial effects is proportional to the number of hops a frame has to undergo to reach its final destination, and the performance advantages offered by Phoenix become evident especially for three hop routes, where our protocol reduces the average latency by as much as 40 ms (16%) with respect to CSMA and by 25 ms (10%) compared to plain decode-and- forward cooperation. We infer that not only does Phoenix assure a larger number of multihop communications (see Fig. 3.13) with higher reliability, but also it is able to deliver the payload much more quickly.

3.5.2.3 Tree Networks

Another scenario in which we have tested the protocols is given by multihop environments where 25 nodes are spread in a $200 \times 200\, m^2$ area and arranged in a tree topology. All terminals send their load towards a sink, and the network can be as deep as six hops. This setting is particularly insightful, since it applies to several applications of interest, such as Wireless Sensor Networks (see Appendix A), and challenging, because of the presence of bottlenecks, especially around the sink. In this case, any packet loss delays all the flows that converge into the bottleneck, and MIMO_NC can be very helpful in "masking" these failures: if the packet is recovered together with the transmission of a new frame, the delay induced by the retransmission will play out less severely on the waiting flows. We show re-

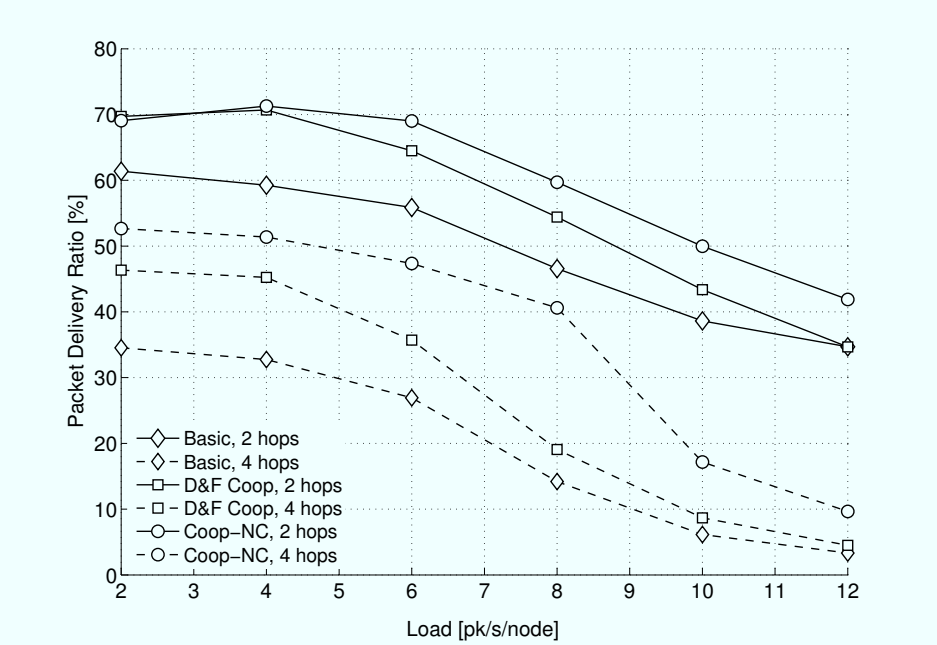


Figure 3.15. *Tree networks: packet delivery ratio vs nominal load [pk/s/node] according to hop distance from the sink.*

sults up to a load of 12 pk/s/node, for which saturation arises. Incidentally, note that such a low value of injected traffic confirms that severe congestion is present in this scenario, highlighting the need for a fast failure recovery procedure.

Fig. 3.15 shows the PDR for nodes at 2 and 4 hops from the destination. It is clear that Phoenix delivers higher performance in all cases, especially for nodes farther from the sink, whose PDR can be twice as large for Phoenix as for CCSMA or CSMA. This is confirmed by Fig. 3.16, which shows that Phoenix outperforms all other protocols by as much as 30% in terms of throughput.

Fig. 3.17 analyzes the aggregate throughput for all the network. In this case the advantage of Phoenix is about 14% over CCSMA, lower than the 20-30% shown in Fig. 3.16. This is due to the fact that one hop neighbors of the sink always have a high throughput no matter what protocol is used (although the cooperative protocols do attain higher performance), and deliver the majority of the traffic. Therefore, CSMA and, to a lesser extent, CCSMA are rather unfair. Phoenix can offer more throughput to the network periphery than CSMA and CCSMA, and thus its gains over the other protocols are more significant than in the single-hop flow scenario. Phoenix is more fair, and therefore one can expect more throughput also at higher loads because the outer regions of the network deliver more traffic than with CCSMA.

To further investigate this aspect, we have computed the Jain's fairness index [63]. The

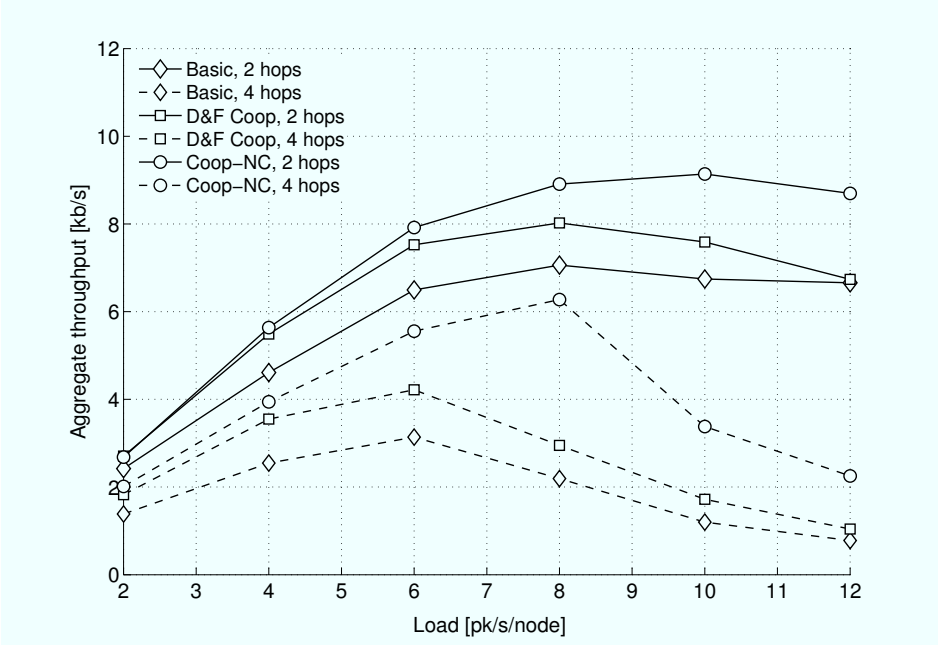


Figure 3.16. *Tree networks: Aggregate throughput vs nominal load [pk/s/node] according to hop distance from the sink.*

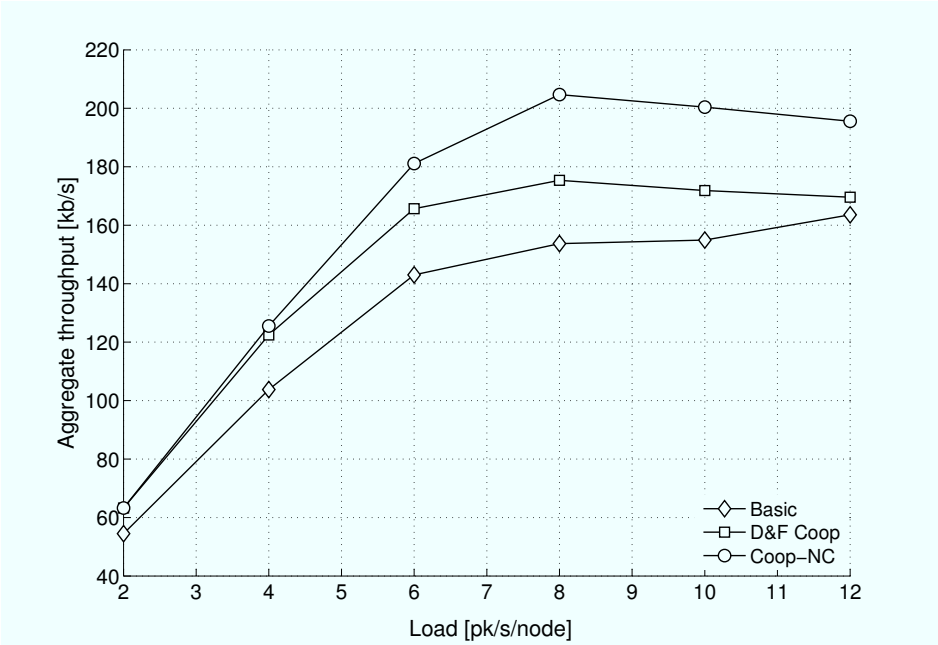


Figure 3.17. *Tree networks: Total aggregate throughput vs nominal load [pk/s/node].*

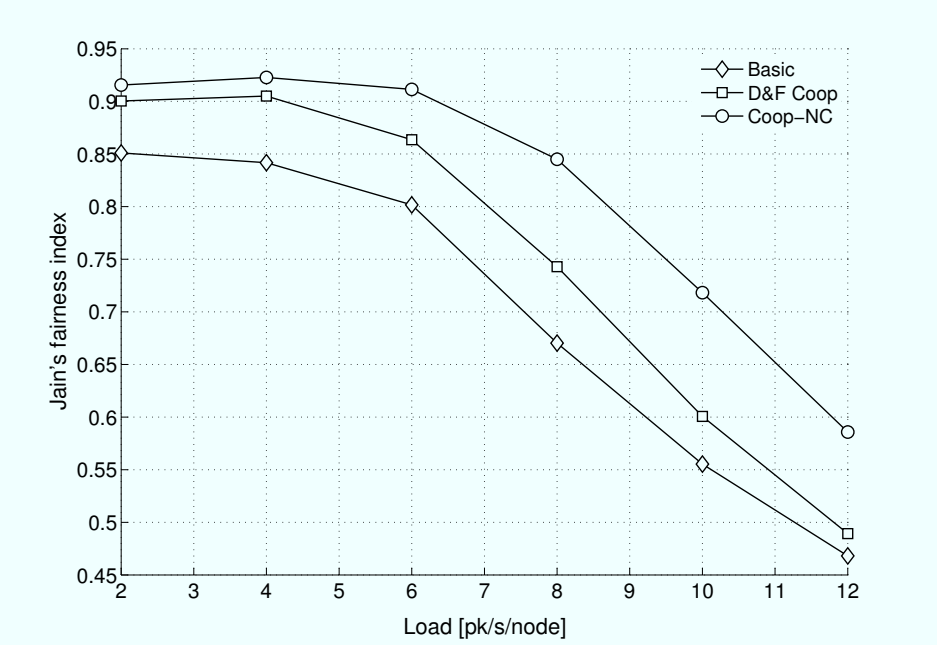


Figure 3.18. *Tree networks: Jain's fairness index vs nominal load [pk/s/node].*

metric, reported in Fig. 3.18 against the nominal load, is equal to 1 when all nodes have the same throughput, whereas the more asymmetric (i.e., unfair) the bandwidth allocation, the lower the index. The plot confirms exactly the intuition discussed above: at higher loads, Phoenix's index is 10-12 percentage points higher than that of CCSMA, let alone CSMA, whose notorious unfairness shows up very early. We point out that our protocol has not been explicitly designed to be fair, and thus this important result is a desirable byproduct that comes for free.

3.5.3 Discussion

Let us now briefly summarize the results discussed in this section. A first implementation of hybrid cooperative-network coded ARQ in carrier sense based ad hoc networks has been presented. The designed protocol, named Phoenix, has been tested in a variety of environments, while further results on its performance in centralized WLANS can be found in [64]. The following lessons and conclusions can be drawn:

- In single hop networks, Section 3.5.2.1, Phoenix yields the best benefits for delay constrained applications, i.e., for low SRL. In addition, the possibility to cooperate and pursue their own interest does not reduce the cooperators' performance;
- In mesh networks (Section 3.5.2.2), gains improve with the route length, because the reduction of queuing times is especially beneficial to multihop traffic;

- In tree networks (Section 3.5.2.3), Phoenix is particularly useful, since such topologies have plenty of bottlenecks, especially close to the sink, and network coded cooperation relieves congestion on them and avoids that packet losses may delay traffic of all upstream nodes. It follows that more fairness is achievable;
- In clustered networks [64], gains over CCSMA are somewhat larger than for random single hop networks (around 12%) and they scale with the network size. Relays are often nodes close to the gateway, and their throughput is not lowered by cooperation. Instead, the number of users with high QoS is significantly expanded with respect to both CSMA and CCSMA. Also coverage is improved because of the larger number of low QoS nodes;
- The protocol is relatively simple compared to CCSMA and also to CSMA. The only remarkable difference with respect to CCSMA lies in the type of cooperation (which is more a PHY problem rather than a MAC issue, and entails only a slight increase in the computational complexity), while both CCSMA and Phoenix must implement a distributed relay election phase. This procedure is heavily based on the IEEE 802.11 backoff mechanism and hence is a rather straightforward software upgrade.

In conclusion, the introduction of network coded cooperation has proven to yield important gains also in more realistic network scenarios than the simple topologies tested in the past and Phoenix has been shown to perform well in a variety of environments.

3.6 Cooperative-Network Coded ARQ in TDMA Based Networks

Section 3.5 has presented a first implementation of network coded retransmissions, focusing on CSMA based networks. While on the one hand remarkable results have been reported, with our solution capable of doubling the gains achieved by decode-and-forward cooperation, the in-depth discussion that we have carried out has provided some first hints towards the fact that completely distributed networking environments that rely on carrier sense may hamper the full potential of advanced collaborative schemes (see, for instance, Fig. 3.10).

In view of this, this section introduces an implementation of hybrid cooperative-NC ARQ in scheduled TDMA systems. By scheduled TDMA we mean a scenario where nodes are grouped into cells,[20] all terminals are linked to a cell master and the master sends out in

[20]Throughout this section the words *cluster* and *cell* are used interchangeably. However, let us remark that cellular networks are not the subject of our study.

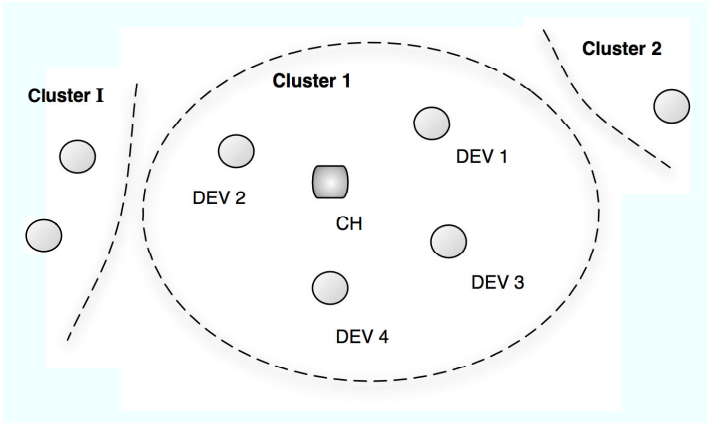


Figure 3.19. *Reference topology for a TDMA scenario.*

every slot the transmission schedule. Such a configuration is of particular interest for Wireless Personal Area Networks (WPANs) [65], where a master commands the communications of its slaves through the exchange of polling packets from the latter and beacons with the schedule from the former.

This further step has two merits. In the first place, the results that will be presented confirm the advantages offered by the proposed idea, as well as its suitability to a variety of scenarios. Secondly, a comparison of the performance of network coded ARQ in carrier sense- and time division-based networks will highlight that not all multiple access schemes are equally suitable to reap the potential of cooperation, laying the basis for the discussions that will characterize Chapters 4 and 5.

3.6.1 Network Scenario

Fig. 3.19 shows the reference scenario for this section. The network comprises a set of clusters $\mathcal{C}_i$, $i = 1, \ldots, I$, each composed by a clusterhead (CH) and $m_i - 1$ devices (DEVs). All the nodes in a cluster are within each other's communication range, so that single-hop peer-to-peer traffic can take place among any pair of terminals, i.e., also the CH generates and receives packets. Moreover, we assume all nodes within a cluster to be frame-synchronized. Communications are based on TDMA, with the CH setting up the transmission scheme for its own cell based on polling requests sent by the DEVs. Such a scenario is representative, for example, of high-rate wireless personal area networks (H-WPANs), as described in the IEEE 802.15.3 standard [66].

We want to highlight a few properties of this scenario that make it particularly suitable for cooperative communications, especially if compared to CSMA. A first advantage is that

the frame schedule is communicated by means of a beacon. Hence, if a node receives the beacon, it is fully aware of which transmissions are planned for the frame. Therefore, if such node fails to decode a PDU addressed to it, it is nevertheless aware of the loss as well as of the details of the packet (e.g., source ID or packet ID), which enables the request of a cooperative retransmission. Instead, as discussed in Section 3.5, if CSMA is employed, a receiver may miss a packet for a variety of reasons (e.g., the node was synchronized on another PDU) and may not be aware that there was a packet intended for it, thus hindering the cooperative process. In addition, a central authority (the CH) can collect all information about the lost packets and coordinate the action of the relays, hence drastically improving the efficiency of cooperation.

3.6.2 Protocols Description

In this section we first describe a basic medium access protocol, called Basic-PAN (B-PAN), that implements a simplified version of the IEEE 802.15.3 standard. Then, we discuss a scheme, named Cooperative-PAN (C-PAN), that modifies such protocol in order to support cooperative retransmissions. Finally, in Section 3.6.2.3, we present and discuss a complete cooperative-network coded medium access strategy.

3.6.2.1 Basic-PAN

With reference to Fig. 3.19, let us focus on the behavior of $\mathcal{C}_i$. When a node joins $\mathcal{C}_i$, a unique local ID is assigned to it, so that in stationary conditions each DEV is identified by an address between 1 and $m_i - 1$, 0 being the ID of the CH. As in 802.15.3-based WPANs, time is divided into successive frames of duration $T_{frame}^{\mathcal{C}_i}$. Each frame, in turn, is made up of three parts, as shown in Fig. 3.20: a beaconing period (BP), during which the CH distributes a schedule describing the communications scheme for the current frame; a channel time allocation period (CTAP) composed by n data slots dedicated to the exchange of peer-to-peer traffic; and a polling period (PP) divided into $m_i - 1$ polling slots and reserved for DEVs to request channel access in the following frame. [21]

At the beginning of the PP of frame $k - 1$, all the nodes in $\mathcal{C}_i$ check the status of their queues. If DEV j has traffic to serve, it exploits polling slot j to send to the CH a packet requesting channel access in the CTAP of frame k. The polling packet (POLL) is made up of up to p transmission requests, each composed by the triplet $< destID, pkID, txAtt >$, where $destID$ is the local ID of the data flow destination, $pkID$ is a unique identifier of the data

[21] The duration of a frame is dependent on the cardinality of a cluster.

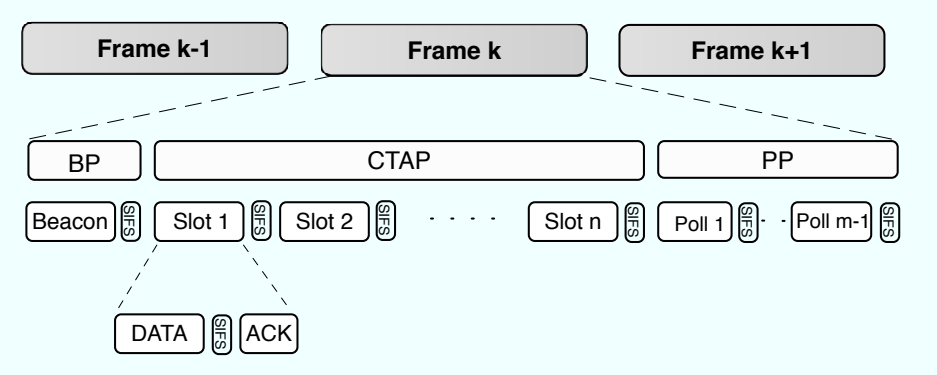


Figure 3.20. *Frame structure for the TDMA-based protocols under analysis*

packet whose transmission is being requested, and $txAtt$ specifies the number of attempts already performed for such packet.

At the end of the PP of frame $k-1$, the CH generates a schedule that assigns the n slots of the following CTAP based on the information received from the various DEVs and on the status of its own queue. In particular, the cluster coordinator randomly allocates data slots to the available data flows, giving priority to packets with higher $txAtt$ values. The computed schedule is then distributed by the CH during the BP of frame k by means of an n-field beacon packet. Each field is in the form $< srcID, destID, pkID >$, where $srcID$ is the local ID of the node that is granted access to the channel, while $destID$ and $pkID$ uniquely identify the packet that is to be served in the corresponding data slot. Upon receiving the beacon, DEVs get to know how to coordinate their medium accesses in frame k. In particular, when a slot assigned to a given transmission starts, the source node sends the corresponding data packet and waits for a feedback, as shown in Fig. 3.20. In case of successful data decoding, the destination responds after a SIFS period with an ACKnowledgement message. If the ACK is received, the communication comes to an end. Otherwise, the source keeps the packet in its queue for later retransmission until the Short Retry Limit value is reached.

3.6.2.2 Cooperative-PAN

As discussed in Section 3.6.2.1, B-PAN relies on a basic ARQ mechanism: in the event of a communication failure, the source is in charge of performing successive retransmissions until either the packet is successfully delivered or the maximum number of attempts is reached. However, the high-rate data transfers that characterize H-WPANs severely hamper the performance of this approach in the presence of correlated channels. From this

viewpoint, cooperative relaying appears extremely appealing, so as to substitute temporal with spatial diversity.

In this perspective, we have designed a medium access policy named C-PAN that suitably extends the functionalities of Basic-PAN. According to this protocol, each node allocates a cooperative queue that can buffer up to n_{coop} PDUs. During a CTAP, a DEV that correctly receives a data packet not addressed to it, continues to listen to the channel for a feedback from the destination. If an ACK for the source is also decoded, the packet is discarded. Otherwise, the DEV assumes that the communication performed during that slot has failed and caches the retrieved data unit in the cooperative buffer.

During polling phases, not only do nodes request channel access for their own traffic, but also they offer their contribution as relays for other links. This is achieved by means of a slightly modified POLL. With C-PAN, the first p_{coop} entries of the packet are in the form $< srcID, destID, pkID, txAtt >$, where $srcID$ and $destID$ describe the source and destination DEVs for an unsuccessful communication overheard and cached by a node in the previous CTAP. The remaining $p - p_{coop}$ entries of the POLL, instead, are structured as in B-PAN, and are dedicated to own generated packets. [22]

At the end of a PP, the CH exploits the received information to compute a schedule for the successive CTAP, in which the first n_c slots can be devoted to cooperation, whereas the remaining $n - n_c$ slots are reserved for basic peer-to-peer communications. Notice that nodes may erroneously propose to cooperate for packets that have successfully been delivered. This is the case when a DEV that has cached a PDU not addressed to it fails in decoding an ACK actually sent by the intended destination. Should such a node act as relay, the corresponding phase would result in a waste of bandwidth. In order to prevent this issue, the cluster coordinator may allocate a slot to a cooperative retransmission only if two conditions are met: i) the original source has requested channel access to perform ARQ for this PDU, i.e., it has not received feedback from the destination; and ii) one or more DEVs have proposed to act as cooperators for such packet. In this case, the CH randomly selects one among the available relays and charges it to perform the hybrid-ARQ phase. We remark that requirement i) is meant to prevent cooperative transmissions that are not needed and that would result in a waste of bandwidth, as in the case of DEVs that erroneously assume a transmission failure due to the missed reception of an ACK.

Once the schedule has been computed, the CH distributes it to the rest of the cluster by means of a beacon. The first n_c entries are in the form $< txID, srcID, destID, pkID >$,

[22] If a DEV has cached $p' < p_{coop}$ packets, it can use the redundant $p_{coop} - p'$ entries of the first part of its POLL to request channel access for its own traffic by setting the $srcID$ field to an invalid value known to the CH.

where $txID$ indicates the ID of the node in charge of performing the retransmission on behalf of the original source identified by $srcID$.[23] The remaining $n - n_c$ entries of the beacon, on the contrary, do not vary with respect to plain B-PAN.

When a node is given a slot to act as relay, it just transmits the corresponding packet without waiting for any feedback. In turn, if the intended destination successfully decodes the data unit, an ACK addressed to the original source is sent.

3.6.2.3 NC-PAN

Let us now discuss how to effectively implement cooperative-network coded ARQ in the scenario under consideration by describing a protocol named NC-PAN that suitably extends the decode-and-forward scheme proposed in Section 3.6.2.2. To this aim, as in the case of carrier sense multiple access, some additional coordination among the nodes in a cluster has to be provided . First of all, DEVs have to determine whether packets that require retransmissions may trigger hybrid cooperative-NC ARQ with sufficiently high success probability. To this aim, a terminal that is not able to decode a data packet addressed to it during a CTAP checks the SINR that affected the failed reception. If this value is below a given threshold Λ_{Th}, the data unit is considered to be eligible only for basic relaying. Otherwise, a hybrid cooperative-NC retransmission could be triggered.

Secondly, with NC-PAN, nodes resort to a slightly modified polling mechanism. The POLL packet is divided in three sections: a *NC-request* part, a *cooperative* part and a *basic traffic* part. The first section is composed by p_{NC} entries in the form $< srcID, destID, pkID, txAtt >$, and can be used by a node to inform the CH of packets for which hybrid cooperative-NC ARQ is likely to be successful, i.e., non-decoded packets addressed to the DEV during the current frame that satisfy the Λ_{Th} constraint. We remark that a node is aware of the parameters of the lost packet because they were included in the previously received beacon. On the other hand, cooperative and basic traffic sections, composed by p_{coop} and $p - p_{coop}$ entries, respectively, are as in C-PAN. The entries of the former are composed by the quadruplet $< srcID, destID, pkID, txAtt >$ and are exploited to notify packets for which a DEV can act as relay, whereas requests for channel access for own traffic are in the form $< destID, pkId, txAtt >$ as usual.[24]

[23] If $n' < n_c$ cooperative retransmissions are to be performed, the remaining $n_c - n'$ slots of the first part of the CPAN are used to serve normal traffic. The CH can signal this by setting the $txID$ field of the corresponding entry to an invalid value, known by all the DEVs.

[24] If a node is not able to fill in the first $p_{NC} + p_{coop}$ entries with cooperative-NC information, it can employ the vacant slots for requesting channel access for its own traffic, following the method discussed in Section 3.6.2.2.

At the end of a PP, the cluster coordinator computes the schedule for the successive frame as follows. First of all, for each packet x in the *NC request* part of the received polls, the CH checks i) if there is a DEV that has proposed to act as relay for x and ii) if such a node has a data unit y addressed to the same destination in the *basic traffic* part of its POLL packet. If both conditions are met, a data slot is allocated to perform a hybrid cooperative-NC retransmission involving x and y. This procedure is iterated until either n_c slots have been assigned or all the NC requests have been processed. In the latter case, the remaining part of the n_c slots dedicated to hybrid ARQ are devoted to basic cooperation and assigned as in C-PAN. Finally, the last $n - n_c$ CTAP slots are used for peer-to-peer traffic and are randomly assigned as usual to the DEVs of the cluster giving priority to transmissions with higher $txAtt$ value.

Once the schedule has been computed, the CH distributes it during the BP by means of a beacon packet whose first n_c entries are in the form $< txID, srcID, destID, pkID, relPkID >$. If node $txID$ has been given a slot to perform a cooperative-NC transmission acting as relay on behalf of DEV $srcID$ for packet $pkID$, the field $relPkID$ uniquely specifies the data unit addressed to the destination that the cooperator should take from its own queue to be combined. Otherwise, if basic relaying is to be performed, the field $relPkID$ is simply set to an invalid value known to all the terminals.

Finally, the acknowledgment procedure is not modified in NC-PAN with respect to the other protocols. If a cooperative-NC retransmission takes place, the destination, due to the working principle of MIMO_NC, may either decode both packets involved or neither of them [48]. In the former case, an ACK to the original source of the retransmitted packet is sent. The DEV that actually performed the retransmission can therefore easily infer that also its own data unit has been successfully retrieved. If, on the contrary, the decoding fails, no feedback is sent and both source and relay take proper actions to further retransmit their own packets as explained for the other protocols.

As a concluding remark, note that the incremental complexity of NC-PAN over C-PAN is quite limited, since the former adds the low-complexity MIMO_NC physical layer, slightly modifies the frame structure and changes the scheduler, the last two items being low-cost software updates. In spite of these limited modifications, Section 3.6.3 will show that the performance improvements are more than worth the effort.

Table 3.3. *Parameters used in our simulations of hybrid cooperative-NC ARQ in TDMA-based networks*

Transmission power	10 dBm
Noise Floor	-91.6 dBm
Detection threshold	-89 dBm
Path loss exponent, α	3.5
Maximum Doppler shift	40 Hz (5 m/s)
Carrier Frequency	2.4 GHz
Data Rate	22 Mbit/s
Short Retry Limit (SRL)	3
Minimum SINR for a coop-NC phase, Λ_{Th}	3 dB
DATA, ACK	2500, 224 bit
Polling pkt *Basic/D&F Coop/Coop-NC*	200/224/272 bit
Beacon *Basic/D&F Coop/Coop-NC*	192/208/220 bit
Number of data slots per frame, n	8
Number of hybrid ARQ slots per frame, n_c	3
Tx req. p per polling pkt *Basic/D&F Coop/Coop-NC*	8/10/10
NC req. per polling packet, p_{NC}	2
Coop. proposals per polling packet, p_{coop}	3
SIFS duration	10 μs
T_{frame} *Basic/D&F Coop/Coop-NC*	1223/1231/1248 μs

3.6.3 Simulation Results

The protocols described in Section 3.6.2 have been tested by means of extensive Omnet++ simulations in clustered networks composed by 6 adjacent non-overlapping cells. Each cell comprises one CH and 5 additional terminals randomly distributed within the circular region, for a total of $m = 6$ terminals per cluster. All the nodes in the network are assumed to be frame synchronized. Every terminal, including CHs, generates single hop traffic addressed to the other members of the cell it belongs to according to a Poisson model with intensity λ (pk/s/node). In our simulations, all the transmitters share the same spectrum (i.e., universal frequency reuse). Moreover, the transmission power level may induce significant inter-cell interference. Therefore, the considered scenario is apt to test the protocols under harsh conditions in terms of both medium access contention and link quality.

The Short Retry Limit (SRL) has been set to 3 for all the protocols. This value has been

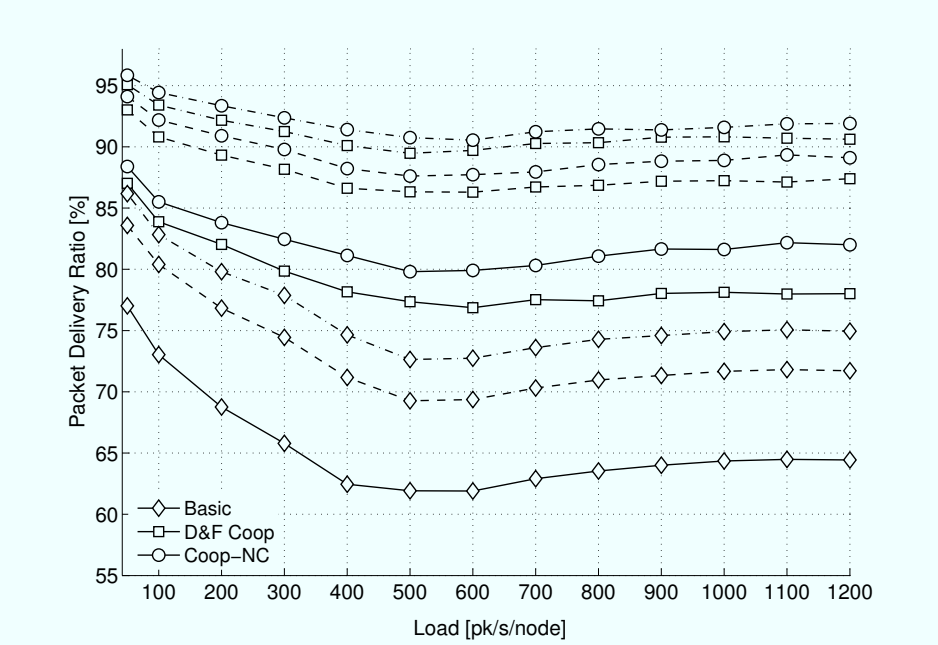


Figure 3.21. *Packet delivery ratio vs nominal load [pk/s/node]. Solid, dashed and dotted lines indicate PDR after one, two and three attempts, respectively.*

chosen so as to achieve a reliability higher than 90% for the schemes that rely on hybrid-ARQ. The wireless environment is subject to path loss with exponent α and to correlated Rayleigh fading with Doppler frequency of 40 Hz, corresponding to a speed of 5 m/s at 2.4 GHz. The complete set of parameters used in our simulations is reported in Tab. 3.3. Incidentally, we remark that our studies take into account the additional overhead required to set up cooperative and network coded phases by considering different packet sizes for the analyzed schemes, and we point out that all the metrics have been obtained by averaging over 25 different topologies, with simulation duration chosen long enough to stabilize the results. Furthermore, consistently with the discussion performed in the rest of this chapter, B-PAN, C-PAN and NC-PAN are referred to in the reported figures as Basic, D&F Coop and Coop-NC, respectively.

Before delving into the outcome of our simulations, we shall observe that the main trends that will be presented follow the arguments carried out in Section 3.5.2, an so do the inferred conclusions. This comes as no surprise, as the schemes that we are considering implement the same communication paradigms faced in Section 3.5. Nonetheless, several remarks and hints that are specific to the scenario under considerations will be pointed out in the remainder of this chapter. Moreover, the present discussion also provides a quantitative insight on the performance of hybrid cooperative-NC ARQ in scheduled systems, prompting signifi-

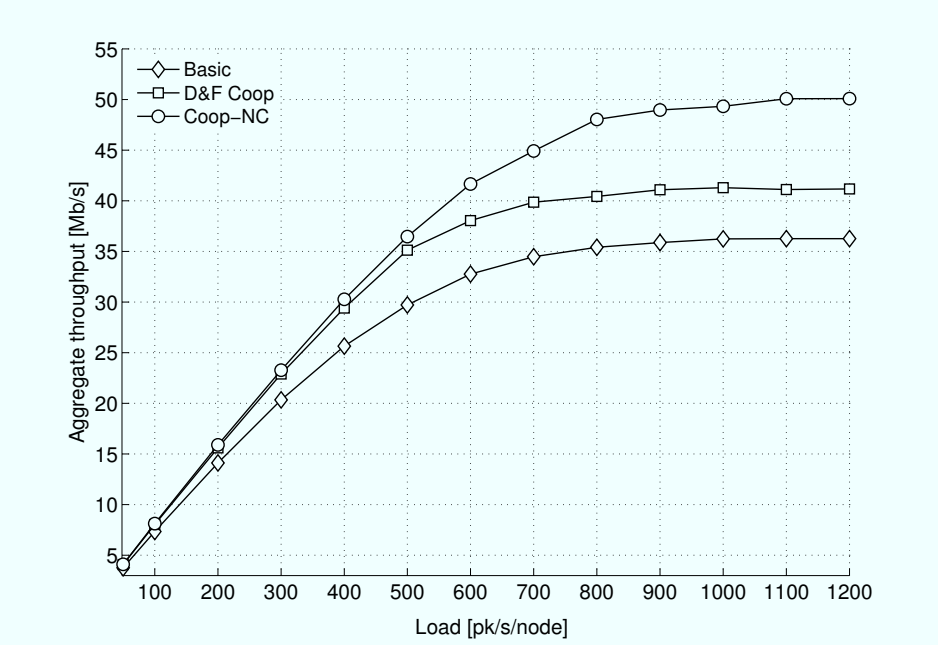


Figure 3.22. *Aggregate throughput vs nominal load [pk/s/node].*

cant differences in terms of achievable gains with respect to what has been seen in CSMA environments.

The first metric of interest is the Packet Delivery Ratio (PDR) as a function of the number of attempts, depicted against the nominal load per node λ in Fig. 3.21. Several facts can be inferred from it. First of all, the diversity gain offered by cooperation yields a significant advantage of 10-15% over B-PAN. In addition, NC-PAN always has a slightly better PDR than C-PAN for all loads and number of attempts. This testifies the soundness of the proposed idea as well as the higher success rate of hybrid cooperative-NC retransmissions. The latter fact is due to the threshold mechanism discussed in Section 3.6.2.3, that triggers such phases only when the success probability justifies the increased overhead. Another important observation is that the advantage in terms of PDR of NC-PAN over C-PAN decreases as the number of retransmissions increases. This is reasonable because when many retransmissions are needed for a packet, the overall interference level at the destination is likely to be very high. Thus, NC phases are unlikely to be triggered, and NC-PAN tends to behave like C-PAN. On the other hand, if the SINR of the corrupted packet is not too low, NC-PAN yields much faster data transfer than the other protocols.

The aggregate throughput is of course another metric of primary importance (Fig. 3.22). Cooperation achieves a gain of at least 13% over B-PAN but NC-PAN outperforms C-PAN and B-PAN by over 20 and 35%, respectively. This significant boost is due to the possibility

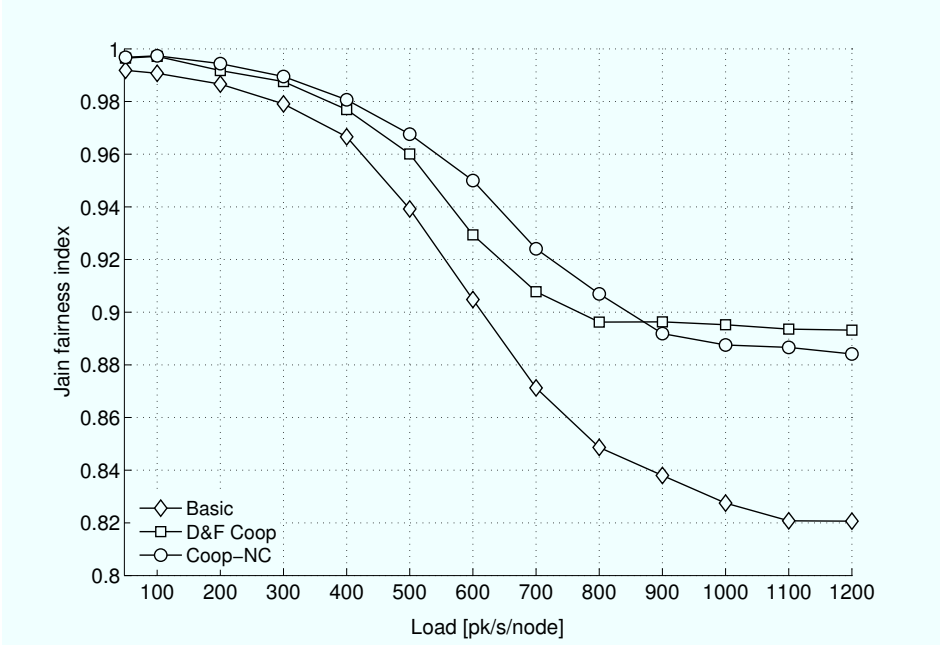


Figure 3.23. *Jain's fairness index vs nominal load [pk/s/node].*

for relays to also deliver their own traffic. A key insight is again the following: in C-PAN, a relay is helping weaker nodes (those who need a retransmission). This implies that terminals with lower SINR, and thus lower per node throughput, have a larger bandwidth share, whereas fewer opportunities for terminals in a good position are available. Instead, in an NC retransmission not only does a relay support lower SINR DEVs, but at the same time it also increases its own throughput. This is especially important at saturation, when bandwidth is scarce.

Jain's fairness index (see Section 3.5.2.3) sheds light on how the throughput is distributed among the terminals (Fig. 3.23). Due to the additional diversity, frames from terminals with poor channel conditions can be delivered more easily, and thus both C-PAN and NC-PAN are more fair than the basic protocol at all loads. Moreover, far from saturation, NC-PAN is more fair than C-PAN. The reason is that, by exploiting NC retransmissions, relays can empty their queues faster. Thus, on the one hand they reduce the level of interference they would generate, and on the other hand they release more bandwidth to other nodes. This clearly helps especially cell-border users to achieve more bandwidth, higher SINR, and hence more throughput. Instead, at saturation the cell-border terminals always have to face strong interference and thus their SINR is anyway low. In C-PAN, nodes in a good position give part of their bandwidth to these weaker terminals, hence achieving a high degree of fairness. In NC-PAN, the relays can nonetheless deliver their traffic also in a cooperative

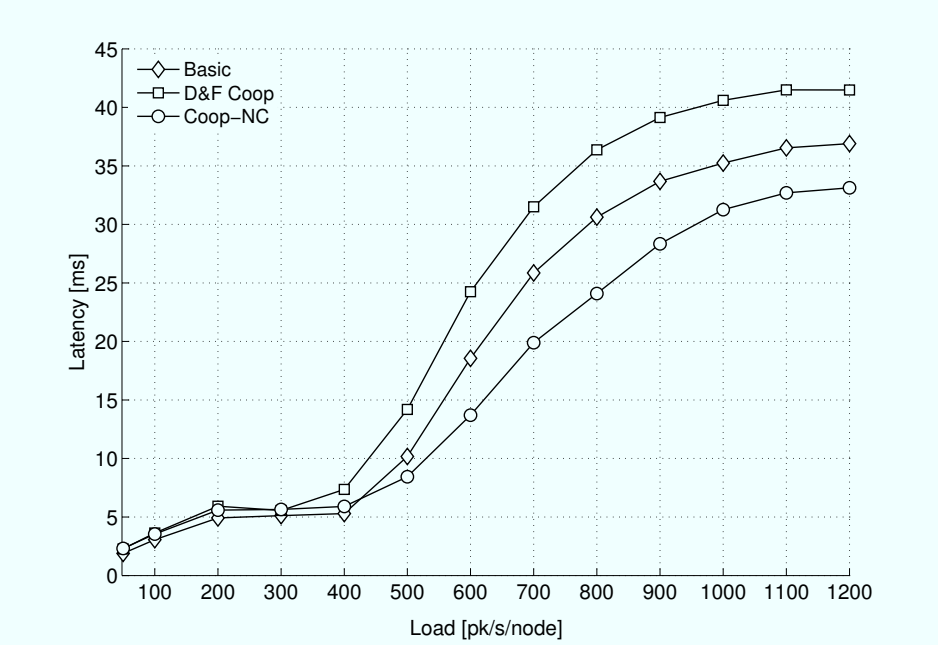


Figure 3.24. *Average latency vs nominal load [pk/s/node].*

retransmission and thus the throughput allocation is somewhat more unbalanced than in C-PAN. The upshot is a lower fairness at saturation. However, this loss is very small (less than 2%), especially if compared to the significant advantages in terms of the other metrics. Moreover, we highlight that this fairness loss does not imply a throughput reduction for the cell-border nodes but instead a larger number of delivered frames for the terminals close to the CH.

The last set of results shows the average delay per successfully delivered packet. The possibility to deliver new traffic in a retransmission avoids the delay that a cooperative phase would induce and this property has a clear impact, as shown in Fig. 3.24. The delay reduction is especially useful for high-throughput nodes, which are often the cooperators. Their packets can be quickly sent to the destination as soon as an NC retransmission opportunity arises, with a delay reduction of 11 and 20% over B-PAN and C-PAN, respectively. Also note that C-PAN performs worse than B-PAN. The reason is that in C-PAN some slots are reserved for retransmissions, hence for packets whose delay is inherently high. This does not happen in B-PAN and thus its delay is on average shorter. It is thus noticeable how not only does the proposed protocol boost throughput with respect to both its competitors, but it also overcomes drawbacks that tend to be induced by basic cooperative strategies.

Incidentally, we remark that the capability of NC-PAN to exploit cooperative transmissions to serve additional traffic leads to gains similar to the ones outlined in Fig. 3.22 also

in terms of energy consumption, not reported here due to space constraints. This result is of particular interest for networks with battery-constrained devices, where the proposed scheme enables more, faster and less energy-hungry transmissions.

The discussed in-depth study of the reference scenario enables to understand the fundamental behavior of the proposed approach. Such an investigation is complemented by an analysis against the number of nodes in each cell at saturation, so as to evaluate the performance of the systems for different network configurations. Tab. 3.4 summarizes the different results for networks composed by 6 cells with 4, 6, 8, 10 and 12 nodes per cell. Several observations are in order. First of all, as the number of terminals increases, the performance worsens for all protocols. The reason is that more and more nodes on the cell-edge are present and this negatively affects all metrics. However, NC-PAN is able to better sustain such degradation, because it enables nodes closer to the CH to deliver their data. This happens less often with C-PAN because the cell-edge terminals require many retransmissions and these are carried out by the relays, who lose opportunities to transmit their own data. Such phenomenon is reflected in the relative gain of NC-PAN over C-PAN in terms of aggregate throughput, which increases from 16% to 28% as the population is expanded from 4 to 12 terminals. It is clear that NC-PAN suffers less heavily from congestion because of its ability to serve nodes in a favorable position. The same reasoning works for all metrics, for instance the delay reduction brought by NC-PAN increases from 10% to 16%.

3.7 Conclusions

In this chapter we have presented a novel form of collaborative behavior capable of improving the performance of wireless networks. The proposed idea extends the paradigm of decode-and-forward cooperation, according to which a node may help its neighbors by acting as relay, performing retransmissions of data units on behalf of the original source so as to obtain spatial diversity at the intended receiver. While such an approach improves decoding probability with respect to plain ARQ solutions, it also asks cooperating terminals to behave selflessly, spending their own resources to help other nodes without any immediate reward and may thus be not too appealing from the perspective of a single node, especially in large networks. We have overcome this drawback by introducing the concept of hybrid cooperative-network coded ARQ. With our solution, relays are allowed to transmit a linear combination of a replica of the original data unit sent by the source and of a packet of taken from their own queue, instead of just a copy of the former, so as to both help surrounding terminals and pursue their own interest at the same time. Such a result is obtained by

Table 3.4. *Performance of the considered protocols as a function of the cluster size*

Nodes/cluster (saturation load)	Protocol	*Throughput [Mb/s]*	*Delay [ms/pk]*	*Jain index*
4 nodes	*Basic*	41.8	22.2	0.77
(1600 pk/s)	*D&F Coop*	45.8	19.2	0.82
	Coop-NC	53.0	16.8	0.83
6 nodes	*Basic*	36.3	41.5	0.82
(1200 pk/s)	*D&F Coop*	41.7	37.5	0.90
	Coop-NC	49.1	34.1	0.88
8 nodes	*Basic*	34.4	60.3	0.84
(1000 pk/s)	*D&F Coop*	38.9	54.2	0.91
	Coop-NC	47.2	49.6	0.90
10 nodes	*Basic*	32.8	80.0	0.85
(1000 pk/s)	*D&F Coop*	37.6	72.3	0.92
	Coop-NC	45.7	66.0	0.91
12 nodes	*Basic*	31.2	91.1	0.86
(800 pk/s)	*D&F Coop*	35.6	101	0.93
	Coop-NC	45.7	67.0	0.92

smartly merging the techniques of cooperative relaying and network coding, and allows to maintain the diversity order provided by the former while serving additional traffic. By means of an analytical model, we have shown that network coded retransmissions have the potential of enabling significant gains over their decode-and-forward counterparts. These trends have been confirmed by extensive simulation campaigns, that have tested link layers capable of supporting the proposed form of cooperation in both carrier sense- and time division-based scenarios, considering distributed and centralized networks as well as single and multi-hop flows.

The thorough study that we have carried out not only has pointed out the soundness of the proposed solution, able to outperform its competitors in a wide range of scenarios, but has also highlighted how the same paradigm may lead to completely different performance gains when implemented according to different medium access policies. Indeed, while our scheme always doubles the gain over plain ARQ with respect to decode-and-forward cooperation, the improvements reach 40% in centralized TDMA-based scenarios, whereas they are bounded to less than 20% in fully distributed CSMA-based networks. Such a remark

hints that a close relationship between cooperative strategies and the multiple access control rules they have to obey exists, and paves the way for the investigations that will be carried out in the following chapters.

Chapter 4

On the Impact of Medium Access Policies on Cooperative Strategies

The discussion carried out in the previous chapter has highlighted how the same collaborative paradigm can lead to significantly different performances when implemented in networks that rely on distinct medium access policies. Such a remark sheds light on the existence of a strong relationship between link layers and cooperation. This aspect, however, has typically been neglected in the research community, as most works deal with rather simplified topologies like the three-node source-relay-destination network, and consider idealized MAC schemes that perfectly coordinate the transmissions among the terminals. Such simplistic assumptions have the merit of enabling advanced analysis techniques, and have led to a thorough understanding of the potential of relaying strategies from a theoretical perspective. Nonetheless, these approaches completely overlook the complex interactions that arise in larger and more realistic scenarios, and may thus provide incomplete or even misleading hints towards the actual development of cooperative systems.

It comes as no surprise, then, that the choice of an underlying multiple access control policy plays a role of paramount importance in view of designing cooperative schemes, as the advantages offered by smart ideas may be hampered and masked by issues that go beyond their intrinsic nature. It is also interesting to point out that these observations apply not only to plain decode-and-forward solutions, but even more to advanced forms of relaying, e.g., hybrid cooperative-NC ARQ. In this perspective, indeed, the more sophisticated and potentially beneficial a cooperative technique, the more its gain can be limited by problems such as relays unavailability or scarce coordination.

Starting from these remarks, in this part of the thesis we carry out an extensive investiga-

tion on the interactions between two different link layers and both plain cooperative hybrid ARQ and its evolved version based on network coding introduced in Chapter 3. In particular, we concentrate on the impact of two actual and widespread MAC schemes, namely IEEE 802.11g [5] and IEEE 802.15.3 [67], epitomizing distributed and centralized wireless systems respectively.

The contribution of the study presented here is twofold. In the first place, it allows to draw some general conclusions on which classes of medium access control are more suitable to effectively support cooperation. From this point of view, it will be highlighted that even though collaborative strategies have been acknowledged to enable significant performance gains, not all link layer policies are able to exploit them at the utmost. Secondly, the thorough discussion carried out on the issues that hinder cooperative hybrid ARQ in different scenarios provides some hints on the design of efficient cooperative MACs.

We start our research in Section 4.1 by identifying the conditions which have to be satisfied for cooperation to successfully take place, regardless of the specific medium access policy being employed, and by investigating the capability of completely distributed, i.e., carrier sense-based, and centralized, e.g., time division-based systems of effectively meeting such requirements. Then, with Section 4.2 we briefly summarize some protocol implementations for decode-and-forward and cooperative-NC ARQ in the two environments. Finally, by means of extensive simulations, we quantify the impact of the considered MACs on the performance of such schemes (Section 4.3).

The main conclusion, reported in Section 4.4, is that the distributed nature of CSMA severely stymies relaying, as it both often prevents nodes from caching packets for later combining and hinders medium access for cooperators. Conversely, even a loosely centralized system like TDMA can provide the necessary coordination to avoid these issues altogether, reaping far higher gains from cooperation and all the more so for its more evolved network coded versions.

4.1 Cooperation in CSMA and TDMA Environments

As discussed in the introduction of the chapter, this part of the thesis focuses on the impact that different medium access control policies can have on relaying strategies based on the decode-and-forward paradigm. Before delving into the details of our investigation, let us briefly recall the key principles of the schemes we will study by considering the topology depicted in Fig. 4.1. In the event of a communication failure between a source $\boldsymbol{S}$ and a destination $\boldsymbol{D}$, basic retransmission policies require the former to perform another try after

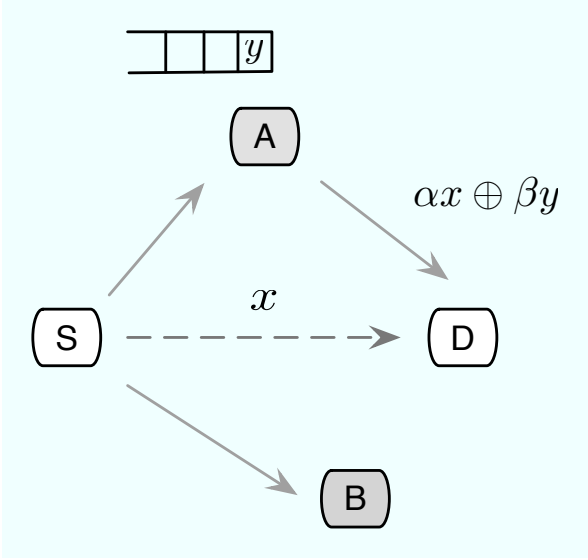


Figure 4.1. *Reference topology for a hybrid ARQ phase for a communication between $\boldsymbol{S}$ and $\boldsymbol{D}$. x is the original packet sent by $\boldsymbol{S}$, while y is a packet taken from $\boldsymbol{A}$'s queue. $\alpha x \oplus \beta y$ represents a linear combination of the two payloads, according to the rules of MIMO_NC ([48] and Section 3.3).*

a certain time interval, whose duration depends on the underlying medium access control. On the contrary, if another node, e.g., $\boldsymbol{A}$ or $\boldsymbol{B}$, has decoded the packet sent by the source, cooperative relaying approaches propose it to immediately perform a retransmission on behalf of $\boldsymbol{S}$. The destination, then, has at its disposal two copies of the same packet received over statistically independent channels, and can take advantage of spatial diversity, e.g., by performing Chase combining.[1]

While this approach has the potential of both shortening recovery phases and improving their reliability, relays are asked to behave in a selfless way, spending their own resources to help other terminals in the hope that similar support will be offered to them when needed. This intrinsic limitation can be overcome by relying on the *hybrid cooperative-NC ARQ* proposed in Chapter 3, according to which if a potential cooperator, say $\boldsymbol{A}$, has decoded x and has in its own queue a data unit y addressed to the same destination, it can exploit a cooperative phase to transmit a linear combination of x and y instead of just a copy of the former. In such a case, resorting to the MIMO_NC physical layer (see Section 3.3), $\boldsymbol{D}$ may be able to retrieve the information content of both packets. This advanced form of cooperation makes it then possible to both benefit from spatial diversity and to opportunistically serve additional traffic.

[1]For a more in-depth discussion on cooperative relaying as well as on hybrid cooperative-NC ARQ, please refer to Section 3.2.

In order to understand the interactions between link layer policies and relaying approaches, we start by identifying the conditions which have to be met for cooperation to successfully take place. From this viewpoint, and with reference to the topology of Fig. 4.1, three key requirements can be identified:

R1) $\boldsymbol{D}$ *has to cache the corrupted version of* x *sent by* $\boldsymbol{S}$*.*

In order to take advantage of a relayed transmission by means of Chase combining or MIMO_NC, the destination has to store the payload coming from the source. Not only does this mean that $\boldsymbol{D}$ has to be available for reception, i.e., neither transmitting nor synchronized to other ongoing communications, but also that it has to be aware that an attempt addressed to it for a specific data unit is being performed. The latter condition can be met either by decoding the header of x, which contains information on nodes involved in the transmission as well as the ID of the packet,[2] or by having been informed in advance of the incoming transmission, as in the case of a scheduled system.

R2) One ore more relay candidates have to be available.

The possibility for a terminal to act as cooperator is subject to specific constraints. In the first place, a potential relay must not be involved in other ongoing communications. Secondly, according to the decode-and-forward paradigm considered throughout this chapter, the node must have successfully decoded the packet originally sent by the source. Finally, a terminal must be granted access to the medium to perform a hybrid ARQ transmission. Such a requirement is not trivial, and may have a severe impact on the availability of relays. Indeed, the communication between $\boldsymbol{S}$ and $\boldsymbol{D}$ does not reserve in general bandwidth for a cooperative stage, and relayed transmissions have to be carefully handled, so as not to harm other ongoing links in the network.

R3) A policy to trigger and define the hybrid ARQ phase has to be specified.

When several nodes are involved in the cooperative processes, additional coordination at the link layer is required. In particular, hybrid ARQ should be triggered by the destination only when it is useful, i.e., $\boldsymbol{D}$ has not decoded the packet sent by $\boldsymbol{S}$, and if it can be successfully supported, i.e., $\boldsymbol{D}$ has cached a corrupted version of the data unit. Transmissions performed under different conditions would only result in undesired interference and wasted bandwidth. Moreover, a strategy to identify who

[2] We point out that the header may be successfully received even if the payload is corrupted, since it can be independently channel encoded and have a separate CRC.

among multiple relay candidates may act as cooperator has to be provided, so as to avoid collisions and inefficiencies.

The capability of a network to cope with these general requirements is tightly related to the control policy used for medium access, which thus are pivotal for the effectiveness of cooperative solutions. Starting from this observation, we investigate the capability of two radically different networking environments, namely *completely distributed*, e.g., carrier sense-based, and *centralized*, e.g., time division-based, systems of effectively supporting relaying schemes. To this aim, we focus on two reference medium access controls: IEEE 802.11g [5] and IEEE 802.15.3 [67]. These protocols are some of the most widely used implementations of the carrier sense- and centralized scheduling-based multiple access policies, respectively. Therefore, their analysis has first of all practical relevance. Also, the differences between the physical layers employed by the two standards, e.g., the use of Orthogonal Frequency Division Multiplexing (present in 802.11g and lacking in 802.15.3), may increase the data rates, but do not fundamentally alter the network capacity, for instance in terms of simultaneous packet receptions. This ensures a comparison that is not biased by non link layer-related aspects. Even more relevant, they employ two radically dissimilar philosophies, and hence enable the investigation of a wide range of issues for cooperative networking. Finally, since 802.11g and 802.15.3 are highly representative of their own class of wireless MACs, such study allows to obtain insight into the essential relationships and problems between CSMA/TDMA and HARQ.

In the remainder of this section we focus separately on such medium access strategies, analyzing their impact on hybrid ARQ implementations.

4.1.1 Carrier Sense Multiple Access

According to the Carrier Sense Multiple Access paradigm,[3] a node that has data to send chooses a random backoff interval and senses the power level on the wireless channel. If the perceived value exceeds a given threshold, the terminal infers the presence of other ongoing links and freezes its backoff until the medium is sensed idle again. Conversely, when the contention window expires with no other communication being detected, the terminal transmits its packet. If the destination successfully decodes the payload, an acknowledgment (ACK) is sent in response, and the data exchange comes to an end. Otherwise, the source starts another backoff interval (following the Binary Exponential Backoff mechanism [5])

[3]For a more detailed description of the 802.11 Distributed Coordination Function, please refer to Section 3.5.1.1.

and iterates the procedure until either a positive feedback is received or a maximum number of attempts (Short Retry Limit, SRL) is reached.

While simplicity and lack of a coordinating unit make this link layer particularly suitable for ad hoc networks, its completely distributed nature poses some intrinsic challenges that hamper effective implementations of cooperative solutions.

In the first place, a terminal becomes aware of a communication it should be involved in only when the transmission is actually performed. This implies that, in case of a data loss, cooperation can solely be enabled when the addressee of the payload successfully receives the header of the incoming packet (*requirement R1*). Such a condition is not easily met in carrier sense-based environments. Indeed, while payload and header may be separately encoded, they experience very similar channels, e.g., in terms of fading and interference. Therefore, the decoding probabilities for the two of them are correlated. In addition, the destination may lose data since it is already synchronized to another ongoing transmission not overheard by the source. Also in this case the incoming packet cannot be cached for later combining, as it is regarded as interference at the receiver. This issue, often referred to as the hidden terminal problem, stems from the asynchronous and decentralized behavior of 802.11, according to which each terminal in the network is asked to decode (and thus can be locked by) any communication taking place in the surroundings, even if not addressed to it.

Carrier sense-based access control also influences the availability of relays (*requirement R2*), for two primary reasons. First of all, the hidden terminal problem may affect potential cooperators as well, as nodes other than the intended destination may fail to decode a packet from $\boldsymbol{S}$, e.g., if they are synchronized to the reception of other data units. Secondly, terminals that could perform a hybrid ARQ phase may not be allowed to transmit. In fact, although the active direct link protects an area centered at $\boldsymbol{S}$, relay candidates, possibly located at the borders of such a region, are likely to sense a significantly higher aggregate interference and may thus be enforced to refrain from accessing the channel.

Finally, we remark that distributed systems lead to non-trivial policies to identify who has to act as cooperator (*requirement R3*). Static approaches that associate a fixed relay to each source-destination pair may turn out to be highly inefficient, as the chosen terminal may be unavailable or experiencing bad channel conditions towards the destination when its action is needed. Conversely, decentralized algorithms to select a cooperator among a set of candidates (i.e., nodes that have decoded the original packet sent by the source and are allowed to transmit) may lead to collisions or suboptimal decisions, since contenders are reciprocally unaware of each other's condition.

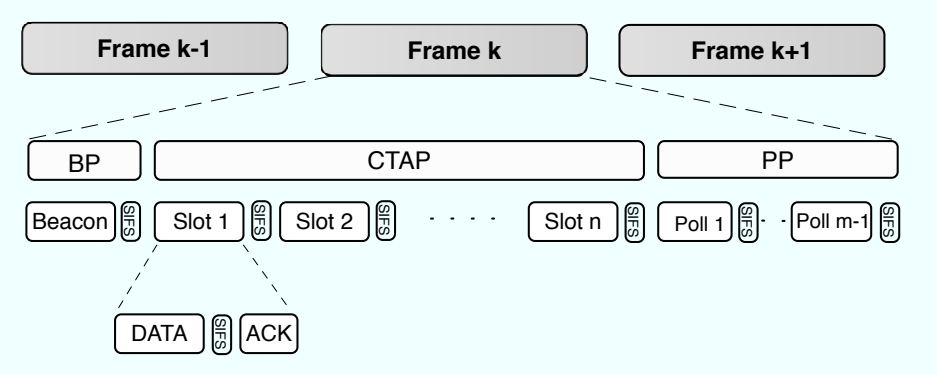


Figure 4.2. *Frame structure for the time division access-based systems under analysis. In this example, the cluster is composed by m nodes, including the clusterhead.*

4.1.2 Time Division Multiple Access

As a reference for time division-based multiple access, we focus once again on high-rate wireless personal area networks (WPANs), as described in the IEEE 802.15.3 standard [66]. These systems, indeed, represent a good scenario for the implementation of hybrid ARQ due to their demand for high throughput even under harsh interference conditions. In WPANs, the network is divided in clusters, each comprising a clusterhead and other wireless devices. All the terminals in a cluster are located within each other's communication range, and single hop peer-to-peer traffic can take place among any pair of nodes. Time is organized in successive frames, each of them subdivided in three periods, as shown in Fig. 4.2: a *beaconing phase* (BP), a *channel time allocation period* (CTAP), and a *polling phase* (PP). During the polling phase of frame k, every node in a cluster is allocated a slot for sending a POLL packet to its clusterhead, requesting channel access for data traffic in frame $k+1$. Once all the polls have been collected and considering the status of its own queue, the cell master runs an algorithm to assign the n data slots of the subsequent CTAP. The computed schedule is then broadcast to the terminals in the cell during the beaconing phase of frame $k+1$. Upon receiving the beacon, all the members of the cluster become aware of how to coordinate medium access for the following data exchanges. In particular, when a slot allocated to a given communication starts, the source node transmits the corresponding data unit and waits for a feedback. If the packet is successfully received, the destination responds with an ACK. Otherwise, no reply is provided, and the source keeps the packet in its queue for later retransmission until the SRL is reached.

The described link layer requires synchronization among terminals within a cluster and introduces additional complexity with respect to the procedures that characterize CSMA.

However, the presence of a coordinating unit significantly favors the implementation of hybrid ARQ solutions. First of all, nodes become aware of all the transmissions that take place in their cell during the beaconing phase, and therefore successful decoding of the packet headers is no longer a requirement for cooperation to take place. From this viewpoint, a destination can cache corrupted versions of data units addressed to it as long as the beacon for the current frame has been received (*requirement R1*). Indeed, receivers can easily avoid being locked to undesired communications, since only one transmission per cluster can be active at a time and packets coming from other cells can be discarded by taking advantage of cluster-based synchronization. Moreover, in contrast to CSMA systems, non-decodable data can be stored even in the presence of extremely unfavorable fading conditions between the source and the destination, as beacon and payload are in general received over different and statistically independent channels. Incidentally, we remark that the absence of the hidden terminal problem can have a beneficial effect on the availability of relay candidates too, since nodes can properly synchronize to, and thus potentially cache, the only transmission taking place in their cell at any given time.

The centralized policy of TDMA also eases the identification and selection process of relays (*requirements R2* and *R3*), as the organization of cooperative phases can be completely entrusted to the clusterhead. Polling processes can easily be extended to include notifications of both lost data units at intended destinations, i.e., hybrid ARQ chances, and successfully decoded packets addressed to other nodes, i.e., availability to act as relay. Upon receiving this information, the cell master can decide whether and how to suitably allot CTAP slots to cooperative communications and notify all the involved terminals exploiting the beaconing phase. Following this approach, inefficient contentions among candidates are avoided, and optimal retransmissions may be possible.

Two concluding observations are in order. In the first place, we notice that the issues discussed in this section for distributed and centralized systems are related to the capability of effectively triggering hybrid ARQ procedures, and thus affect both plain, e.g., decode-and-forward, and advanced, e.g., network coded, cooperative techniques, possibly undermining their beneficial contributions. Secondly, the problems that have been highlighted stem from the interaction of several nodes, and therefore become more pronounced in larger networks or at high traffic loads, which are exactly the conditions in which cooperation could be exploited at its utmost.

4.2 Protocols Description

In order to evaluate the impact of medium access control policies on cooperation, we focus on the implementation of both decode-and-forward and cooperative-Network Coded ARQ in CSMA- and TDMA-based systems. To this aim, we consider the protocols introduced in Chapter 3 with some minor variations. For the ease of the reader, we sketch again in this section a brief description of their key features. A complete presentation of the schemes can be found in Sections 3.5.1 and 3.6.2, respectively.

4.2.1 Carrier Sense Multiple Access

4.2.1.1 Decode-and-forward cooperation

Let us refer to the topology depicted in Fig. 4.1, and assume that the transmission of x between $\boldsymbol{S}$ and $\boldsymbol{D}$ following the procedures of the 802.11g DCF has not succeeded. If the destination has cached a corrupted version of the payload, a Not ACKnowledgement (NACK) packet is broadcast, asking for hybrid ARQ. Nodes that receive this request and that have successfully decoded the data unit coming from the source enter a distributed contention based on carrier sense to select who has to act as relay. To this aim, each candidate picks a random backoff, whose duration, in slots, is uniformly drawn from the set $\{0, 1, \ldots, CW_{rel}\}$, and listens to the medium. If the power level on the channel exceeds a threshold $\Gamma_{CS}^{(rel)}$, the terminal infers that either somebody else is cooperating or other communications are going on in the surroundings and refrains from transmitting. Otherwise, the node wins the contention and, at the end of the backoff, sends a copy of x to $\boldsymbol{D}$. In case of Chase combining success at the destination, an ACK addressed to $\boldsymbol{S}$ is transmitted. Else, no feedback is provided, and the source will perform another attempt according to the Binary Exponential Backoff mechanism.

4.2.1.2 Cooperative-Network Coded ARQ

With reference to Fig. 4.1, let $\boldsymbol{D}$ have cached a corrupted version of x, and assume that $\boldsymbol{A}$ has in its own traffic queue a packet y addressed to the same node. As discussed in Section 4.1, in such a situation the Coop-NC paradigm proposes to exploit a hybrid ARQ phase to convey both messages to the destination. However, if the quality of the stored version of x at $\boldsymbol{D}$ is extremely poor, a network-coded retransmission would have too low a success probability with respect to decode-and-forward cooperation (refer to Chapter 3). In order to cope with this aspect, the NACK packet can be modified to include an *NC_flag*

field. Before asking for hybrid ARQ, the destination checks the average Signal to Noise and Interference Ratio (SINR) that characterized the cached corrupted payload. If this value is below a given threshold Λ_{Th}, the *NC_flag* is set to 0, and only basic cooperation is permitted. Otherwise, the flag is set to 1, signaling that a coded retransmission can be successfully supported. Terminals that receive the request for cooperation and that have decoded the packet sent by the source enter the distributed contention procedure discussed earlier. If the *NC_flag* is not active or the chosen relay does not have traffic for $\boldsymbol{D}$, decode-and-forward is carried out as described in Section 4.2.1.1. On the contrary, when Coop-NC ARQ can be supported both at the destination and at the winner of the contention, the latter, say $\boldsymbol{A}$, generates a linear combination of x and y following the principles of network coding and transmits the obtained packet. In this condition, as a result of the working principles of MIMO_NC [48], $\boldsymbol{D}$ can either retrieve both the encoded data units or neither of them. If decoding fails, no feedback is provided. Instead, when Coop-NC ARQ succeeds, a single ACK is sent. In any case, both the original source and the relay can infer the outcome of the cooperative phase and act consequently, i.e., discarding their packet or keeping it in queue for later retransmission. Incidentally, notice that, in contrast to the implementation proposed in Section 3.5.1.3, we do not consider the possibility of using network coded retransmissions to serve traffic addressed to secondary destinations, i.e., the case in which y is for a node other than $\boldsymbol{D}$. This choice has been made in order to simplify the protocols and takes into account the results of some studies we performed, which showed how the aforementioned feature does not lead to significant improvements.

4.2.2 Time Division Multiple Access

4.2.2.1 Decode-and-forward cooperation

The basic TDMA scheme described in Section 4.1.2 can easily be extended to take advantage of decode-and-forward techniques. In the solution that we propose, each node keeps a so called cooperative queue to cache packets for which it may act as relay. In particular, during CTAPs, terminals store successfully decoded data units that are not addressed to them and for which they do not overhear an acknowledgment from the intended receiver. Moreover, in order to increase the success probability of retransmissions, a node, say $\boldsymbol{A}$, proposes itself as relay only for destinations it shares a good link with (i.e., only if the SINR of the last packet decoded over the $\boldsymbol{D}$-$\boldsymbol{A}$ channel was higher than a threshold Λ_{coop}).

The polling phase is used not only to ask for channel access to serve own traffic, but also to notify availability to perform hybrid ARQ. This can be achieved by reserving p_{coop}

out of the p entries of a POLL packet [4] to describe payloads cached in the cooperative queue during the current frame. Once all the polls have been received, the clusterhead can proceed with the schedule computation, allocating up to n_{coop} of the n slots in the successive CTAP to cooperative communications. Notice that nodes may erroneously propose to act as relays for data units that have been successfully delivered. This may happen, for instance, in case of unfavorable channel conditions between the terminal and the intended addressee of the packet, which may lead to the loss of the positive feedback sent by the latter. Should such a decode-and-forward phase be allocated, it would result in a waste of energy and bandwidth. In order to prevent this issue, the cell master [5] allots cooperative slots only if the original data source requests access to perform basic ARQ for the packet, i.e., it has not received an ACK, and one or more terminals are available as relays.

The designed schedule is then distributed as usual by means of a beacon. When a node is reserved a slot for hybrid ARQ, it simply transmits the corresponding packet, without waiting for any feedback. Conversely, if Chase combining succeeds, the destination sends an ACK to the original source of the data unit.

4.2.2.2 Cooperative-Network Coded ARQ

As discussed in Section 4.2.1.2, network coded retransmissions should be triggered only when the SINR on the corrupted payload available at the destination is sufficiently high. However, differently from the immediate feedback implemented in CSMA, we propose to exploit polling to identify potential Coop-NC phases. To this aim, the POLL packet is split in three sections: an *NC-request* part, a *cooperative* part, and a *basic traffic* part. The first part is composed by p_{NC} fields, describing the non-decoded packets addressed to the node that have been cached during the current frame with average SINR greater than Λ_{Th}. The successive p_{coop} entries are used by the source of the POLL to propose itself as cooperator as in the decode-and-forward case. Finally, the remaining $p - p_{NC} - p_{coop}$ fields are devoted as usual to requests for channel access to serve own traffic. When computing the schedule for a frame, the clusterhead scans through the *NC-request* sections of all the received polls, and tries to allocate hybrid coop-NC phases for each of the notified packets. In particular, network coded channel access is reserved for a data unit if there is at least one relay candidate for it with a packet for the same destination in the *basic traffic* part of its POLL. This procedure is iterated until either n_c slots have been assigned or no other Coop-NC transmissions

[4] For a detailed description of the packet structures please refer to Section 3.6.2.2.

[5] The words *cell* and *cluster* will be used interchangeably.

can be supported. In the latter case, the remaining part of the n_c slots reserved to hybrid ARQ are distributed, if possible, to decode-and-forward cooperation as in Section 4.2.2.1.

Network coded retransmissions are handled as usual, with the relay node sending a linear combination of the involved payloads and the destination either responding with a single ACK, which can be interpreted by both cooperator and original source, or providing no feedback at all.

4.3 Simulation Results and Discussion

In this section, an extensive investigation is carried out by means of simulations so as to quantify the impact of the considered access control strategies on the performance of cooperative techniques. Three schemes are compared, and each of them is instantiated for CSMA and TDMA: on the one hand "Basic", representing the standard solutions discussed in Section 4.1; on the other hand "Coop" and "NC", implementing hybrid ARQ policies as explained in Section 4.2. Moreover, we have designed and studied protocol bounds that idealize some aspect of medium access contention in the two networking environments under consideration. The rationale behind this approach is twofold. Indeed, not only can the performance analysis of the bounds provide an insight on the share of potential gains attained by the implementations that we propose, but also it sheds light on the intrinsic limitations of different link layers to cooperation. As for CSMA, the bound that we consider resembles the one proposed in Section 3.5.2,[6] and always selects the best relay among the available candidates, i.e., the node that experiences the most favorable channel conditions to the destination. Moreover, the selection process is idealized so that all potential candidates are immediately informed of how the hybrid ARQ phase will be organized, i.e., no contention is required and collisions among relays are avoided. In TDMA, conversely, the protocols have been bounded by letting terminals have unlimited cache for packets involved in cooperative processes and include in their polls the complete description of their queues, at no additional cost in terms of overhead. Indeed, with this approach clusterheads can maximize the number of allocated decode-and-forward and network coded retransmissions.

The simulation scenario is a network composed by 6 cells, each comprising 6 nodes, and all results are averaged over 25 different topologies simulated in Omnet++. The system works under universal frequency reuse, i.e., all nodes transmit and receive on the same frequency. Therefore, the presence of several clusters induces a high level of co-channel

[6]However, in contrast to the bound described in Chapter 3, the scheme considered here does not resort to network-coded retransmissions involving secondary destinations.

interference and tests the network in demanding conditions. The nodes generate Poisson traffic for other members of the same cluster, and packets are sent over correlated Rayleigh fading channels. The Short Retry Limits (SRLs) have been chosen so as to achieve comparable packet delivery ratios at saturation (around 98%). This led to 4 attempts for the basic protocols and 3 for the cooperative approaches. All relevant simulation parameters are reported in Table 4.1.

We distinguish two main types of metrics: end-user and cooperative. The former are those quantities that directly affect the end-user satisfaction, i.e., aggregate throughput, delay, transmit energy consumption. The latter are those metrics that are not immediately relevant for the application layer, but shed light on the relationship between the access scheme and cooperation. Examples are the reasons why a cooperative retransmission is not carried out or the percentage of hybrid ARQ phases that employ NC. Since the discussion of these two families of performance criteria is quite different, they are presented in two distinct subsections.

A basic remark common to all graphs is the following. The two basic transmission policies (802.11 and 802.15.3) are very different, and hence a comparison of the absolute values of end-user performance would be glib. Therefore, the plots always show normalized metrics. Moreover, due to the different data rates, the injected traffic is significantly higher in the TDMA than in the CSMA setting . Hence, it is more insightful to refer the metrics to the normalized load λ^*, defined as the absolute load in terms of generated packets per node per second divided by the saturation load, i.e., the per node load at which 98% of the saturation aggregate throughput is achieved.

4.3.1 End-User Metrics

The performance in terms of aggregate throughput, delay and transmit energy consumption is analyzed for the eight considered protocols. As stated before, a comparison of the absolute values is not insightful and hence the plotted curves in this subsection are the relative gains of the different protocols with respect to CSMA/TDMA-Basic, for CSMA and TDMA systems, respectively. Incidentally, also notice that the behavior of the schemes in scheduled systems resembles the one discussed in Section 3.6.3, whereas the analysis of CSMA-based protocols differs from the one provided in Section 3.5.2 due to the different topologies being considered here.

The aggregate throughput is depicted in Fig. 4.3. Several observations are in order. First of all, at very low loads the gain is about 7-9% and is almost independent of the specific pro-

Table 4.1. *Parameters used in our simulations*

	CSMA	TDMA
Transmission power [dBm]	10	10
Noise Floor [dBm]	-91.6	-91.6
CS threshold, Γ_{CS} [dBm]	-89	//
CS threshold for relay contention, $\Gamma_{CS}^{(rel)}$ [dBm]	-89	//
Detection threshold [dBm]	-89	-89
Path loss exponent	3.5	3.5
Maximum Doppler shift [Hz]	40 (5 m/s)	40 (5 m/s)
Saturation load [pk/s/node]	128	960
Slot, DIFS, SIFS duration [μ s]	20, 128, 28	//, //, 20
Carrier Frequency [GHz]	2.4	2.4
Data Rate [Mbit/s]	6	22
Initial maximum contention window (slots)	128	//
Short Retry Limit - *Coop* and *NC*	3	3
Short Retry Limit - *Basic*	4	4
Number of slots used for relay contention, CW_{rel}	32	//
Minimum SINR to trigger a Coop-NC phase, Λ_{Th} [dB]	3	3
Minimum SINR with a destination to act as relay, Λ_{coop} [dB]	//	5
Simulation Time after transient [s]	9	9
DATA header [bit]	272	//
Payload [bit]	4000	4000
ACK/NACK [bit]	112	224
Polling pkt *Basic*/*Coop*/*NC* [bit]	//	200/224/272
Beacon *Basic*/*Coop*/*NC* [bit]	//	192/208/220
Number of data slots per frame, n	//	8
Number of hybrid ARQ slots per frame, n_{coop}	//	3
Tx req. p per polling pkt *Basic*/*Coop*/*NC*	//	8/10/10
NC req. per polling packet, p_{NC}	//	2
Coop. proposals per polling packet, p_{coop}	//	3

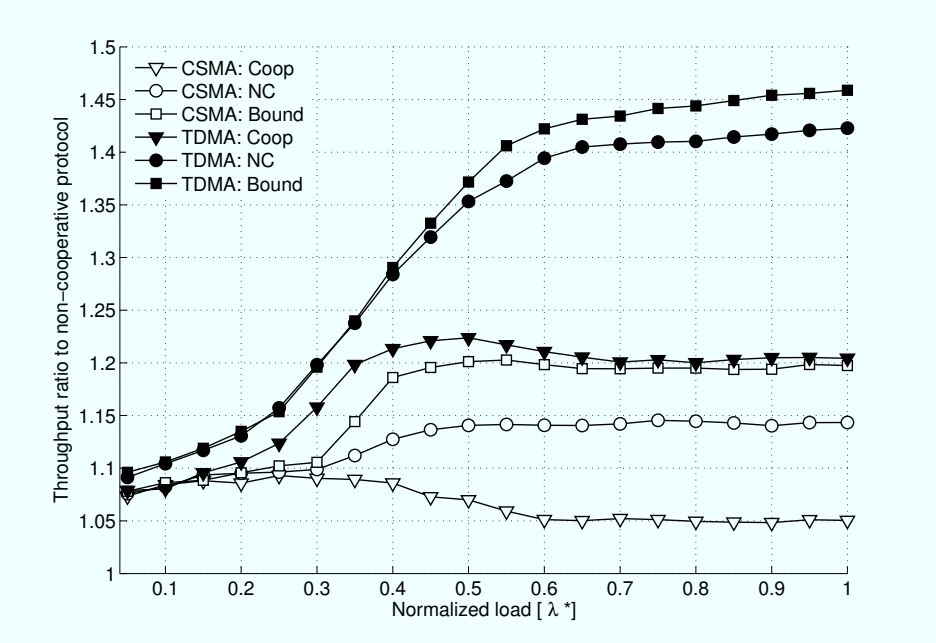


Figure 4.3. *Ratio of aggregate throughput for cooperative schemes over the basic protocol vs. normalized load λ^*.*

tocol and of the medium access policy. The rationale is that for very light traffic the system is just noise limited and no network coding opportunities are possible. Thus the different protocols are simple cooperative systems with no medium access contention and exhibit very similar performance because they work in very comparable environments. As the traffic increases, CSMA-Coop and TDMA-Coop gains first rise and then fall. The reason is that the throughput of the Basic protocols grows more for heavier loads and hence the relative gain of the Coop systems decreases.[7] Such a behavior is due to the more fair resource distribution of the Coop protocols: in Coop, the nodes with good channels (and hence potentially high throughput) use part of their resources to support terminals in worse conditions in their retransmissions. The outcome is a more balanced distribution of the bandwidth, at the expense of a slightly reduced throughput. It is important to notice that this issue does not affect Coop-NC ARQ, as coded retransmissions boost the throughput also of the high performing nodes when they help other terminals.

It is also remarkable that TDMA cooperative systems attain about twice as large an improvement than their CSMA counterparts, which is true for the other aggregate metrics, too. Accordingly, TDMA-NC is able to achieve over 40% gain on TDMA-Basic and 20% on TDMA-Coop, far more than the 15% and 5% improvement of CSMA-NC and CSMA-

[7] A more in-depth discussion on the issue of unfairness can be found in Sections 3.5.2 and 3.6.3.

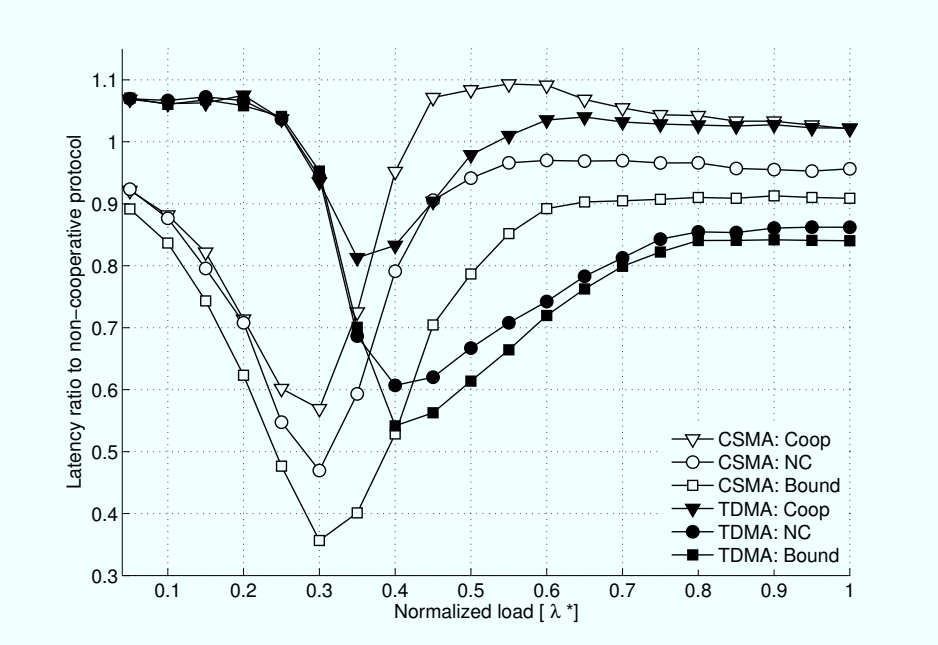


Figure 4.4. *Ratio of latency for cooperative schemes over the basic protocol vs. normalized load λ^*.*

Coop over CSMA-Basic, respectively. Such advantage also holds with the Bounds, proving that the TDMA environment is inherently more suitable for cooperation and confirming the discussion of Section 4.1. The later paragraphs will explain the reasons behind this performance. In addition, Both NC-enhanced protocols are close to the respective Bounds, thus verifying the efficiency of the schemes proposed in Sections 3.5.1.3 and 3.6.2.3.

Another key end-user metric is the MAC delay (Fig. 4.4). At low loads, TDMA-Coop and TDMA-NC actually underperform with respect to TDMA-Basic. The reason is due to channel correlation: with low traffic, the SINR is almost equal to the SNR and hence its variability is due almost exclusively to the channel time variability. In TDMA, the time between two retransmissions can be quite short (one frame), while in CSMA it may be much longer (one backoff interval). Hence, in CSMA the retransmissions experience less correlated SNR and thus also nodes with unfavorable position may in the end enjoy some time diversity (albeit after a long time). Instead, in TDMA at light loads the packets are often either immediately delivered at the first attempt or dropped.

At medium loads, cooperation starts being efficient and yields very significant improvements in both scenarios. At heavy loads, many retransmissions become network coded and thus the NC protocols have truly an edge over their decode-and-forward counterparts, reducing the delay by up to 10% and 18%, respectively. Such result is possible because of the shorter queuing time experienced by all packets and the absence of delay due to channel

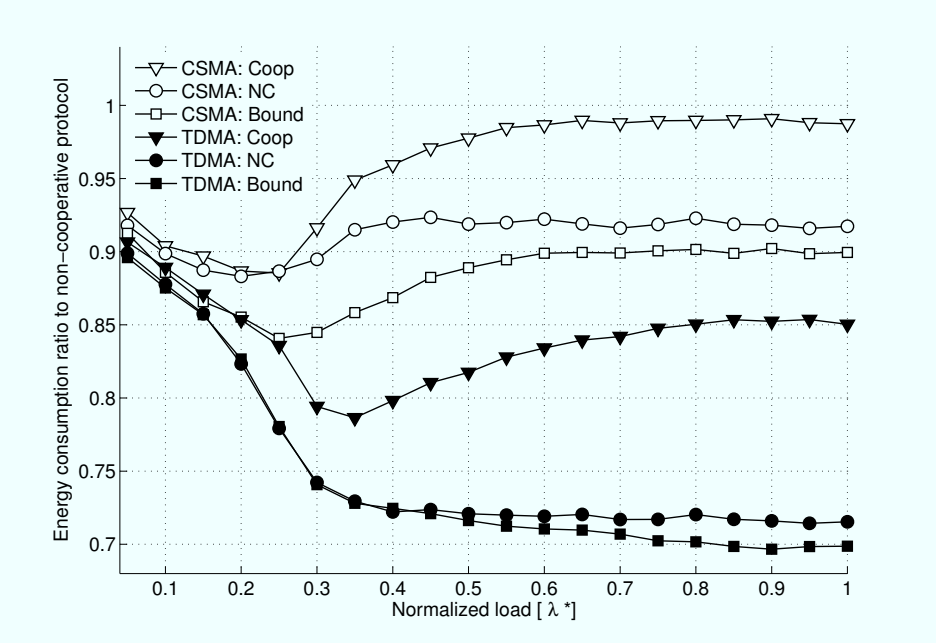


Figure 4.5. *Ratio of transmission energy consumption for cooperative schemes over the basic protocol vs. normalized load λ^*.*

contention for the retransmitted frames. It is also remarkable that at saturation both TDMA-Coop and CSMA-Coop have slightly worse delay than the reference protocol. This is again due to their more fair behavior towards terminals with severe loss rates: high-throughput nodes sacrifice part of their bandwidth (and of their low-delay packets) to help other terminals which intrinsically suffer a worse delay. Instead, NC is able to reduce the impact of this drawback and hence the latency is shorter. As a concluding remark on the delay, the better performance of TDMA-Bound over CSMA-Bound at saturation proves once again that TDMA is a more suitable environment than CSMA for advanced cooperation.

The final end-user metric is transmit energy consumption, depicted in Fig. 4.5 and defined as the ratio of the total energy spent in transmission to the number of delivered data bits. The observations follow what has been discussed so far: the metrics improve at medium load, while for high loads the unfairness of CSMA-Basic or TDMA-Basic may curb the gains. The NC protocols are however immune to this effect because of the aforementioned mechanisms. Notice again the significant improvement achieved by TDMA-Bound (and TDMA-NC) over CSMA-Bound, which testifies once again the inherent superiority of cooperation in a TDMA environment.

In our work, we have also carried out several simulation studies over relevant parameters in order to understand how general the conclusions drawn so far are. Specifically,

we have investigated the impact of different number of nodes per cell, of the carrier sense threshold to drop the relay contention $\Gamma_{CS}^{(rel)}$ (for CSMA) and of the maximum number of cooperative slots per frame n_{coop} (for TDMA). Only the qualitative results will be reported here due to space constraints. By varying the cell density it has been observed that the relative gains of CSMA-NC and TDMA-NC increase as the number of nodes per cluster grows. The reason is that the additional interference will generate more retransmissions and thus the recovery capacities of HARQ and NC retransmissions are all the more important in these cases. Such effects are more noticeable in CSMA based protocols, where the access to the medium is less coordinated and thus the amount of interference is more clearly dependent on the number of terminals per cell.

As for MAC dependent parameters, in CSMA it is arguable that there exists an optimal value for $\Gamma_{CS}^{(rel)}$. Too low a value will force too many relays drop the contention, while an exceedingly large threshold will generate excessive interference, again thwarting cooperation. The optimal performance is achieved at -86 dBm (3 dB above the value used in the simulations), but the gain is relatively limited (additional 2% in throughput) and the performance does not critically depend on such choice. As far as centralized systems are concerned, the relative gain of TDMA-NC over the other protocols reaches a maximum when at most 3 cooperative slots per frame are allocated. Setting too few (2 slots) will reduce the effectiveness of cooperation and NC, while too many (4 slots) will slightly reduce the performance gain, since too much time is devoted to performing retransmission rather than injecting new traffic.

In conclusion, this discussion on the end-user metrics has shown that TDMA based cooperative protocols are marked by fundamentally better performance than comparable CSMA based systems. Such finding is also supported by the parametric studies, because the relative gains do not appreciably vary as the system configuration is modified. The next paragraphs will explain the roots of this difference.

4.3.2 Metrics on the Efficiency of Cooperation

In Section 4.3.1, we described the improvements brought by cooperation and NC HARQ to metrics like throughput and delay. However, we did not explore why cooperation would be more efficient in one environment rather than another or what are the elements that prevent these techniques from achieving their full potential. This part wants to fathom these issues and investigate several metrics specifically related to cooperation and Network Coding. The results reported here are computed on CSMA-NC and TDMA-NC.

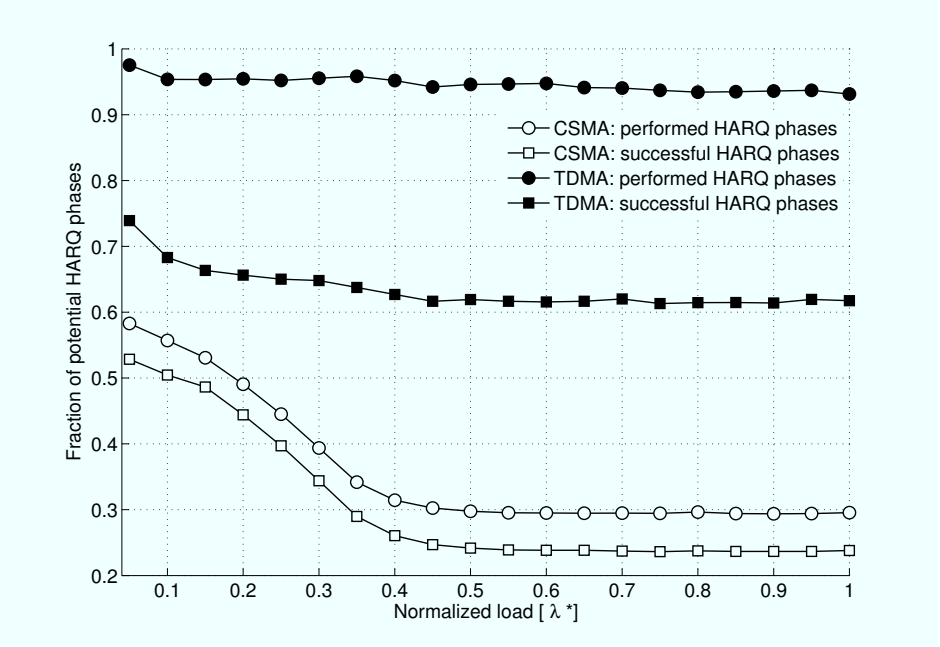


Figure 4.6. *Efficiency of hybrid ARQ vs. normalized load λ^*.*

The first question that comes to mind is how efficient is cooperation, i.e., how often a cooperative retransmission (simple or network coded) is carried out given that both source and destination need it, and how often such a retransmission is successful. The answers are reported in Fig. 4.6. The first apparent observation is that in TDMA a cooperative retransmission is performed far more often than in CSMA. This is due to the centralized structure of the cell, where one node collects all information and can schedule the necessary relays. In CSMA this information gathering is not present and thus there may be available relays, but because of the carrier sense mechanism they may not act. This is confirmed by the fall of the CSMA curves at high loads: the harsher interference may often fool carrier sense based cooperative mechanisms. Instead, such a behavior is virtually absent in TDMA, as polling packets and allocated cooperative data will always be sent and will not be postponed or canceled because of the medium access policy. It may also be noticed that the success rate in CSMA is higher than in TDMA, and the reason lies in the different way a node may become cooperator. In our implementation of CSMA-Coop and CSMA-NC, a terminal joins the relay election procedure not only if it correctly received the packet sent by the source, but also if it decodes the NACK sent by the intended receiver. This implies that the SINR to the destination cannot be too weak. On the other hand, TDMA applies the milder requirement SINR $> \Lambda_{coop}$. Thus, on average the relay-destination channel used for HARQ will be worse than in CSMA and cooperation will be somewhat less successful. However, note that the

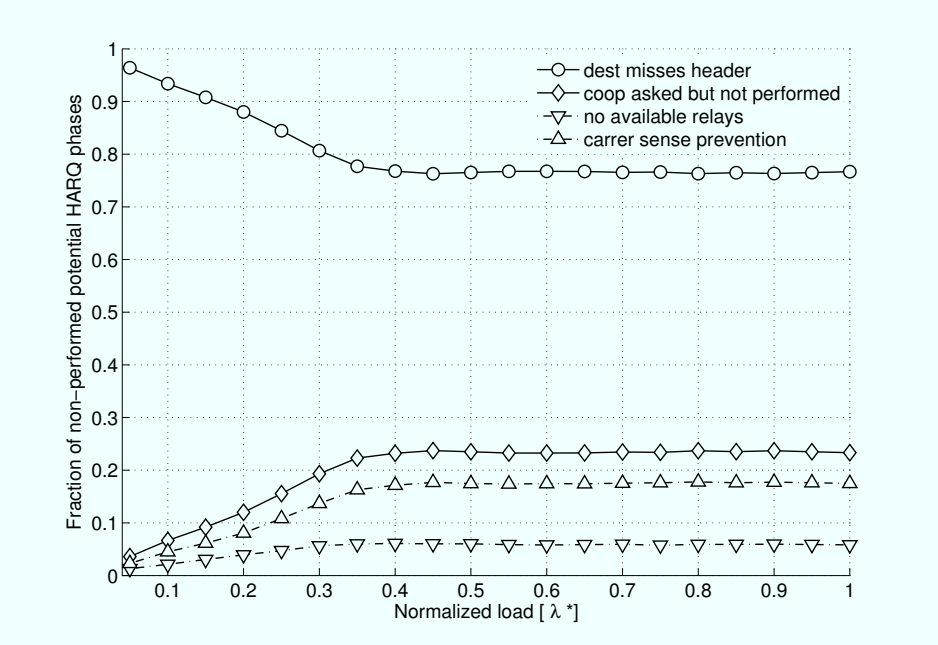


Figure 4.7. *Reasons that prevent hybrid ARQ when it is needed in a CSMA environment vs. normalized load λ^*.*

success rate in TDMA cooperative retransmissions is about 2/3, while in CSMA it is about 3/4, hence not dramatically higher.

If all potential cooperative phases were performed, the top curve for CSMA and TDMA would be equal to one for all loads. The gap between the actual and ideal curves for CSMA is explained in Fig. 4.7, which illustrates the causes that prevent a cooperative retransmission from being carried out. In CSMA two main reasons may lead to this: i) the destination does not decode the data packet header (and thus does not even request a retransmission), or ii) the relaying is requested but is not performed. The latter event is in turn due to two problems: ii-a) there may be no available cooperators, or ii-b) there may be such nodes but nobody wins the contention because the carrier sense makes all participants drop. The clear message from Fig. 4.7 is that the vast majority of the non performed retransmissions is due to the loss of the data header. In this case, the destination is not even aware of the sender's identity, and thus cannot ask for help. This is an intrinsic limit of CSMA and will be a major difference with respect to the scheduled system. On the other hand, the lack of relays is not particularly relevant, since nodes close to the source have probably decoded the source's packets because of the carrier sense mechanism. However, carrier sense does take its toll during the relay election phase.

While Fig. 4.7 explored the reasons why cooperation is not performed in CSMA, Fig. 4.8

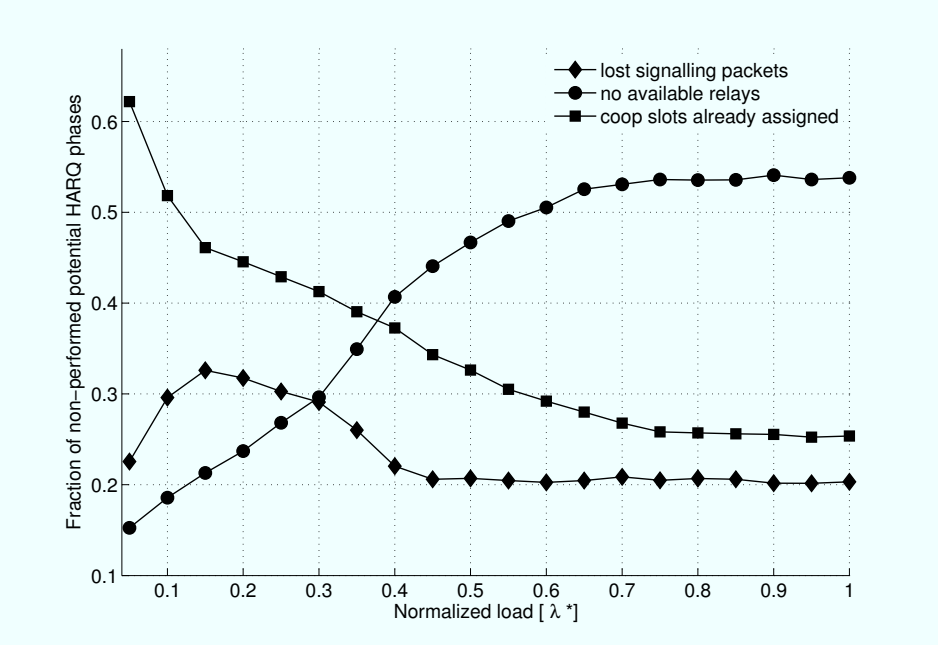


Figure 4.8. *Reasons that prevent hybrid ARQ when it is needed in a TDMA environment vs. normalized load λ^*.*

analyzes the same phenomenon for TDMA. Three main problems may avert a cooperative phase: i) a signaling packet (i.e., a beacon or a polling packet) is not received, ii) no relays are available, or iii) all the slots reserved for cooperation in the next frame have already been allocated. For high loads, the main cause is the lack of helpers (i.e., no node correctly decoded the lost frame), which is in direct contrast to CSMA, where this problem is negligible in comparison to the header loss or the failure of the election procedure. Instead, at low loads interference is not a major issue and thus most cooperative phases do find a relay. Hence, the main issue may be a lack of sufficient slots reserved to cooperation, but this happens in a really negligible fraction of cases (see Fig. 4.6). In conclusion, the comparison between Fig. 4.7 and Fig. 4.8 (for CSMA and TDMA, respectively) suggests that the main reasons for missed cooperation in CSMA and TDMA are the header loss and the lack of relays, respectively. Note that each cause is negligible or even impossible in the other scheme, and this highlights the deep differences between CSMA and TDMA towards cooperation.

It is reasonable to wonder whether the substantially larger gains of TDMA over CSMA may be due to a larger relay population, i.e., whether the MAC schemes affect the number of relays that join the contention. Fig. 4.9 gives an essentially negative answer to the question, as it shows the histogram of the number of nodes that join a relay election in CSMA and TDMA. The average number of such terminals is quite similar in both environments. Hence

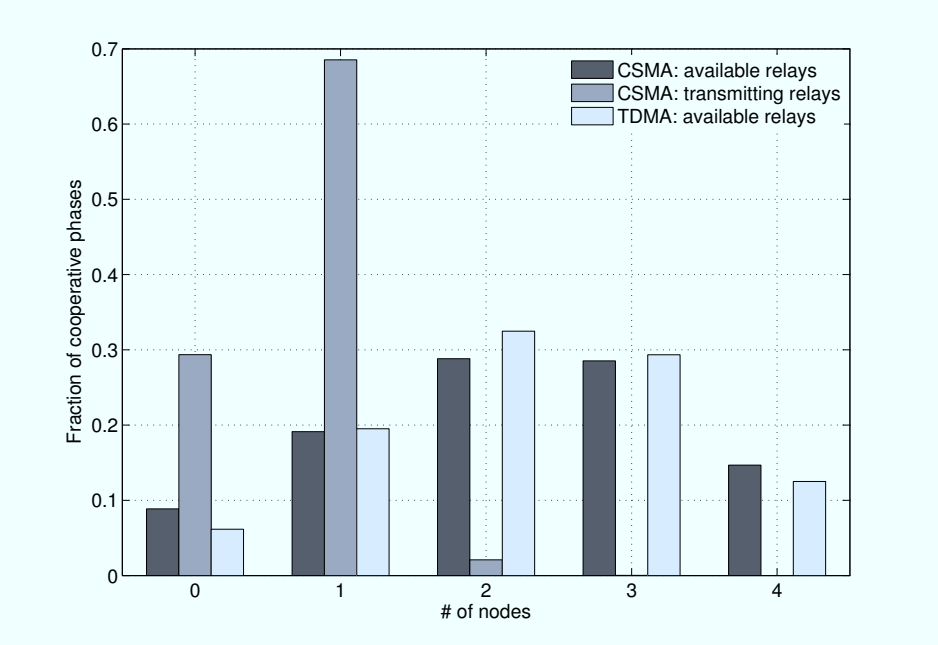


Figure 4.9. *Estimated statistical distribution of different classes of relays in CSMA and TDMA.*

the improvement of TDMA is due to the better coordination among nodes, not to topological reasons or intrinsically different levels of interference.

The previous graphs analyzed the relationship between the two different MAC schemes and cooperative techniques. The last two plots will focus on the efficiency of NC HARQ. The first question is how often are NC retransmissions possible and how many of them are actually carried out. Fig. 4.10 shows the percentage of potential NC phases and of actually performed NC retransmissions out of all the cooperative stages. It is clear that NC occurs more often in CSMA rather than in TDMA and this is due to the fact that in CSMA a cooperative phase is requested only if the destination recovered at least the header. Once again, this implies a minimum quality level on the source-destination channel, and increases the probability of meeting the Λ_{Th} requirement. Instead, no such constraint exists in TDMA, and the destination can request cooperation after decoding the beacon but missing the data packet. Such condition may enable to ask for cooperation also at lower SINR, where NC is not recommended. Let us remark though, that putting Fig. 4.6 (percentage of actually performed HARQ phases) and Fig 4.10 (fraction of NC HARQ) together, it turns out that at high loads $30\% * 65\% = 20\%$ and $90\% * 40\% = 36\%$ of all potential cooperative retransmissions are network coded for CSMA and TDMA respectively. This shows that NC occurs twice more often in TDMA than in CSMA, and this justifies why NC protocols are more effective in TDMA than in CSMA (see Section 4.3.1).

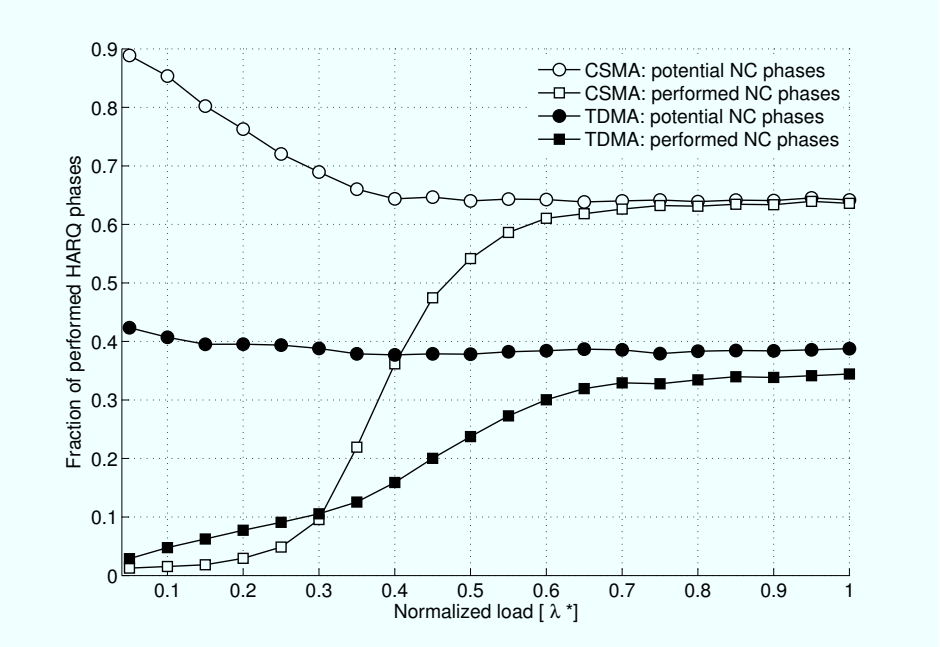


Figure 4.10. *Efficiency of advanced hybrid ARQ techniques vs. normalized load λ^*.*

The last figure of interest is Fig. 4.11, which explains why plain cooperation is performed instead of a NC retransmission. In TDMA, three main problems may arise: the source-destination SINR is too weak; no relay has a packet to transmit for this destination; or the SINR between relay and destination is also too low. In CSMA this last fact does not happen, since a node can become a relay only if it decodes the destination's NACK, thus putting a lower bound on the transmission quality. At light loads, the lack of packets for the destination is the major impairment towards NC in CSMA, while at high loads (as the interference grows) the low source-destination SINR is the foremost problem. The same phenomenon occurs in TDMA, but the weak source-destination SINR becomes a secondary problem for this MAC only at loads lighter than those shown in Fig. 4.11.

4.4 Conclusions

This chapter has performed a first exploration of the fundamental contrasts between carrier sense- (CSMA) and time division-based (TDMA) medium access control with respect to hybrid ARQ. It has been highlighted that basic but essential differences between these two environments have a deep impact on the efficiency of cooperation and especially of its more advanced versions, like network coded retransmissions. The completely distributed nature of CSMA has been shown to severely stymie relaying, as it both often prevents nodes

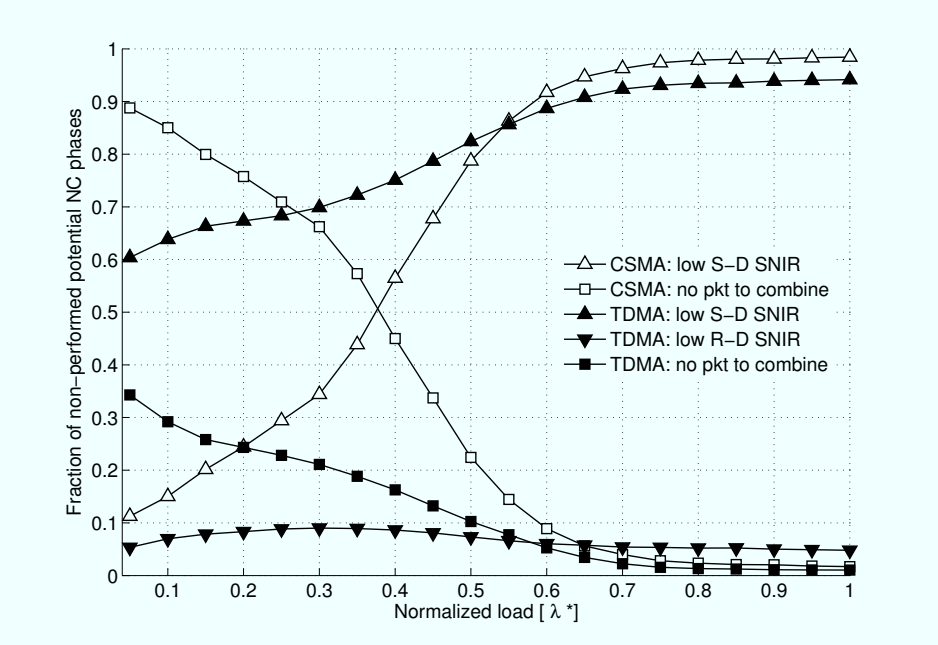


Figure 4.11. *Reasons that prevent advanced hybrid ARQ when it would be possible in a CSMA and a TDMA environment vs. normalized load λ^*.*

from caching packets for later combining and hinders medium access for cooperators. Conversely, TDMA avoids such issues altogether by resorting to a centralized scheduler, and achieves a much higher efficiency. The better coordination enables TDMA to attain far more noticeable performance gains.

Chapter 5

Some Insights on Cooperation in Carrier Sense Based Ad Hoc Networks

Chapter 4 has shown by means of extensive simulation studies how CSMA-based link layers hinder effective implementations of cooperative solutions that rely on the decode-and-forward paradigm. Nonetheless, carrier sense stands as one of the most interesting medium access policies for wireless environments, as it perfectly embodies the distributed and self-organizing paradigm that characterizes ad hoc networks. Also, CSMA is extremely widespread, e.g., it is the standard for WLANs, and finds lots of practical applications. Therefore, even though channel impairments like noise, fading and shadowing as well as the well known hidden and exposed terminal problems [68] may limit its effectiveness, such a multiple access control still represents a significant and insightful test for relaying protocols. In this perspective, we devote this chapter to further analyzing the relationship between collaborative strategies based on relaying and CSMA in non-centralized ad hoc networks.

We focus our work on two main aspects. On the one hand, we investigate how the interference distribution induced by this rather simple MAC influences statistics that are of fundamental importance for the efficiency of cooperation, such as the number and the position of nodes available for relaying. On the other hand, following the approach that also characterized the previous chapters of this thesis, we analyze some practical issues that are generally neglected in theoretical studies, e.g., header decoding and interaction of multiple simultaneous communications within a network, and we show how they affect the potential advantages offered by cooperative schemes.

In order to provide a comprehensive discussion, and so as to draw broadly applicable conclusions, we consider two classes of strategies, namely *reactive* and *proactive*. The for-

mer set refers to policies that trigger the intervention of relays only in the event of a missed decoding at the destination for the transmission performed by the source, under the rationale of enhancing plain failure recovery procedures by taking advantage of spatial diversity. Conversely, we group as proactive those strategies that exploit channel state information to pre-emptively split a communication in two phases, transferring first the payload from the source to a relay node and then letting such terminal send redundancy to the intended destination, with the aim of maximizing the sum rate of the link. Incidentally, notice that the decode-and-forward paradigm as well as the hybrid cooperative-network coded approach discussed in Chapters 3 and 4 fall under the category of reactive schemes. Therefore, some of the observations that will be inferred here also apply to those schemes, and will confirm the discussions carried out earlier in this thesis.

In particular, some of the key observations that we will present and discuss are the following:

- the channel sensing mechanism biases the geographical distribution of the available relays, reducing the probability of finding any of them in the *optimal relay region*[1] and thus affecting the effectiveness of cooperation;
- such drawbacks do not depend on the implementation of a specific relaying policy, but rather stem from the intrinsic interaction of carrier sense multiple access and cooperation;
- relayed transmissions, if not properly handled, may result in unexpected interference perceived at neighboring receivers, with detrimental effects on the overall network performance;
- practical issues, such as header decoding and receiver synchronization, hamper cooperative techniques in large-scale non-centralized ad hoc networks.

Furthermore, one of the main contributions of this work is to highlight the importance and the effect of the environment in which cooperation, and other advanced techniques, have to operate. The methodology we adopt makes use of analysis and simulation, and we believe that both components are fundamental to the investigation presented in this chapter. Indeed, on the one hand, the analysis of simplified scenarios enables the identification of basic issues and features of the framework, and thus allows a proper design of a cooperative scheme. Nevertheless, a detailed simulation of a large network, besides highlighting many

[1] By optimal relay region we mean the geographical region in which the existence of an available relay provides the best link performance.

practical aspects that analysis cannot account for, provides a realistic environment for testing the effectiveness of the considered protocols, and is therefore a key step.

The rest of this chapter is organized as follows. Section 5.1 describes the system model that we consider together with the notation used in this part of the thesis. Then, in Section 5.2, we introduce an analytical framework for reactive cooperation, shedding some light on the issues induced by CSMA. Such remarks are exploited in Section 5.3 to design a protocol capable of reaping at the utmost the potential of relaying in the considered environment. The performance of the proposed scheme is then thoroughly studied by means of a simulation campaign in Section 5.4. The second part of the chapter, in turn, is dedicated to proactive cooperation, which is introduced in Section 5.5. In this context, Section 5.6 proposes a scheme that implements the pre-emptive relaying paradigm, while Section 5.7 illustrates an analytical approach to study the effectiveness of such a solution in carrier sense-based environments. Finally, Section 5.8 investigates the reported protocol in large and more realistic networking scenarios.

5.1 System Model and Notation

Throughout this chapter we focus on plain carrier sense multiple access, epitomized by the IEEE 802.11 DCF without channel negotiation, while an investigation including collision avoidance is left as part of future work. For a complete description of the basic CSMA link layer, the reader can refer to Section 3.5.1.1.

In order to keep our scenario general and to have a framework compatible with many other theoretical studies, we model wireless links considering Rayleigh fading, so that the received power η_{n_1,n_2} over the $\boldsymbol{N}_1$-$\boldsymbol{N}_2$ channel is an exponential random variable with mean $P\delta_{n_1,n_2}^{-\alpha}$, where P is the transmission power, δ_{n_1,n_2} is the Euclidean distance between the two terminals and α is the path loss exponent. Moreover, we consider a channel capacity model for packet decoding, so that the failure probability in retrieving an L-bit payload for a transmission that starts at t_s and lasts for T seconds can be expressed as $\Pr\{\mathcal{L} < L\}$, where $\mathcal{L}$ is the number of decoded information bits given by:

$$\mathcal{L} = \int_{t_s}^{t_s+T} \mathcal{C}(\gamma(t))dt\,, \tag{5.1}$$

$C(\gamma(t)) = B\log_2(1+\gamma(t))$ being the instantaneous link capacity between sender and receiver with bandwidth B and signal to interference and noise ratio (SINR) $\gamma(t)$.

In the analytical computations that will be performed, we denote scalars with regular

font, whereas vectors are represented in bold. Moreover, the notation $\mathcal{K}^x(x)$ indicates quantity $\mathcal{K}$ conditioned on the variable x.

5.2 Impact of CSMA on Relays Distribution

We start our investigation by focusing on *reactive* cooperative schemes. These techniques, indeed, represent an interesting case study both because of the simple yet smart principle that they propose and because of the large deal of research that has been devoted to the analysis of their potential [28–30]. As discussed in the previous chapters (see, for instance, Section 3.2), such a communication paradigm proposes, in case of a decoding failure over a direct link, a neighboring node of the source-destination pair to act as relay by immediately sending redundancy to the latter on behalf of the former, so as to replace the time diversity that characterizes basic ARQ with spatial diversity.

As a first step in the direction of understanding the reciprocal influence between carrier sense-based medium access and reactive cooperation, we develop an analytical framework to highlight some key interactions in simple topologies. To this aim, let us consider a scenario with a source $\boldsymbol{S}$, a destination $\boldsymbol{D}$, and an interfering terminal $\boldsymbol{I}$ deployed on a region $\mathcal{A}$ at positions $p_s = \{x_s, y_s\}$, $p_d = \{x_d, y_d\}$ and $p_i = \{x_i, y_i\}$, respectively. All data transmissions in the network are performed at a fixed information bitrate ρ_{data} and involve a payload of L bits, thus lasting for $T = L/\rho_{data}$ seconds. Furthermore, for the sake of mathematical tractability, we assume fading coefficients to remain constant for the whole duration of a data exchange.[2] Without loss of generality, $\boldsymbol{S}$ accesses the channel at time $t_s = 0$, while the birth time t_i of the interfering communication is uniformly distributed in $[-T, T]$, see Fig. 5.1. Recalling the system model described in Section 5.1, the probability that a node $\boldsymbol{N}_2$ senses the medium idle given that terminal $\boldsymbol{N}_1$ is transmitting, i.e., the carrier sense constraint, is given by:

$$\mathcal{F}(p_{n_1}, p_{n_2}) = \Pr\{\eta_{n_1,n_2} + N < \Lambda\} = 1 - e^{-\frac{\Lambda - N}{P\delta_{n_1,n_2}^{-\alpha}}}, \tag{5.2}$$

where N is the noise floor and Λ is the carrier sense threshold.

We initiate our study by computing, for a topology configuration $\boldsymbol{p} = \{p_s, p_d, p_i\}$, the probability $\mathcal{I}(\boldsymbol{p})$ that the interferer induces an outage at the destination when carrier sense is employed as the medium access policy, thus triggering the need for cooperation. In order for this to happen, two conditions have to be met. In the first place, both $\boldsymbol{S}$ and $\boldsymbol{I}$ have

[2] Let us remark that the link capacity $C(t)$ reported in (5.1) may vary during a communication even in the presence of a constant fading, due to changes in the perceived aggregate interference level.

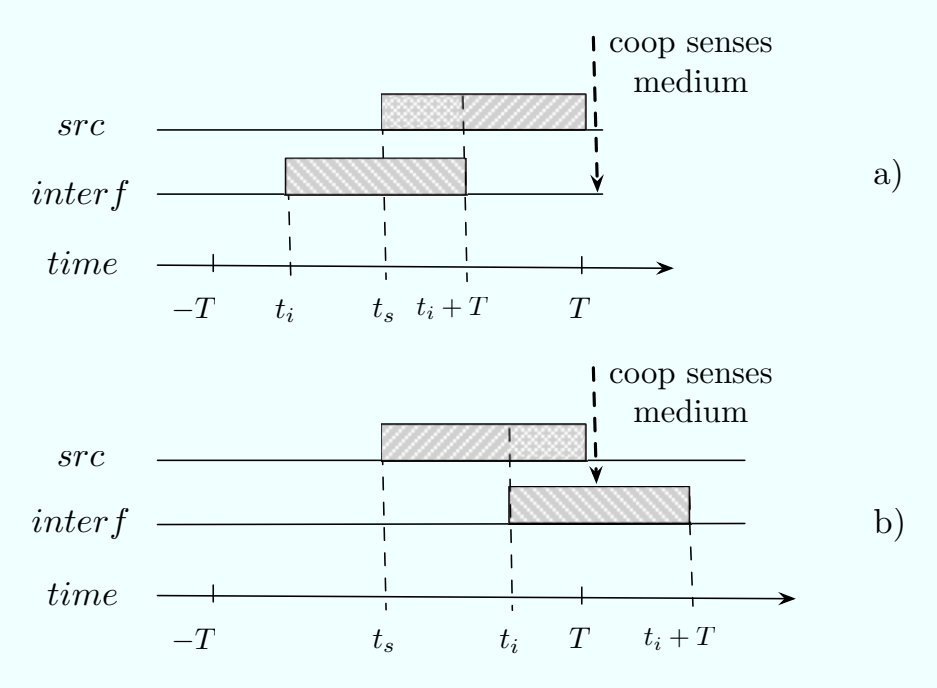


Figure 5.1. *Time diagram for the considered scenario.* $\boldsymbol{S}$ *accesses the channel at* t_s*, while* $\boldsymbol{I}$ *tries to send data at* t_i*. All communications last for* T *seconds, and the cooperating terminal senses the channel immediately after the end of* $\boldsymbol{S}'$ *transmission. a)* $t_i < t_s$*, b)* $t_i \geq t_s$*.*

to be granted access to the channel. This implies that either the source, for $t_i \leq 0$, or the interferer, for $t_i > 0$, are required to sense a power level on the medium below the carrier sense threshold despite the other node's ongoing transmission. Secondly, the interference level at $\boldsymbol{D}$ during the overlapping part of the two communications has to be high enough to prevent a successful decoding of the payload. The two events are statistically independent, since the former only involves the $\boldsymbol{I}$-$\boldsymbol{S}$ link, whereas the latter depends on the $\boldsymbol{S}$-$\boldsymbol{D}$ and $\boldsymbol{I}$-$\boldsymbol{D}$ channels. Therefore, conditioning on the birth time t_i of the interfering transmission, and recalling the CSMA constraint reported in (5.2), the outage probability can be factorized as:

$$\mathcal{I}^{t_i}(\boldsymbol{p}, t_i) = \mathcal{F}(p_i, p_s) \cdot \mathcal{O}^{t_i}(\boldsymbol{p}, t_i)\,, \tag{5.3}$$

where $\mathcal{O}^{t_i}(\boldsymbol{p}, t_i) = \Pr\{\mathcal{L}^{t_i}(t_i) < L\}$ is the probability of decoding less than L information bits at $\boldsymbol{D}$ conditioned on t_i. Notice that for both $t_i \leq t_s$ and $t_i > t_s$, $T - |t_i|$ seconds of $\boldsymbol{S}'$ payload are affected by $\boldsymbol{I}$'s transmission, whereas $|t_i|$ seconds are interference-free. Thus, applying (5.1), and defining as $\boldsymbol{\eta} = \{\eta_{s,d}, \eta_{i,d}\}$ the vector of the powers received at $\boldsymbol{D}$, the number of retrieved information bits at the destination evaluates to:

$$\mathcal{L}^{t_i}(\boldsymbol{\eta}, t_i) = |t_i| \cdot \mathcal{C}\left(\frac{\eta_{s,d}}{N}\right) + (T - |t_i|) \cdot \mathcal{C}\left(\frac{\eta_{s,d}}{N + \eta_{i,d}}\right)\,. \tag{5.4}$$

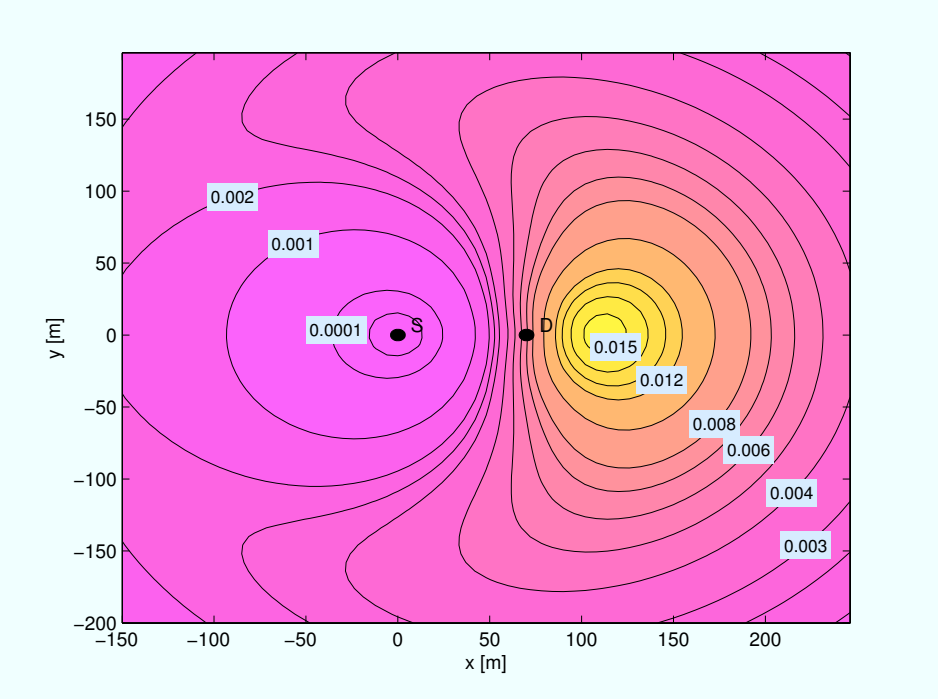


Figure 5.2. *Probability that an interfering node located at p_i causes an outage event at $\boldsymbol{D}$.* $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.

The explicit derivation of $\mathcal{O}^{t_i}(\boldsymbol{p}, t_i)$, though conceptually simple, is rather involved, and is reported in Appendix B.1. Taking advantage of this result, i.e., (B.2), we can compute the sought outage probability by numerical evaluation of:

$$\mathcal{I}(\boldsymbol{p}) = \frac{1}{2T} \cdot \int_{-T}^{T} \mathcal{I}^{t_i}(\boldsymbol{p}, t_i)\, dt_i \,. \tag{5.5}$$

Fig. 5.2 depicts $\mathcal{I}(\boldsymbol{p})$ for different positions of $\boldsymbol{I}$, when the source and the destination are placed at $p_s = \{0, 0\}$ and $p_d = \{60, 0\}$, respectively, with the coordinates being expressed in meters, and when all the other system parameters are set according to Tab. 5.1. It is possible to observe that the simple CSMA policy induces a strong bias in the spatial distribution of interferers when cooperation is needed. Indeed, the region in which $\boldsymbol{I}$ is most likely to induce a failure lies along the line connecting $\boldsymbol{S}$ and $\boldsymbol{D}$, right behind the destination. This result is rather intuitive, since such positions are at the same time geographically close to $\boldsymbol{D}$, so that $\boldsymbol{I}$ can heavily interfere with the reception, and sufficiently far from $\boldsymbol{S}$, so that the carrier sense constraint can be met with high probability.

Even more interesting for our discussion is the fact that such an interferer distribution significantly influences the availability and the position of relay nodes, with potentially detrimental effects on reactive cooperation. To investigate this aspect, let us consider a terminal $\boldsymbol{C}$ located at $p_c = \{x_c, y_c\}$, and assume that it may act as helper for the $\boldsymbol{S}$-$\boldsymbol{D}$ link if two

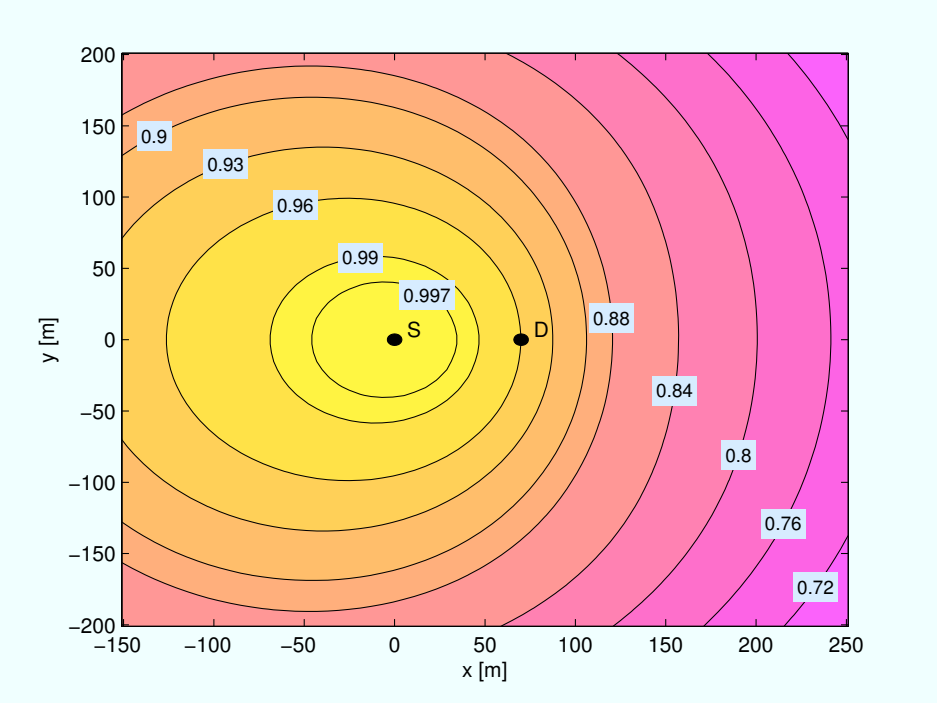


Figure 5.3. *Probability that a node located at p_c is available for cooperation given that the interfering transmission has already come to an end. $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.*

conditions are satisfied: i) it has decoded the payload sent by the source despite the interfering transmission which induced an outage at the destination; and ii) it senses the medium idle at time $t = T$ and is therefore allowed to send redundancy to $\boldsymbol{D}$. Both requirements clearly depend on the position of $\boldsymbol{I}$, which, in turn, varies with the birth time of the interfering communication, as discussed earlier and as expressed by (5.5). Therefore, defining $\boldsymbol{p}' = \{p_s, p_d, p_c\}$ as a vector describing the topology under consideration, the spatial distribution of cooperators can be determined by first evaluating the probability $\mathcal{H}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$ that $\boldsymbol{C}$ meets i) and ii) conditioned on p_i and t_i, and by then averaging this distribution over the interference conditions. In particular, the probability $\mathcal{H}^{t_i}(\boldsymbol{p}', t_i)$ that $\boldsymbol{C}$ is available given that the interfering transmission starts at t_i can be computed as:

$$\mathcal{H}^{t_i}(\boldsymbol{p}', t_i) = \frac{\int_{\mathcal{A}} \mathcal{H}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i) \cdot \mathcal{I}^{t_i}(\boldsymbol{p}, t_i)\, dp_i}{\int_{\mathcal{A}} \mathcal{I}^{t_i}(\boldsymbol{p}, t_i)\, dp_i}\,, \tag{5.6}$$

with $\mathcal{I}^{t_i}(\boldsymbol{p}, t_i)$ given by (5.3). In turn, the unconditional probability $\mathcal{H}(\boldsymbol{p}')$, i.e., the sought cooperators distribution, is obtained as:

$$\mathcal{H}(\boldsymbol{p}') = \frac{1}{2T} \int_{-T}^{T} \mathcal{H}^{t_i}(\boldsymbol{p}', t_i)\, dt_i\,. \tag{5.7}$$

Table 5.1. *Parameters used in our studies*

Transmission power	10 dBm
Noise Floor	-102 dBm
CS thresholds, Λ, Λ_{rel}	-100 dBm
Detection threshold	-96 dBm
Path loss exponent, α	3.5
Maximum Doppler shift	40 Hz (5 m/s)
Carrier Frequency	2.4 GHz
Bandwidth, B	1 MHz
Headers and control pkt information bitrate, ρ_{ctrl}	0.532 Mbps
Data pkt information bitrate, ρ_{data}	2.1 Mbps
Minimum src-dest capacity for proactive cooperation, ρ_{min}	0.95 Mbps
Slot, DIFS, SIFS duration	10, 128, 10 μs
Initial contention window index, CW_{start}	5
Number of slots used for relay contention, CW_{rel}	16
SRL - cooperative, basic protocols	4, 5
Levels of quantization for $\gamma_{s,d}$ in the NACK frame for DHARQ	8
Interference margin for reactive cooperation ε	0.06
Interference margin for proactive cooperation, ε'	0.15
Payload	5000 bit
Header/ACK/NACK	112 bit
Simulation Time	50 s

In order to derive $\mathcal{H}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$, it is useful to distinguish the cases for $t_i \leq 0$ and $t_i > 0$. In the former condition, as reported in Fig. 5.1, $\boldsymbol{I}$'s transmission has already come to an end when $\boldsymbol{C}$ potentially senses the medium for channel access. Therefore, requirement i) is always met, and the only constraint for the node to act as cooperator is the decoding of the payload sent by $\boldsymbol{S}$ given the interference level generated by $\boldsymbol{I}$. This event, in turn, can be analyzed following the same procedures we discussed for the study of the outage at $\boldsymbol{D}$, and we immediately get:

$$\mathcal{H}_{-}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i) = 1 - \mathcal{O}^{t_i}(p_s, p_c, p_i, t_i) \ , \tag{5.8}$$

where the subscript on $\mathcal{H}_{-}^{p_i,t_i}$ indicates that we are considering values of $t_i \leq 0$, and $\mathcal{O}^{t_i}$ $(p_s,$ $p_c,$ $p_i,$ $t_i)$ can be obtained by computing (B.2), reported in Appendix B.1, on the topology

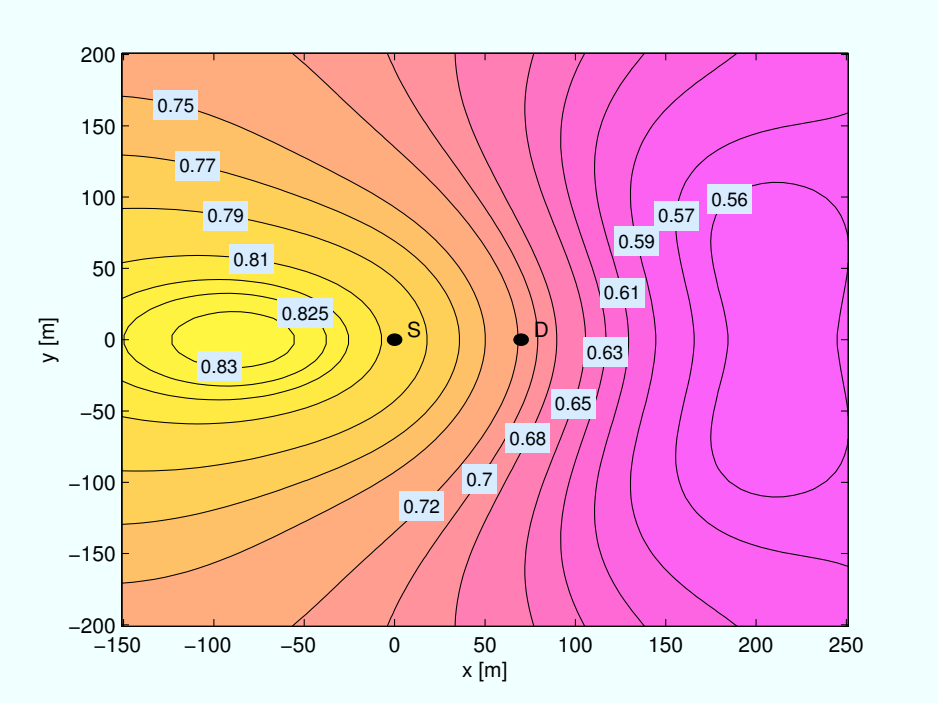


Figure 5.4. *Probability that a node located at p_c is available for cooperation given that the interfering transmission is still active.* $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.

$\{p_s, p_c, p_i\}$. Conversely, for $t_i > 0$, $\boldsymbol{I}$ is still transmitting when the relay candidate checks the power level on the medium, and thus the interference level perceived at $\boldsymbol{C}$, i.e., $\eta_{i,c}$, determines both the payload decoding and the carrier sensing events. This implies that, unlike what has been done in the first part of our discussion, the conditional probability $\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$ cannot be factorized to take into account separately of requirements i) and ii).[3] Its derivation, consequently, becomes rather involved, and for the sake of clarity is postponed to Appendix B.2. In this section, instead, we focus on the discussion of the co-operator distribution obtained by applying the presented analytical approach to a topology with source and destination 60 m apart, located once again at $p_s = \{0, 0\}$ and $p_d = \{60, 0\}$ respectively.

As a starting point, Fig. 5.3 reports the probability that a terminal is available as relay conditioned on having the interfering transmission start before t_s, i.e., (5.7) restricted to the integration domain $[-T, 0)$. In this situation, as discussed, the only requirement $\boldsymbol{C}$ has to meet is to have decoded the payload sent by the source. From the plot it is apparent how such a constraint is satisfied with higher probability when the terminal is located in the

[3] In accordance to the notation introduced earlier, the subscript on $\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$ specifies the fact that we are dealing with interfering transmissions characterized by $t_i > 0$.

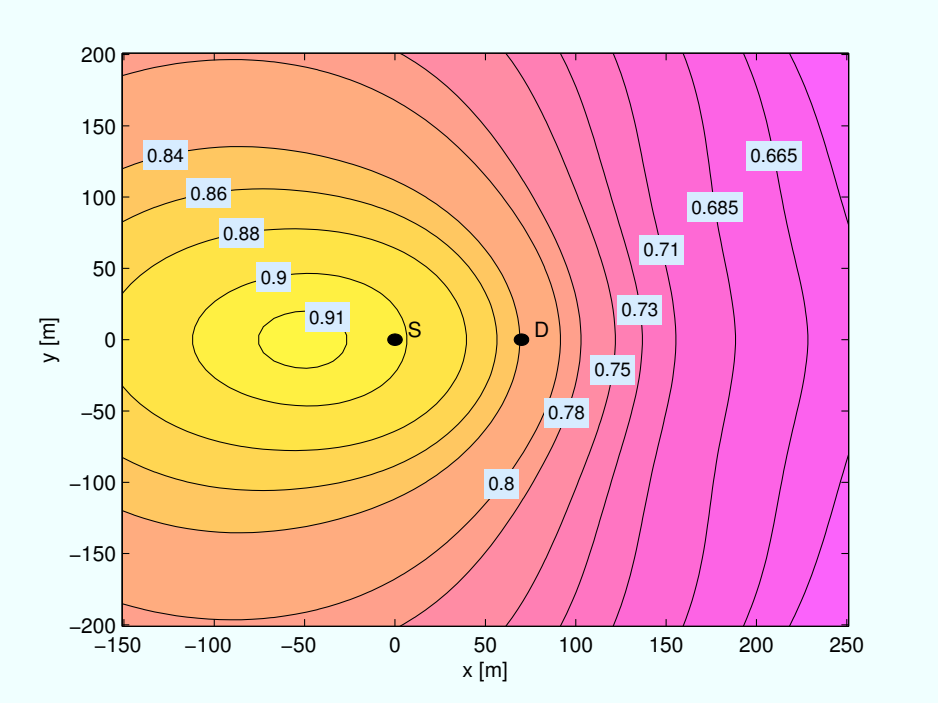


Figure 5.5. *Average probability that a node in p_c is available for cooperation.* $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.

proximity of $\boldsymbol{S}$. This result stems from the interferer distribution induced by CSMA and presented in Fig. 5.2, which hampers payload retrieval especially for terminals which are close to $\boldsymbol{D}$.

The spatial distribution of cooperating nodes becomes even more biased when $\boldsymbol{I}$'s transmission is still active at the end of the $\boldsymbol{S}$-$\boldsymbol{D}$ data exchange, as shown in Fig. 5.4, which depicts (5.7) computed on the restricted integration domain $[0, T]$. In such a case, relaying may not take place even when the packet sent by $\boldsymbol{S}$ has been correctly received at $\boldsymbol{C}$, since the terminal may refrain from cooperating due to the carrier sense constraint, so as not to harm the communication being performed by $\boldsymbol{I}$. The impact of this condition on the availability of relays is significant even in the minimal topology considered here. Indeed, when only packet decoding from $\boldsymbol{S}$ is requested, finding a cooperator in the proximity of the destination is only approximately 5% less probable than having it close to the source (see Fig. 5.3). Conversely, such a gap more than doubles when $\boldsymbol{C}$ is also subject to CSMA (see Fig. 5.4).

Finally, Fig 5.5 reports the spatial distribution $\mathcal{H}(\boldsymbol{p}')$ of potential cooperators averaged over the whole birth time interval of the interfering communication. As a first remark, we notice that nodes available for cooperation are concentrated close to $\boldsymbol{S}$, and thus offer a limited advancement towards the destination. As will be discussed in Section 5.4, and as confirmed by further analytical results not reported here due to space constraints, this effect

increases as the distance between $\boldsymbol{S}$ and $\boldsymbol{D}$ becomes larger. This stems from the fact that the closer the source to the destination, the more the area reserved by the carrier sense mechanism to the ongoing transmission extends over the addressee's neighborhood, improving decoding capabilities of relay candidates surrounding $\boldsymbol{D}$. However, cooperation is more likely to be required when the distance between the communicating terminals is large, due to the poorer link quality induced by the higher path loss.

We can therefore already infer from this simple analysis that CSMA has an unfavorable impact on cooperative schemes, since, when relaying is needed, cooperators are likely to be far from the destination they want to reach.

5.3 Distributed HARQ in CSMA-Based Networks

The analytical framework developed in Section 5.2 has highlighted how CSMA influences the geographical position of nodes available for reactive cooperation, concentrating them in the proximity of the source rather than close to the destination of a wireless link. This result, though obtained in simple topologies, offers important insights on the design of relaying protocols in carrier sense-based networks.

In the first place, we can infer that the biased distribution of relay candidates discourages the use of solutions that implement some form of *cooperative routing* [69]. These approaches, indeed, exploit cooperation to build opportunistic paths toward the final destination, under the rationale of finding relays that provide a significant geographical advancement towards the addressee of the payload. Simulations of a simple cooperative routing protocol in a detailed network scenario, whose results are not reported here due to space constraints, have confirmed this intuition.

From this viewpoint, instead, distributed Hybrid Automatic Repeat reQuest (HARQ) [70] appears as an ideal solution to reap at the utmost the benefits of cooperation when CSMA is used as medium access policy. According to these schemes, also known as *coded cooperation* protocols, the destination collects fragments of a codeword obtained encoding the payload and received from the source and from relay nodes. Not only is coded cooperation able to exploit the greatest diversity order compared to all other approaches, but also it is likely to undergo a smaller performance reduction as a consequence of the worse position of relays, achieving its best results exactly when the cooperators are concentrated towards the source [71,72].

Starting from these remarks, in this section we present a medium access protocol named DHARQ that extends plain CSMA to take advantage of distributed HARQ of type II in

non-infrastructured ad hoc networks. As in the non cooperative case, according to DHARQ a source $\boldsymbol{S}$ accesses the channel after the carrier sense-based backoff procedure (see Section 3.5.1.1). Let us assume that no information on the channel connecting $\boldsymbol{S}$ to its addressee is available, so that the node sends its payload to the destination $\boldsymbol{D}$ at a fixed information bitrate ρ_{data}. If the transmission succeeds, an ACK at rate $\rho_{ctrl} < \rho_{data}$ is sent by the addressee, and the communication comes to an end. On the contrary, if the reception fails, two conditions may have occurred at $\boldsymbol{D}$, as discussed in Section 3.5.1.2: i) the header has not been successfully decoded, or ii) the header has been decoded, yet the payload was corrupted.[4] The former case may happen if the receiver's radio is synchronized to another ongoing transmission, e.g., due to the hidden terminal problem, or in the presence of harsh channel and interference conditions. In such a situation, the destination is not even aware of the attempt performed by $\boldsymbol{S}$. Therefore, no feedback can be sent, and the source relies on the basic CSMA approach for retransmissions, performing other attempts after suitable backoff intervals, until either the payload is successfully delivered or the SRL is reached. Conversely, if condition ii) occurs, the destination caches the corrupted version of the payload and transmits a Not ACKnowledgement (NACK) packet at rate ρ_{ctrl}, explicitly asking for an immediate retransmission from a relay node. Not only does the NACK frame trigger a hybrid ARQ procedure, but also it includes an indicator on the quality of the reception at $\boldsymbol{D}$. This is achieved by means of a k bit field in the frame that contains a quantized version $\tilde{\gamma}_{s,d}$ of the average SINR perceived by the node during data reception. Any terminal $\boldsymbol{C}$ that receives the NACK packet and that has correctly decoded the payload transmitted by the source performs adaptive rate selection by computing the maximum information bitrate ρ^c_{rel} it can use to perform a successful coded cooperative retransmission. To this aim, $\boldsymbol{C}$ evaluates the number $\mathcal{L}$ of decodable information bits at the destination after a relaying phase assuming constant SINRs during packet receptions and applying (5.1):

$$\mathcal{L} = T_{s,d} \cdot \mathcal{C}(\tilde{\gamma}_{s,d}) + T_{c,d} \cdot \mathcal{C}(\gamma_{c,d}) \,, \tag{5.9}$$

where $T_{s,d} = L/\rho_{data}$, $T_{c,d} = L/\rho^c_{rel}$, and $\gamma_{c,d}$ is the average SINR perceived by the relay candidate during the reception of the NACK packet. Note that (5.9) describes the number of decodable information bits at $\boldsymbol{D}$ if node $\boldsymbol{C}$ acts as cooperator considering an estimate of the relay-destination channel $\gamma_{c,d}$ obtained at the former rather than at the latter terminal. However, even if the fading coefficient for the $\boldsymbol{C}$-$\boldsymbol{D}$ link is symmetric, the two overall SINRs may be significantly different, as the relay node is likely to experience a lower interference

[4] Note that condition ii) may occur since header and payload are typically independently encoded and have a separate CRC.

level, since it is typically more protected by the carrier sense region induced by the source's transmission. In order to cope with this problem, DHARQ selects the cooperative bitrate in a conservative fashion, assuming that $L(1+\varepsilon)$ information bits have to be decoded at $\boldsymbol{D}$. Thus, node $\boldsymbol{C}$ computes ρ^c_{rel} from (5.9) so as to satisfy the condition $\mathcal{L} \geq L(1+\varepsilon)$, obtaining:

$$\rho^c_{rel} = \frac{\rho_{data} \cdot \mathcal{C}(\gamma_{c,d})}{(1+\varepsilon) \cdot \rho_{data} - \mathcal{C}(\tilde{\gamma}_{s,d})} \,. \tag{5.10}$$

Once the maximum sustainable rate has been determined, the terminal compares it to the one of the direct $\boldsymbol{S}$-$\boldsymbol{D}$ link. If $\rho^c_{rel} \leq \rho_{data}$, the node gives up the procedure and goes back to its activity, since its contribution would last longer than a retransmission attempt made by the source, strongly reducing the advantages of cooperation. On the other hand, if $\rho^c_{rel} > \rho_{data}$, the terminal enters a distributed contention to act as the relay by starting a random backoff whose duration in slots is uniformly drawn in the interval $[1, \mathrm{CW}_{rel}]$. During this period, the candidate performs carrier sense. If the aggregate power level on the channel exceeds a given threshold Λ_{rel}, the node assumes that another potential cooperator has chosen a shorter backoff window, and abandons the contention. If, instead, the countdown expires with an idle medium, the terminal transmits incremental redundancy at the rate ρ^c_{rel} it has computed, and then goes back to its own activity. Let us stress that Λ_{rel} is a fundamental parameter for our cooperative approach. In fact, while a conservative threshold would limit the number of nodes that can act as cooperators, an aggressive choice may cause unexpected and harmful interference to other neighboring transmissions, reducing the spatial reuse in the network. A more in-depth discussion on these tradeoffs will be provided in Section 5.4.

If the destination successfully decodes the data, it replies with an ACK at rate ρ_{ctrl} addressed to $\boldsymbol{S}$. Otherwise, no feedback is sent. The lack of an explicit acknowledgment is interpreted by the source as a failure of the cooperative mechanism, and triggers successive attempts following the procedure described so far, until the packet is successfully received or the SRL is reached.

Let us now make some remarks on the described protocol. DHARQ extends in a simple way the plain IEEE 802.11 DCF so as to take advantage of coded cooperation. From this viewpoint, the relay selection procedure that we propose has been designed trying to adhere to the carrier sense paradigm as well as to the completely distributed nature of non-infrastructured ad hoc networks. Also, the presented medium access scheme triggers cooperation by sending a NACK only when it is actually needed, i.e., when the destination has failed to decode the data packet but has been able to cache a corrupted version of the payload. On the one hand, this approach introduces additional overhead. Nevertheless,

the transmission of NACK frames prevents cooperative transmissions that could not be exploited, e.g., because $\boldsymbol{D}$ cannot perform packet combining, and that would only result in additional and harmful interference. Moreover, the provided feedback makes it possible to perform rate adaptation and to identify the relay node among terminals that are not involved in other communications and that are experiencing channel conditions with $\boldsymbol{D}$ that satisfy a minimum quality level, i.e., good enough to decode the NACK.

In this context, it is also worth noticing that the proposed solution does not apply any criterion for identifying cooperating nodes, but rather randomly chooses one terminal from a set of available candidates. Relay selection procedures have been widely studied in the literature, both from a theoretical perspective, finding optimal solutions [73], and from a more practical angle, identifying suboptimal or heuristics strategies capable of working well in realistic networking scenarios [74]. However, we have chosen not to take advantage of such approaches in DHARQ in order to keep its implementation as simple as possible and to highlight the impact of carrier sense on the efficiency of the plain cooperative paradigm. Moreover, the effects of different relay selection strategies will be discussed in Section 5.4 by considering an idealized scheme that bounds the performance of the proposed protocol always resorting to the best possible candidate, i.e., the node that minimizes the overall transmission time.

Incidentally, we also remark that, for the sake of mathematical tractability, the framework developed in Section 5.2 did not consider NACK packets, only asking nodes to retrieve the payload from the source and to sense the medium idle in order to act as cooperators. This simplification does not significantly affect the insights presented on the cooperators distribution, as frames sent at rate ρ_{ctrl} are decoded with high probability in the whole surroundings of $\boldsymbol{S}$ and $\boldsymbol{D}$. Therefore, the conclusions inferred on the impact of carrier sense-based medium access on cooperation also apply to the DHARQ protocol described in this section.

Finally, observe that the contention scheme implemented by DHARQ resembles the one discussed for CCSMA and Phoenix in Chapter 3. Nonetheless, the two approaches are significantly different. Indeed, the schemes considered in the previous part of the thesis tried to overcome some issues of the well known decode-and-forward paradigm, whereas DHARQ implements the distinct coded cooperation approach and has been designed with the specific aim of maximizing the relaying gains in a CSMA environment starting from the insightful discussions of Section 5.2, which pointed out an unavoidable bias in the distribution of helpers due to the medium access control.

5.4 Simulation Results for Reactive Cooperation

In order to test the effectiveness of reactive cooperation in large scale CSMA-based environments, extensive Omnet++ [61] simulations have been performed. In our study, we have considered networks composed by 35 nodes spread over a $300\,m \times 300\,m$ area. Single hop flows have been analyzed, with each terminal generating Poisson traffic addressed to its neighbors with intensity λ (kbit/s/node). Such a configuration provides enough spatial separation to allow the presence of simultaneous communications within the network and is thus representative to test protocols under harsh channel contention, interference and hidden terminal conditions. The wireless environment is subject to correlated Rayleigh fading, with Doppler frequency of 40 Hz, corresponding to a 5 m/s user speed at 2.4 GHz. The information bitrate ρ_{data} for the transmission of data packets has been chosen to guarantee a success probability of 0.99 at a distance of $60\,m$ with 4 successive independent attempts in the absence of interference, and has been set to 2.1 Mb/s. In turn, the information bitrate ρ_{ctrl} for headers and control packets has been set to 0.53 Mb/s, so as to achieve a decoding probability of 0.95 with a single transmission.

In our simulations, we have analyzed the performance of DHARQ and of a plain CSMA strategy. In order to have a fair comparison between these schemes, we have tuned the protocol parameters so as to achieve a similar reliability. In particular, a Packet Delivery Ratio of approximately 95% has been obtained by setting the SRL to 5 for CSMA, whereas 4 successive attempts were enough for DHARQ. The standard set of network and protocol parameters used in our studies are reported in Tab. 5.1. All the results that will be presented have been obtained averaging the outcome of 50 independent simulations, so that the 95% confidence intervals, although not shown for readability, never exceed 3% of the estimated value.

As a starting point for our studies, we have tested CSMA and DHARQ in networks composed by a single source node surrounded by a few neighboring terminals. This setting reproduces the average topology for a link in the reference network described earlier, yet it avoids the harmful effects induced by concurrent transmissions. The outcome of these investigations, not reported here due to space constraints, supported the trends predicted by many theoretical studies, e.g., [71], with distributed hybrid ARQ exhibiting a throughput that more than doubled the one achieved by plain CSMA.

However, when tested in more articulated and realistic scenarios reactive cooperation has not been able to reproduce similar gains. This is apparent in Fig. 5.6, where the aggregate throughput against the nominal load λ in the considered 35-node topologies is shown.

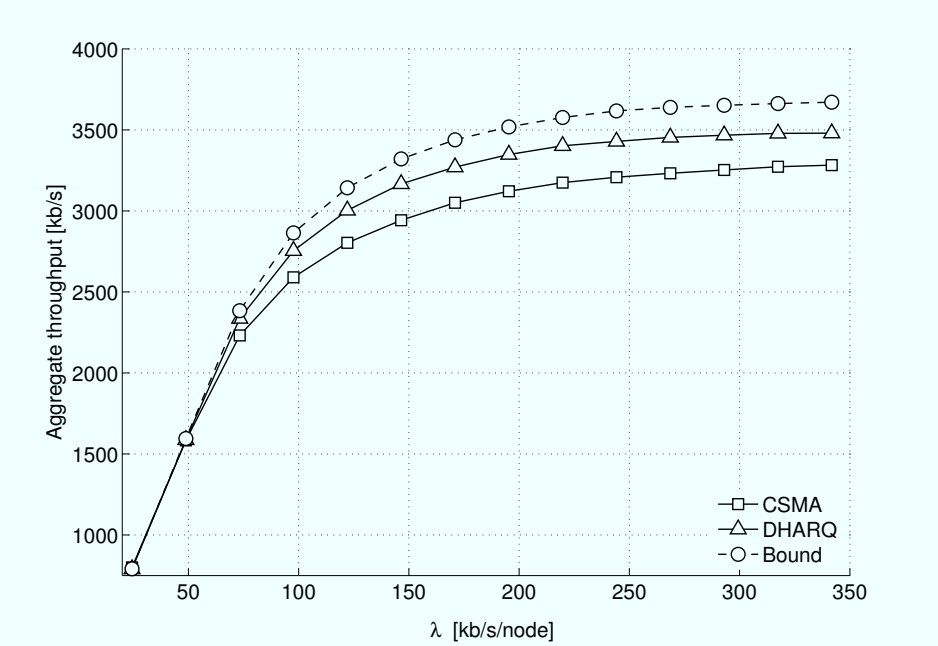


Figure 5.6. *Aggregate throughput vs nominal load expressed in kb/s/node.*

In this case, DHARQ outperforms basic CSMA by less than 10% at saturation. This observation is further corroborated by the dashed curve reported in the plot, which describes the behavior of a bound for the class of protocols that implement the cooperative paradigm under analysis. In particular, this scheme always selects the best relay among the available candidates, i.e., the node that experiences the most favorable channel conditions to the destination, and it resorts to an idealized selection process, so that all the potential candidates are immediately informed at no cost in terms of overhead of how the hybrid ARQ phase will be organized, i.e., no contention is required and collisions among relays are avoided. Fig. 5.6 highlights that even in these ideal conditions the throughput improvement offered over CSMA never exceeds 15%. Such results, then, suggest that the potential of relaying strategies is severely hindered by some effects that arise in large carrier sense-based ad hoc networks.

In order to further investigate this aspect, let us consider the outcome of a single transmission. As discussed in Section 5.3, when a source sends a packet, three events can occur at the addressee: i) the transmission succeeds; ii) the destination is not able to decode the header of the packet; or iii) the header is decoded but the payload is corrupted. Cooperation can take place only if condition iii) is met, since the destination needs to cache a corrupted version of the incoming data to take advantage of relayed retransmissions. Fig. 5.7 shows the distribution of these events in our simulations. For low traffic, the success rate is rather

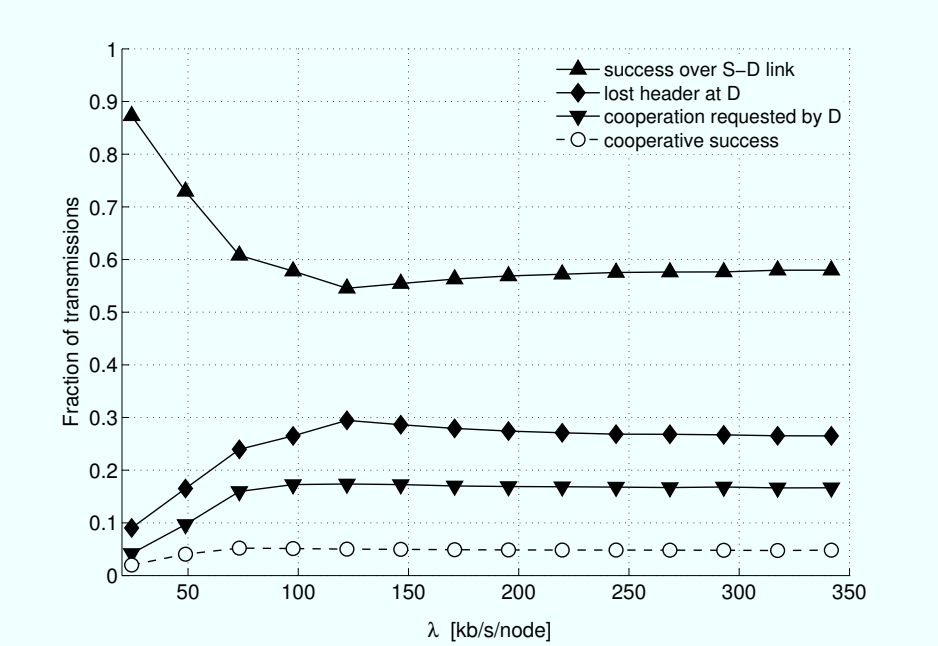


Figure 5.7. *Outcome of a single transmission vs nominal load expressed in kb/s/node. At the destination, three conditions can occur: i) the transmission succeeds (*success over $\boldsymbol{S}$-$\boldsymbol{D}$ link*); ii) the destination is not able to decode the header of the packet (*lost header*); and iii) the header is decoded but the payload is corrupted (*cooperation requested*).*

high, and thus cooperation is seldom of use. Conversely, as λ is raised, the probability of decoding the payload with a direct transmission decreases, due to the higher interference level. Nonetheless, the worse reception quality does not lead to a significant increase of the number of cooperative phases that may be expected. In fact, at saturation, transmission failures are mostly induced by the loss of the packet header, almost 30% of the communications, whereas relaying processes are curbed to less than 20% of the cases. In this perspective, it is important to observe that the incapability of retrieving packet headers does not primarily stem from channel impairments, as control frames are sent at a conservative rate ρ_{ctrl}, but rather from the completely distributed nature of the medium access policy under consideration. With CSMA, indeed, a destination may not perceive packets addressed to it since it is already synchronized to other ongoing communications, e.g., because of the hidden terminal problem. Let us remark that while this effect could not be taken into account in the simple scenarios analyzed in our theoretical studies, it plays a key role in larger and more congested networks.

The reduced fraction of times in which a destination may successfully support relaying clearly represents a major limitation to the performance of DHARQ. However, Fig. 5.7 also

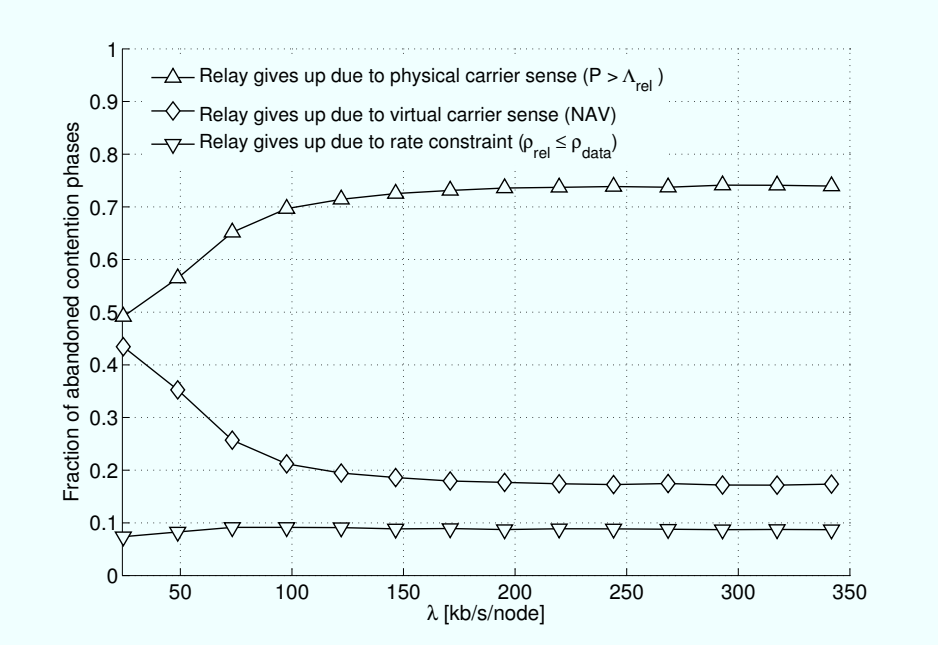


Figure 5.8. *Impact of the reasons that may lead a relay candidate to abandon a contention vs nominal load expressed in kb/s/node.*

highlights that, even when requested, cooperation is successful only in one-fourth of the cases (dashed line). From this viewpoint, when a NACK is sent, the distributed hybrid ARQ mechanism that we propose may fail because of three reasons: i) no node has been able to decode the packet transmitted by the source; ii) some candidates are present, but all of them give up the contention; or iii) at least one terminal acts as relay, but the destination is still not able to decode the payload, e.g., because of a collision. Our simulations have shown that the dominant motive is by far factor ii), as nodes that abandon the election procedure are responsible for more than 60% of the non-performed cooperative phases. In turn, as reported in Fig. 5.8 against the nominal load, terminals refrain from relaying most likely because they perceive a power level on the medium which is above the carrier sense threshold Λ_{rel} (80% of the cases), whereas a much smaller number of withdrawals are due to the rate constraint described in Section 5.3, or to virtual carrier-sense.[5] From this discussion we can once more infer that proactive cooperation is hampered by the intrinsic nature of carrier sensing, which does not grant channel access to nodes that would be available for relaying.

Finally, it is important to observe that not only does CSMA prevent a large share of co-

[5]Packet headers contain information on the duration of the communication they initiate. Therefore, nodes that decode a header not addressed to them become aware that the channel will be busy for a given time, and suitably update their Network Allocation Vector (NAV) [5], deferring any channel access for the reserved period.

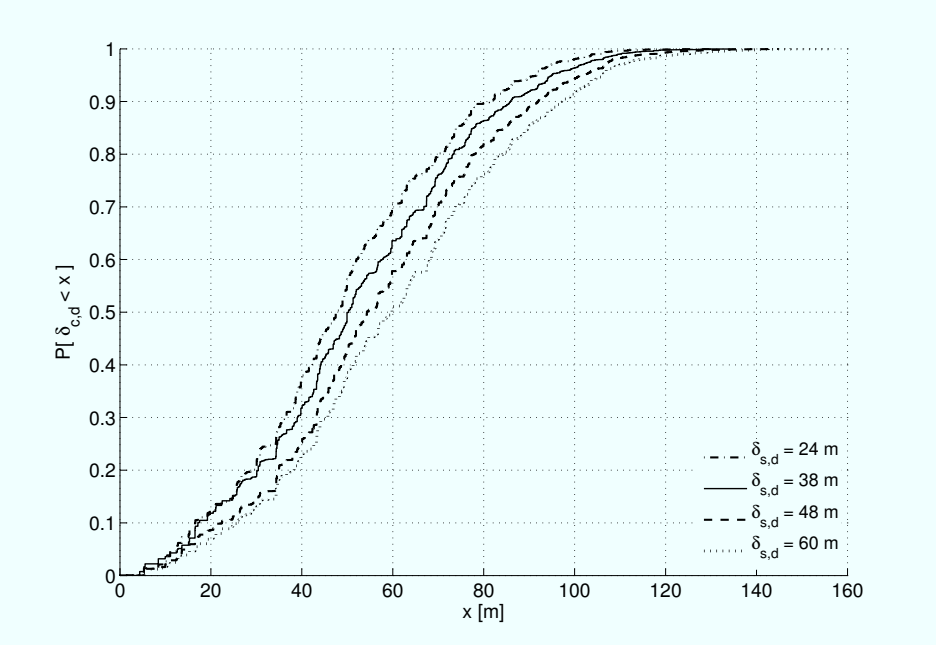


Figure 5.9. *Cumulative Distribution Function of the distance of relay candidates from the destination in reactive cooperative schemes. Different curves represent different source-destination distances.*

operative phases, but also it reduces their effectiveness when they actually take place. As discussed in Section 5.3, if an $\boldsymbol{S}$-$\boldsymbol{D}$ communication fails, interferers are most likely located toward $\boldsymbol{D}$ (see Fig. 5.2), since the area surrounding the source is protected by the carrier sense mechanism. This, in turn, tends to concentrate nodes that decode the payload, i.e., relay candidates, in the proximity of $\boldsymbol{S}$. Thus, cooperators are likely to offer a limited (if any) advancement towards the destination. This has a negative effect on hybrid ARQ, since a higher path-loss induces longer retransmissions and therefore reduces the advantages of quick failure recovery offered by cooperation. We have delved into this aspect by performing a specific set of simulations. In particular, we considered topologies in which a source and a destination were located at the center of the network at distance $\delta_{s,d}$, whereas the rest of the terminals were randomly distributed in the $300\,m \times 300\,m$ area. Fig. 5.9 depicts the cumulative distribution function (CDF) of the distance from $\boldsymbol{D}$ of relay candidates, i.e., of nodes that decode the source's payload given that the reception at the destination fails, for values of $\delta_{s,d}$ ranging from 25 to 60 m. As a first remark, the plot confirms the bias in the positions of potential cooperators: on average, nodes that offer an advancement towards the destination with respect to the source are available only in 30% of the cases. Moreover, we notice that the smaller $\delta_{s,d}$, the higher the probability of finding relays that are in the proximity of the destination. This stems from the fact that, the closer $\boldsymbol{S}$ to $\boldsymbol{D}$, the more

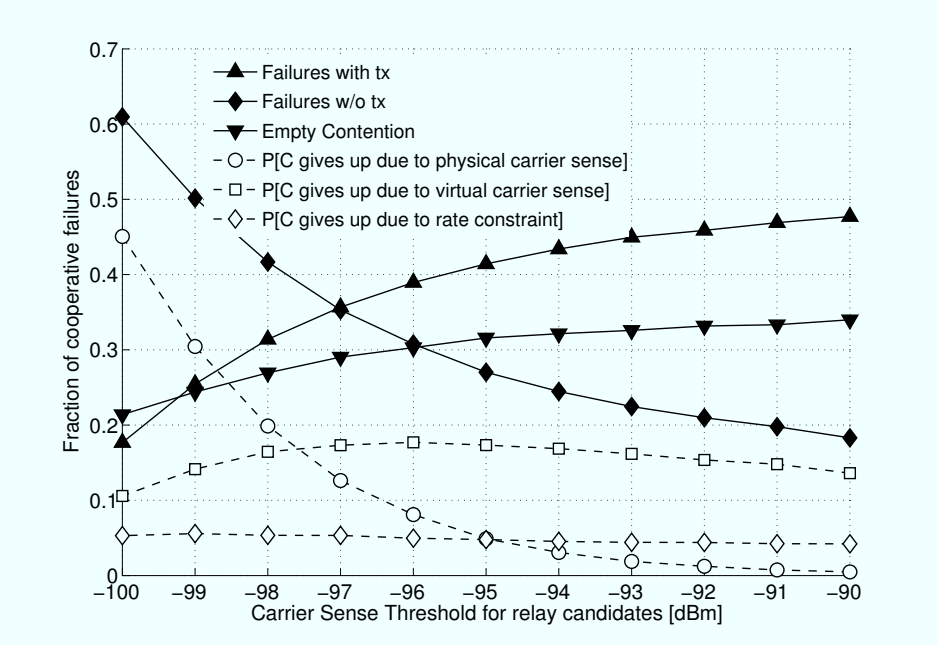


Figure 5.10. *Efficiency of reactive cooperative phases for different values of the carrier sense threshold for relay contentions.*

the area that carrier sense reserves for the ongoing communication extends over the destination's surroundings, making it possible for terminals in this region to participate in the relay contention. On the other hand, this favorable effect is not able to trigger significant improvements, since it facilitates relaying exactly when cooperation tends to be triggered seldom, due to the more favorable average channel conditions for links with reduced values of $\delta_{s,d}$.

The results discussed so far have been obtained by setting the carrier sense threshold used during relay contention equal to the one used for basic channel access, i.e., $\Lambda_{rel} = \Lambda$. In order to increase the number of cooperative phases, one could think of a more aggressive approach, letting terminals act as cooperators even though they detect some activity on the channel. In our analysis, we have considered the effectiveness of such a strategy by testing DHARQ with different values of Λ_{rel}. Let us first consider the results reported in Fig. 5.10. Here, the solid lines investigate why a relaying request (i.e., a NACK packet) is not followed by a cooperative success. As discussed earlier, such a condition may happen either because no node has decoded the payload sent by the source (*empty contention*), or because all the candidates give up the contention (*failure w/o tx*), or because at least one relay actually transmits but the destination is not able to decode the payload, likely due to a collision (*failure with tx*). As Λ_{rel} is increased, nodes that have cached the packet sent

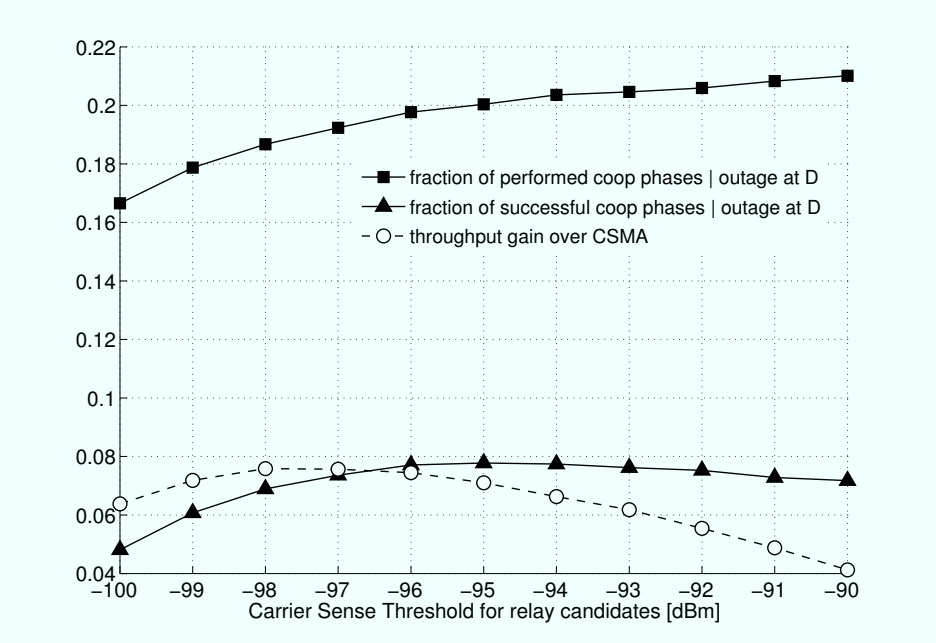


Figure 5.11. *Impact of reactive cooperative phases and aggregate throughput gain for DHARQ with respect to CSMA for different values of the carrier sense threshold for relay contentions.*

by the source tend to become more aggressive, and more cooperative phases take place, as shown by the fact that the impact of failures without transmissions significantly decreases, passing from 60% to as low as 20%. This behavior is further confirmed by the dashed lines in the plot, which report the reasons that lead nodes to refrain from cooperating and sum up to the *failure w/o tx* curve: a higher carrier sense threshold prevents relay candidates from withdrawing from the contention. However, this approach has two main consequences. First of all, the number of collisions among relays increases, since candidate nodes may not overhear each other's attempt and may access the channel simultaneously. We stress that while on the one hand concurrent relay transmissions are likely to significantly worsen the effectiveness of cooperation, on the other hand they also represent a waste of network resources in terms of energy and bandwidth. Moreover, a higher number of relayed packets tends to raise the overall interference, and thus to lower the average decoding probability in the system. Not only does this affect reception at destination nodes, but also it reduces the number of potential relays that are able to successfully retrieve the packet sent by the source. Fig. 5.10 confirms this observation, showing how the number of times in which no terminal has been able to decode the payload, i.e., empty contentions, increases with Λ_{rel}.

The aforementioned tradeoffs reflect at a system level, as reported in Fig. 5.11. When Λ_{rel} is raised, more relays tend to actually perform retransmissions (square-marked curve). This

initially leads to an increases the fraction of successful cooperative phases as well (triangle-marked curve), which pass from the 5% that characterized the reference implementation discussed in Fig. 5.7 to up to 8% of the overall communications. However, as soon as the carrier sense threshold goes beyond a certain limit (around -97 dBm in our settings), the drawbacks induced by collisions and by the higher level of interference prevail, worsening the overall efficiency of the cooperative mechanism. As a consequence, a similar trend is followed by the aggregate throughput for DHARQ, whose gain over plain CSMA plummets when relays become too aggressive. These results suggest that the optimal working point for the system can be achieved for a value of Λ_{rel} slightly higher than the one used for basic channel access. Nevertheless, the achievable improvements are limited, as the gains over the benchmark protocol never exceed 10%.

In conclusion, the simulation studies reported in this paper have thoroughly highlighted a detrimental effect of CSMA on reactive cooperation, induced both by the bias in the spatial distribution of relay nodes discussed in Section 5.2 and by other simple yet unavoidable networking mechanisms that would affect any implementation of a real-word cooperative scheme in a carrier sense-based environment.

5.5 Cooperation in the Availability of Channel State Information

In the first part of this chapter we have considered wireless communication schemes characterized by a source node transmitting at a fixed information bitrate and by a relay terminal offering support by sending redundancy to the destination in the event of a failure over the direct link. This paradigm does not require any prior knowledge of the channels connecting the terminals involved in the data transfer, and therefore significantly eases the design and the implementation of cooperative protocols in the non-infrastructured and completely distributed scenarios that are the focus of our study. Recently, however, several works have studied from a theoretical perspective a different class of relaying solutions that take advantage of Channel State Information (CSI) to tune transmission rates, reporting significant gains over plain ARQ in ad hoc networks [38, 75, 76]. In particular, with reference to the usual $\boldsymbol{S}$ (source), $\boldsymbol{D}$ (destination), $\boldsymbol{C}$ (cooperator) topology, let us assume that some knowledge on the SINRs that would characterize data exchanges over the $\boldsymbol{S}$-$\boldsymbol{D}$, $\boldsymbol{S}$-$\boldsymbol{C}$ and $\boldsymbol{C}$-$\boldsymbol{D}$ channels is available. Under these hypotheses, the terminal that has data to send may decide whether to transfer its payload directly to the addressee or to split the communication into two parts, delivering first the information to the relay and then letting such terminal send redundancy to the destination, so as to minimize the overall transmission

time while providing the desired reliability by performing rate adaptation. This approach determines in advance the need for cooperative support, and thus it sets up a *proactive relaying* paradigm, as opposed to the *reactive* policy of DHARQ (see Section 5.3). Also, it has the merit of exploiting CSI to optimize the usage of channel capacity for cooperative links.

In light of this discussion, and with a view to get a more comprehensive understanding of the interactions between carrier sense-based medium access and cooperation, we devote the second part of our study to the investigation of these type of strategies.

From this viewpoint, a first and key remark is that the effectiveness of a protocol implementing proactive cooperation is tightly correlated to the accuracy of the channel knowledge available at the nodes. It is clear that in a realistic scenario, and in particular with a fully decentralized framework, the retrieval of up-to-date information on channel qualities is not feasible due to the large amount of overhead that would be required, especially when the source has to choose among several relay candidates. On the other hand, a practical and interesting solution that tries to overcome these issues has been provided in [38], where a carrier sense-based infrastructured wireless ad hoc network is considered. The authors investigate a scenario where all nodes contend for a shared channel to deliver data to a common access point. In the proposed protocol, each source has a predefined cooperator, that is identified exploiting channel knowledge gathered from past transmissions rather than triggering dedicated negotiation procedures. This strategy has been shown to enable remarkable gains in centralized environments, yet its efficiency may be severely hindered in the non-infrastructured and non-hierarchical networks with all-to-all traffic that we consider. In fact, in our scenarios each terminal has multiple possible destinations, and thus it selects a particular addressee less frequently than in the centralized case. This implies that a knowledge of the achievable rate associated to a link learnt from previous transmissions is more likely to be outdated. Moreover, the presence of multiple simultaneous communications within the network area induces rapidly varying interference conditions, further hampering solutions that rely on a fixed cooperator.

Starting from these remarks, in our study we assume that all the up-to-date channel knowledge needed for setting up a cooperative link is always available at nodes without any overhead. The merit of this approach is twofold. On the one hand, the solutions that will be presented and discussed represent a bound for the class of protocols that implement proactive cooperation in non-centralized ad hoc networks, and thus they allow to draw broadly applicable conclusions. On the other hand, neglecting the drawbacks that would stem at the link layer from procedures implemented to gather CSI better emphasizes the

impact of issues related to carrier sense-based medium access on relaying strategies.

We start our investigation by describing in Section 5.6 two protocols, referred to as CSMA-CSI and Coop-CSI, that exploit a perfect awareness of channel and interference conditions to extend plain CSMA and to realize proactive cooperation, respectively. Then, in Section 5.7 an analytical framework is developed to highlight the key effects induced by CSMA on such schemes in a simple scenario. Finally, Section 5.8 is devoted to simulations that compare the performance of the considered medium access protocols in large and more realistic networking environments.

5.6 Proactive Cooperative Protocols

5.6.1 Carrier Sense Multiple Access with CSI: CSMA-CSI

Let us consider a source $\boldsymbol{S}$ that has data to deliver to a destination $\boldsymbol{D}$. As in the plain IEEE 802.11 DCF, with CSMA-CSI the node performs the backoff mechanism described in Chapter 3.5.1.1. Once channel access has been granted, say at time t_0, $\boldsymbol{S}$ evaluates the maximum information bitrate $\rho_{s,d}$ it can use for reliably communicating with $\boldsymbol{D}$ based on its knowledge of the instantaneous SINR $\gamma_{s,d}(t_0)$. In order to cope with potential fluctuations that may affect the SINR over the transmission of the L-bit payload, e.g., due to changes in the aggregate interference perceived at the destination or due to the variability of the channel coefficient, the source follows a conservative approach, determining the sustainable bitrate as if $L(1+\varepsilon')$ information bits had to be retrieved at $\boldsymbol{D}$. Recalling (5.1), $\rho_{s,d}$ is computed as:

$$\rho_{s,d} = \frac{\mathcal{C}(\gamma_{s,d}(t_0))}{(1+\varepsilon')} \ , \tag{5.11}$$

for a transmission time of $T_{s,d} = L/\rho_{s,d}$ seconds.

Once this calculation has been performed, the source checks whether the $\boldsymbol{S}$-$\boldsymbol{D}$ channel satisfies a minimum quality constraint, so as to avoid communications over links characterized by extremely poor conditions that would result in an inefficient usage of the bandwidth. From this viewpoint, if $\rho_{s,d} < \rho_{min}$, or, equivalently, if $T_{s,d} > T_{max}$, the node refrains from transmitting and behaves as if the communication had failed, i.e., it increases the counter of the attempts performed for the current packet and re-initiates the backoff procedure. Conversely, if $\rho_{s,d} \geq \rho_{min}$ ($T_{s,d} \leq T_{max}$), $\boldsymbol{S}$ transmits the payload at the maximum sustainable information bitrate and waits for a reply from $\boldsymbol{D}$. In the event of a successful reception, the destination replies with an ACK sent at rate ρ_{ctrl} and the communication comes to an end. Otherwise, no feedback is provided, and the source iterates the described mechanism until either the payload is delivered or the SRL is reached.

5.6.2 Cooperative CSMA with CSI: Coop-CSI

With reference to the transmission of an L-bit data unit over the $\boldsymbol{S}$-$\boldsymbol{D}$ link, let $\mathcal{R}_s = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ \ldots, \ \boldsymbol{C}_n\}$ be the set of neighbors of the source node.[6] As in the non-cooperative case, $\boldsymbol{S}$ initiates its transmission process by performing channel sensing and by computing the minimum sustainable transmission time $T_{s,d}$ over the direct link with the destination. However, when Coop-CSI is used as medium access policy, the source also checks whether it is possible to reliably deliver the payload in a shorter time by relying on the cooperating help of a surrounding terminal, splitting the transmission in two successive phases. Following this approach, the first part of the communication is devoted to a reliable data transfer from $\boldsymbol{S}$ to one of its neighbors. The destination, in turn, caches such a packet even if the used bitrate was too high for it to successfully retrieve the information content. Conversely, the second phase of the transmission is reserved to the relay node, which sends a different part of the original codeword obtained re-encoding the data unit sent by $\boldsymbol{S}$, and providing the redundancy needed for payload decoding at $\boldsymbol{D}$ according to the distributed hybrid ARQ rationale discussed in Section 5.3.

From this perspective, a node $\boldsymbol{C}_i \in \mathcal{R}_s$ is considered as a candidate for relaying if two conditions are met at the time at which the source wants to access the channel: i) $\boldsymbol{C}_i$ is available for communication, i.e., it is not involved in any other ongoing link, and ii) $\boldsymbol{C}_i$ senses the medium idle, i.e., it is allowed to transmit. Let now $\mathcal{R}'_s \subseteq \mathcal{R}_s$ be the set of terminals that satisfy these requirements. In order to detect whether a two-hop link that offers improvement over the direct transmission is available, the source starts by computing for each $\boldsymbol{C}_i \in \mathcal{R}'_s$ the maximum sustainable information bitrate for the first phase of a cooperative communication, based on the current value of the SINR γ_{s,c_i}:

$$\rho_{s,c_i} = \mathcal{C}(\gamma_{s,c_i})/(1+\varepsilon') \ . \tag{5.12}$$

The corresponding transmission time is $T_{s,c_i} = L/\rho_{s,c_i}$ and, according to (5.1), the destination may exploit the $\boldsymbol{S}$-$\boldsymbol{C}_i$ data exchange to retrieve $\mathcal{L}_{1,i} = T_{s,c_i} \cdot \mathcal{C}(\gamma_{s,d})$ information bits. Therefore, the overall duration of the two-hop solution with $\boldsymbol{C}_i$ as cooperator can be computed as:

$$T^i_{split,i} = T_{s,c_i} + \frac{L - \mathcal{L}_{1,i}}{\rho_{c_i,d}} \ , \tag{5.13}$$

where the second term represents the time needed for the $\boldsymbol{C}_i$-$\boldsymbol{D}$ transmission to allow reli-

[6] By neighbors we generally mean terminals that may support a node in the cooperative process. As an example, the set $\mathcal{R}_s$ may be determined at $\boldsymbol{S}$ by keeping track, possibly in a dynamic fashion, of the nodes it can reliably communicate with.

able decoding at the destination, and where $\rho_{c_i,d} = \mathcal{C}(\gamma_{c_i,d})/(1+\varepsilon')$ is the maximum sustainable bitrate during the second phase of the communication.

Upon completing these calculations, the source decides in which way to deliver its payload evaluating

$$T^* = \min_{\boldsymbol{C}_i \in \mathcal{R}'_s} \{T_{split,i}, T_{s,d}, T_{max}\} \ . \tag{5.14}$$

In particular, if $T^* = T_{max}$, no solution providing sufficient reliability is available. In this case, the node refrains from transmitting, increases the counter for the number of attempts performed for the current packet and re-initiates the backoff procedure. Conversely, if $T^* = T_{s,d}$, $\boldsymbol{S}$ simply sends the data unit over the direct link at rate $\rho_{s,d}$, and the communication follows the principles of CSMA-CSI. Finally, when $T^* = T_{split,j}$, with $j = \text{argmin}_i\{T_{split,i}\}$, the source starts the data exchange by sending the payload to $\boldsymbol{C}_j$ at rate ρ_{s,c_j} computed through (5.12). Notice that the header of such a packet contains information describing the structure of the transmission, i.e., source, destination, chosen relay and ID of the data unit. In this way, if the final addressee decodes the header, a corrupted version of the incoming data can properly be cached, and the terminal becomes aware that additional redundancy will be sent over a cooperative link. Therefore, as soon as $\boldsymbol{C}_j$ retrieves the information sent by $\boldsymbol{S}$, the payload is re-encoded so as to form a different part of the original codeword, and the relay immediately delivers it to $\boldsymbol{D}$ with a transmission at rate $\rho_{c_j,d}$. If the destination decodes the data unit, an acknowledgment is sent, and the communication comes to an end. Otherwise, the source iterates the described procedure until either successful delivery is achieved or the SRL is reached.

In conclusion, let us remark that besides relying on perfect channel and interference knowledge as its non-collaborative counterpart, Coop-CSI also idealizes some aspects of medium access contention, as we assume all the terminals involved in a data exchange to be immediately and seamlessly aware of how the communication will be structured. On the one hand such an assumption falls under the line of reasoning, discussed at the beginning of this section, of dealing with schemes that represent a bound for the class of protocols that implement proactive cooperation, so as to draw conclusions that are both broadly applicable and stem from the intrinsic interactions between carrier sensing and cooperation rather than by specific implementation details. On the other hand, we also notice that the proposed protocol does not embody the proactive relaying paradigm at too abstract a level, as Coop-CSI could easily be modified in order to be practically implemented, e.g., by following the approaches presented in [38, 39].

5.7 An Analytical Framework for Proactive Cooperation

Let us start our analysis of schemes that implement proactive cooperation by investigating the performance gain they can offer over basic solutions when nodes have to obey no medium access control policy, i.e., assuming perfect coordination among terminals. To this aim, we focus on the protocols described in Section 5.6 disregarding the constraints of CSMA. In particular, we consider for the moment a topology composed by a source $\boldsymbol{S}$ and a destination $\boldsymbol{D}$ displaced in a region $\mathcal{A}$ at positions $p_s = \{x_s, y_s\}$ and $p_d = \{x_d, y_d\}$ respectively, and by a relay terminal $\boldsymbol{C}$ deployed in the middle of the line connecting them.[7] For the sake of mathematical tractability, we assume fading coefficients to remain constant for the whole duration of a communication, and, following the traditional approach used in theoretical works, we do not model a source of interference subject to carrier sense, but we rather assume an interference level with average power σ_i^2 at both $\boldsymbol{D}$ and $\boldsymbol{C}$.

The basic transmission policy that we analyze, as discussed, always uses the direct link to deliver data from the source to the destination. The maximum sustainable rate in this case is given by $\rho_{s,d} = \mathcal{C}(\gamma_{s,d})$, with $\gamma_{s,d} = \eta_{s,d}/(N + \iota_d)$.[8] Here, both the desired received power $\eta_{s,d}$ and the interfering power ι_d at the destination are modeled as exponential random variables, with mean values $P\delta_{s,d}^{-\alpha}$ and σ_i^2, respectively. The throughput τ_{direct} offered by this approach can be computed by averaging the achievable information bitrate over fading and interference, obtaining:

$$\tau_{direct} = \frac{P\delta_{s,d}^{-\alpha}}{P\delta_{s,d}^{-\alpha} - \sigma_i^2} \cdot \left[\mathcal{G}\left(-\frac{N}{P\delta_{s,d}^{-\alpha}} \right) - \mathcal{G}\left(-\frac{N}{\sigma_i^2} \right) \right] , \tag{5.15}$$

where N is the noise floor, and $\mathcal{G}(a)$ is defined as:

$$\mathcal{G}(a) = \frac{B}{\ln(2)} \cdot e^{-a} \int_{-a}^{+\infty} \frac{e^{-t}}{t} dt \, . \tag{5.16}$$

On the other hand, the cooperative protocol chooses whether to employ a direct or a two-hop link so as to minimize the overall transmission time. In this case, the sustainable bitrates involving the relay terminal are given by $\rho_{s,c} = \mathcal{C}(\gamma_{s,c})$ and $\rho_{c,d} = \mathcal{C}(\gamma_{c,d})$, and the channel capacities are determined by the SINRs $\gamma_{s,c} = \eta_{s,c}/(N + \iota_c)$ and $\gamma_{c,d} = \eta_{c,d}/(N + \iota_d)$, where $\eta_{s,c}$, $\eta_{c,d}$ and ι_c are again modeled as exponential variables with mean values $P\delta_{s,c}^{-\alpha}$, $P\delta_{c,d}^{-\alpha}$ and σ_i^2, respectively. Applying (5.13), the time required for data delivery over a cooperative

[7] As will be discussed later, this configuration maximizes the gains offered by proactive cooperation.

[8] Notice that the margin factor ε' discussed in Section 5.6 is not considered here, since fading and interference are assumed to remain constant for the whole duration of a transmission.

link can easily be computed as:

$$T_{split} = L \cdot \frac{\mathcal{C}(\gamma_{s,c}) + \mathcal{C}(\gamma_{c,d}) - \mathcal{C}(\gamma_{s,d})}{\mathcal{C}(\gamma_{s,c}) \cdot \mathcal{C}(\gamma_{c,d})} \ . \tag{5.17}$$

Starting from this result, and introducing the random vector $\mathbf{v} = \{\eta_{s,d}, \eta_{s,c}, \eta_{c,d}, \iota_c, \iota_d\}$, two regions can be identified: $\Delta_{split}(\mathbf{v}) = \{\mathbf{v} \mid \mathcal{C}(\gamma_{s,c}) \geq \mathcal{C}(\gamma_{s,d}), \mathcal{C}(\gamma_{c,d}) \geq \mathcal{C}(\gamma_{s,d})\}$ and $\Delta_{direct}(\mathbf{v}) = \Delta^c_{split}(\mathbf{v})$,[9] so that for $\mathbf{v} \in \Delta_{split}$ a relayed communication is performed, whereas for $\mathbf{v} \in \Delta_{direct}$ the protocol resorts to the direct link. It follows that the average throughput τ_{coop} for the cooperative solution can be computed as:

$$\begin{aligned} \tau_{coop} &= \int_{\Delta_{split}} \frac{\mathcal{C}(\gamma_{s,c}) \cdot \mathcal{C}(\gamma_{c,d})}{\mathcal{C}(\gamma_{s,c}) + \mathcal{C}(\gamma_{c,d}) - \mathcal{C}(\gamma_{s,d})} \ f_{\mathbf{v}}(\mathbf{v}) d\mathbf{v} + \\ &+ \int_{\Delta_{direct}} \mathcal{C}(\gamma_{s,d}) \ f_{\mathbf{v}}(\mathbf{v}) d\mathbf{v} \, , \end{aligned} \tag{5.18}$$

where $f_{\mathbf{v}}$ is the joint probability density function of vector $\mathbf{v}$, factorizable as the product of the densities of the involved exponential variables,[10] and where the integrand over Δ_{split} has been obtained from (5.17).

Fig. 5.12 reports the ratio of τ_{coop} to τ_{direct} with network and protocol parameters set as in Tab. 5.1. Here, the x axis represents the average interference perceived at $\boldsymbol{C}$ and $\boldsymbol{D}$, i.e., σ_i^2, while the y axis indicates the distance $\delta_{s,d}$ between source and destination. As a general remark, the plot clearly shows that proactive cooperation has the potential to enable important improvements over simpler communication schemes in small topologies and when medium access is perfectly coordinated among nodes. Under these assumptions, indeed, the average throughput offered by relaying strategies is more than twice the one achieved by basic solutions for a wide set of networking parameters. Moreover, it is possible to observe how, for a given distance between source and destination, relaying is exploited at its utmost for intermediate values of σ_i^2. The reason is that for very low interference the direct link already offers extremely good performance and thus there is little room for improvement. On the other hand, when σ_i^2 raises above a critical value not only does the capacity offered by the direct link deteriorate, but also the quality of the $\boldsymbol{S}$-$\boldsymbol{C}$ and $\boldsymbol{C}$-$\boldsymbol{D}$ channels plummets, dilating the time required for a two-hop communication more than proportionally and consequently shrinking the achievable throughput gain. Finally, Fig. 5.12 also suggests that, when a fixed level of interference is considered, cooperation performs better for larger values of $\delta_{s,d}$, as higher path-losses affecting the direct link favor data transfers over two faster hops.

[9] With $\mathcal{B}^c$ we indicate the complementary set of $\mathcal{B}$.

[10] Such a decomposition holds since all the components of $\boldsymbol{v}$ involve spatially disjoint, and therefore statistically independent, channels.

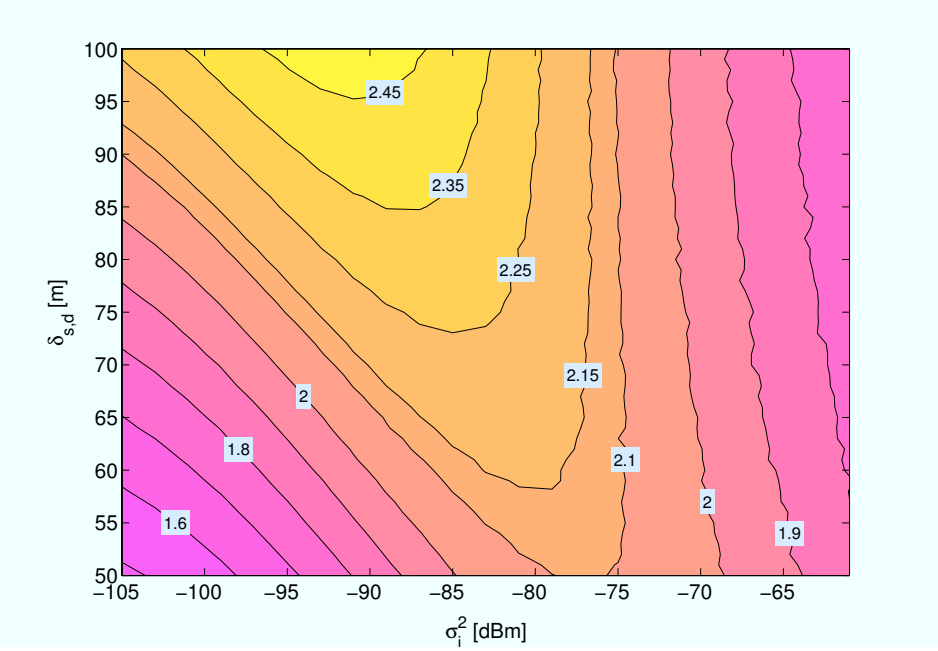


Figure 5.12. *Average throughput gain of proactive cooperative over non-cooperative solutions for different values of $\delta_{s,d}$ and for different interference levels when no medium contention and uniform interference are considered. The cooperator is assumed to be located in the middle of the line connecting $\boldsymbol{S}$ and $\boldsymbol{D}$.*

We now broaden our investigation by focusing on how carrier sense-based medium access can impact proactive relaying strategies. To this aim, we refer again to a scenario composed by a source $\boldsymbol{S}$, a destination $\boldsymbol{D}$, an interferer $\boldsymbol{I}$ and a potential cooperator $\boldsymbol{C}$ placed within a region $\mathcal{A}$ at positions $p_s = \{x_s, y_s\}$, $p_d = \{x_d, y_d\}$, $p_i = \{x_i, y_i\}$ and $p_c = \{x_c, y_c\}$, respectively. Furthermore, for the sake of mathematical tractability, we assume that $\boldsymbol{I}$ accesses the channel simultaneously with $\boldsymbol{S}$ if not locked by carrier sense, and that its transmission lasts for the whole duration of the communication between the source and the destination.

A first important observation regards the bias induced by CSMA on the position of the interfering node. Recalling the system model presented in Section 5.1, the probability that a terminal located at p_i is allowed to access the channel given a transmission being performed by $\boldsymbol{S}$ is given by $\mathcal{F}(p_s, p_i)$, reported in (5.2).

$\mathcal{F}(p_s, p_i)$ exhibits a radial symmetry centered in p_s, with an exponentially decreasing probability of sensing the medium as idle for larger values of $\delta_{s,i}$. It follows that, as expected, $\boldsymbol{I}$ tends to be concentrated far away from the source. In the perspective of our discussion it is important to notice that, while this distribution is meant to protect the direct link so as to enhance its decoding probability, it also severely hinders the effectiveness of cooperative mechanisms. In order to understand this effect, let us recall that a relayed transmission with

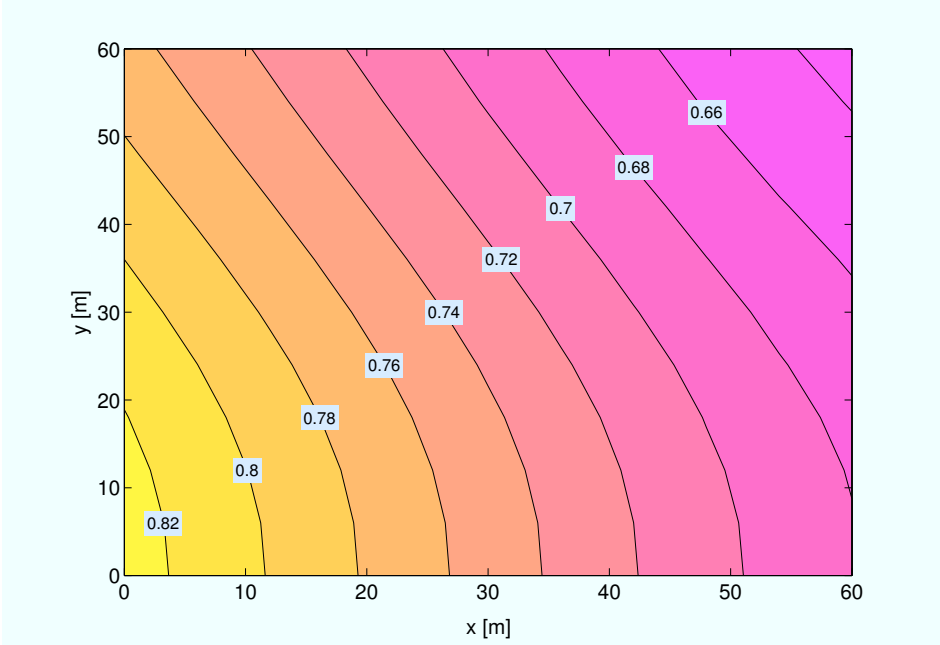


Figure 5.13. *Probability that a node in p_c senses the medium idle in a proactive cooperative scenario. $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.*

the help of $\boldsymbol{C}$ is triggered according to Coop-CSI only if two conditions are met: i) $\boldsymbol{C}$ senses the medium idle, and ii) the time required for delivering data over the $\boldsymbol{S}$-$\boldsymbol{C}$-$\boldsymbol{D}$ path is shorter than the duration achievable by means of a direct link between source and destination, i.e., $T_{split} < T_{s,d}$.

Consider first the former requirement. Conditioning on the position of the interferer, the probability that a node in p_c is allowed to access the medium can be expressed as $\mathcal{F}(p_i, p_c)$. Therefore, letting $\boldsymbol{p} = \{p_s, p_d, p_c\}$ be the topology vector under consideration, the probability $\mathcal{M}(\boldsymbol{p})$ that $\boldsymbol{C}$ meets requirement i) can be obtained by averaging the conditional value over the spatial distribution of $\boldsymbol{I}$, obtaining:

$$\mathcal{M}(\boldsymbol{p}) = \frac{\int_{\mathcal{A}} \mathcal{F}(p_i, p_c)\, \mathcal{F}(p_s, p_i)\, dp_i}{\int_{\mathcal{A}} \mathcal{F}(p_s, p_i)\, dp_i}\,. \tag{5.19}$$

The results of the numerical evaluation of (5.19) for a region $\mathcal{A}$ spanning $x \in [-150, 200]$, $y \in [-200, 200]$, a topology $p_s = \{0, 0\}$, $p_d = \{60, 0\}$ (all the coordinates being expressed in meters), and parameters set according to Tab. 5.1 are shown in Fig. 5.13.[11] The plot clearly highlights how terminals allowed to take part in a relaying phase while obeying the princi-

[11] While $\mathcal{A}$ has been chosen wide enough to also take into account the influence of interfering nodes located far from the source-destination pair, the figures shown here plot the identified distributions restricted to a part of the region for the sake of clarity.

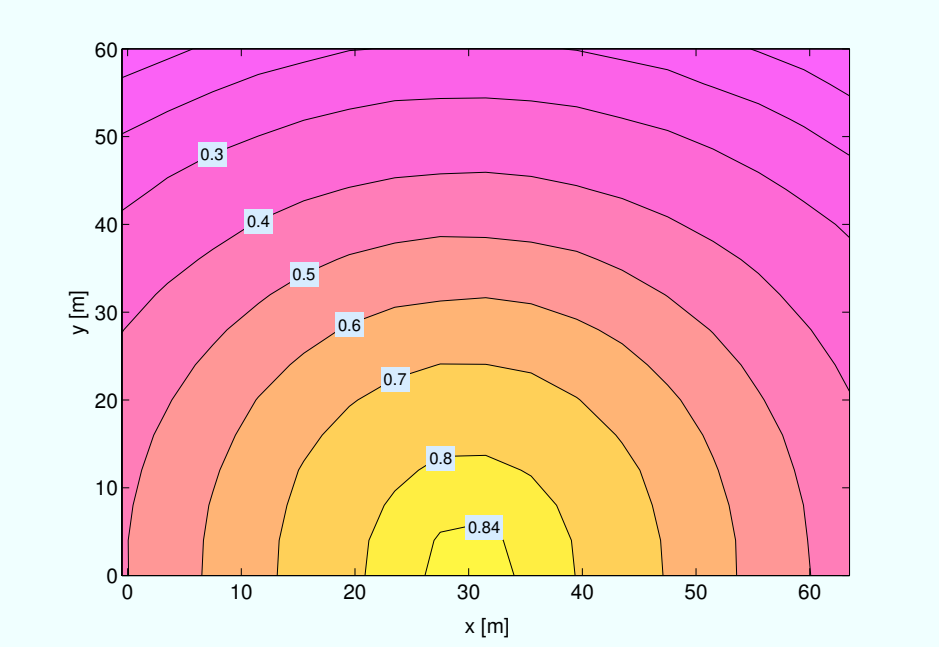


Figure 5.14. *Probability that a node in p_c satisfies the rate constraint to trigger proactive cooperation.* $p_s = \{0, 0\}$, $p_d = \{60, 0\}$.

ples of CSMA are concentrated in the proximity of the source, as a consequence of the lower interference level induced by the medium access policy in such a region.

On the other hand, focusing on requirement ii), let $\mathcal{R}(\boldsymbol{p})$ be the probability that a cooperative communication involving $\boldsymbol{C}$ offers improvement over the direct link. Recalling the approach discussed in the first part of this section for the throughput derivation, $\mathcal{R}^{p_i}(\boldsymbol{p}, p_i)$, i.e., $\mathcal{R}(\boldsymbol{p})$ conditioned on the position of the interferer, can be evaluated by simply integrating $f_{\mathbf{v}}(\mathbf{v})$ over $\Delta_{split}(\mathbf{v})$. In this case, however, the random variables ι_c and ι_d, i.e., the components of $\mathbf{v}$ that describe the interference level perceived at $\boldsymbol{C}$ and $\boldsymbol{D}$, are no longer i.i.d. with mean value σ_i^2. Instead, in order to take into account the bias caused by CSMA, we model them as independent exponential random variables with mean values $P\delta_{i,c}^{-\alpha}$ and $P\delta_{i,d}^{-\alpha}$, respectively. Under these assumptions, the sought probability can be computed by averaging $\mathcal{R}^{p_i}(\mathbf{p}, p_i)$ over p_i, obtaining:

$$\mathcal{R}(\boldsymbol{p}) = \frac{\int_{\mathcal{A}} \mathcal{R}^{p_i}(\mathbf{p}, p_i)\ \mathcal{F}(p_s, p_i)\ dp_i}{\int_{\mathcal{A}} \mathcal{F}(p_s, p_i)\ dp_i}\ . \tag{5.20}$$

Fig. 5.14 reports the values obtained for $\mathcal{R}(\boldsymbol{p})$ in the topology under consideration. From the plot, it is apparent how cooperation is much more likely to be triggered when the relay is located in the middle of a line connecting source and destination. In fact, if the cooperator is in the proximity of the source, the path losses that affect the $\boldsymbol{S}$-$\boldsymbol{D}$ and the $\boldsymbol{C}$-$\boldsymbol{D}$ channels

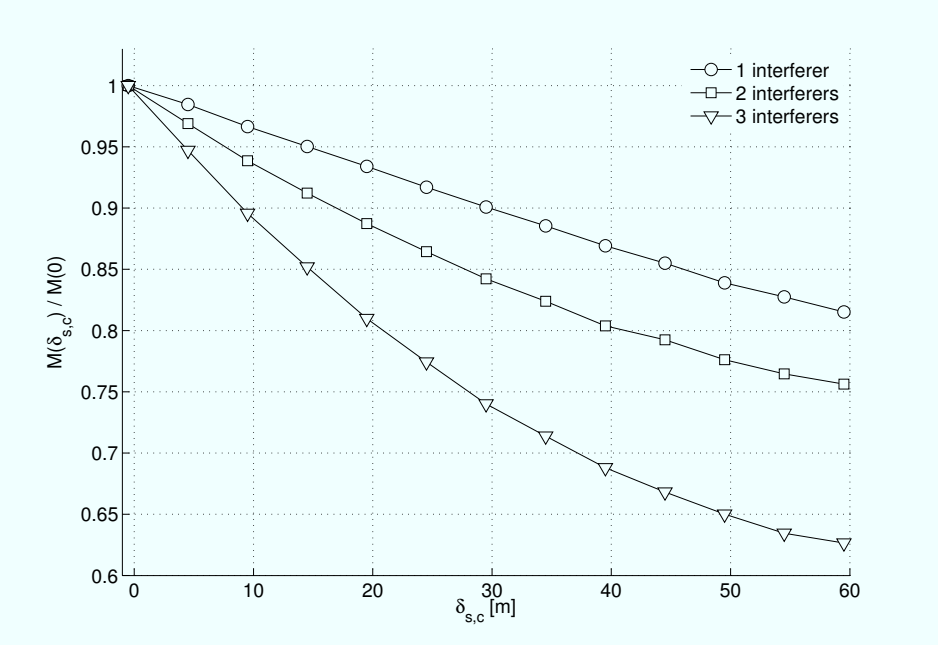


Figure 5.15. *Ratio $\mathcal{M}(\delta_{s,c})/\mathcal{M}(0)$ of the probability for a node lying on the line connecting source and destination at distance $\delta_{s,c}$ from the former to sense the medium idle to the same quantity computed at the position of the source. $k \in \{1,2,3\}$ interferers are distributed over region $\mathcal{A}$ with $\boldsymbol{S}$ and $\boldsymbol{D}$ located at $p_s = \{0,0\}$ and $p_d = \{60,0\}$, respectively.*

are alike, and so are the sustainable transmission times, thus limiting the achievable gain for a two-hop communication. For symmetry, a similar reasoning holds when $\boldsymbol{C}$ is close to the final addressee of the payload.

The analytical framework presented so far makes it possible to draw some important conclusions on the interaction between CSMA and proactive cooperation. As a first observation, comparing the results of Figs. 5.13 and 5.14, we can infer that the employed medium access control policy restricts the set of relay candidates to nodes that are not likely to offer significant benefits in terms of sum rate for payload delivery. This, in turn, reduces the number of triggered cooperative phases (see Section 5.6), with a detrimental effect on the potential gains offered by relaying. Notice that such a consideration is not specific to the protocol implementation that we have proposed, but rather it intrinsically stems from the uneven interference distribution yielded by the carrier sense paradigm.

It is also interesting to remark that even though this effect is already manifest in the basic four-node topology studied up to now, its impact is exacerbated when more realistic scenarios are taken into account. We have investigated this aspect by applying the developed framework also to networks with multiple interfering terminals deployed over $\mathcal{A}$ and

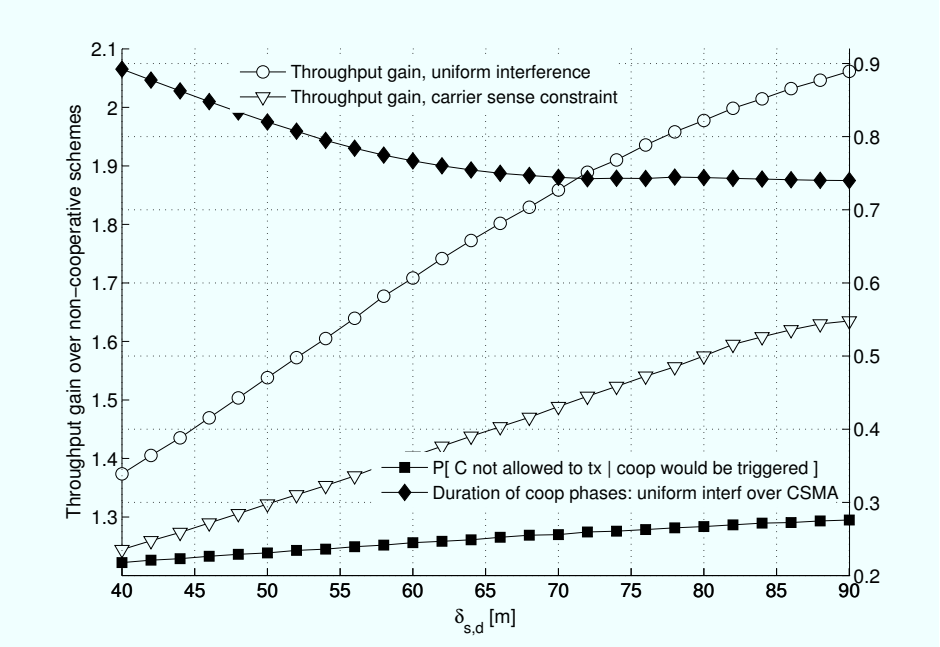


Figure 5.16. *Comparison of proactive cooperation with and without carrier sense constraints for different values of $\delta_{s,d}$. Left axis, white markers: throughput gain over non-cooperative solution. Right axis, black markers: potential cooperative phases lost because the relay senses the medium busy (square markers), and ratio of the average duration of a cooperative transmission with no MAC to the same quantity when CSMA is considered (circle markers).*

subject to CSMA. To this aim, Monte Carlo simulations have been performed, distributing interfering nodes so that each of them satisfies the carrier sense constraint both for the transmission of $\boldsymbol{S}$ and for the transmissions of other active interferers. For the sake of simplicity, we have considered a potential cooperator located along the line connecting $\boldsymbol{S}$ and $\boldsymbol{D}$, so that the probability $\mathcal{M}(\boldsymbol{p})$ that $\boldsymbol{C}$ senses an idle medium, i.e., that it satisfies requirement i), only depends on its distance from the source. Fig. 5.15 reports the ratio $\mathcal{M}(\delta_{s,c})/\mathcal{M}(0)$, i.e., the sought distribution normalized to the probability for a node in p_s to sense the medium idle despite the active interfering communications, for $p_s = \{0, 0\}$ and $p_d = \{60, 0\}$. The results confirm that even a contained increase in the number of interferers further concentrates relay candidates, i.e., nodes that sense the medium idle, in the proximity of the source, thus reducing the chances of enabling efficient cooperative phases. In particular, while with a single interferer (see also Fig. 5.13) the probability that a relay is allowed to access the channel is just 10% lower when located in the most favorable position for throughput maximization, i.e., $p_c = \{30, 0\}$, than when positioned close to the source, the gap increases to more than 15% and 25% for networks with two and three interferers, respectively.

The influence of CSMA on the availability of cooperators shown by Figs. 5.13, 5.14 and

5.15 has a direct impact on the overall network performance. From this viewpoint, it is insightful to compare the throughputs achieved by the same relaying strategy under uniform and carrier sense-biased interference conditions. To this aim, let us refer again to a topology composed of a source-destination pair at distance $\delta_{s,d}$ and of a relay node uniformly distributed over region $\mathcal{A}$. By means of Monte Carlo simulations we have computed the average throughputs offered by the plain and cooperative strategies described in Section 5.6 both when interference with average power σ_i^2 affects $\boldsymbol{D}$ and $\boldsymbol{C}$, and when an interferer $\boldsymbol{I}$ subject to carrier sense based medium access control is distributed over $\mathcal{A}$.

The curves with white markers in Fig. 5.16 report the obtained results in terms of performance gain of the relaying schemes over their non-cooperative counterparts in the two scenarios as a function of $\delta_{s,d}$. In order to have a fair comparison, for each value of the distance between $\boldsymbol{S}$ and $\boldsymbol{D}$, σ_i^2 in the non-CSMA case has been set equal to the average interference level perceived at the destination when $\boldsymbol{I}$ has to obey the medium access policy. The plot clearly stresses the influence of the considered MAC, as the achievable throughput gain reduces by up to 40% when carrier sense is implemented. Also notice that the performance gap in the two scenarios becomes more pronounced for higher values of $\delta_{s,d}$. As highlighted in the first part of our discussion (see Fig. 5.12), a larger distance between $\boldsymbol{S}$ and $\boldsymbol{D}$ favors relayed transmissions with respect to data delivery over direct links. Nonetheless, such a beneficial effect is counterbalanced by the biased distribution of potential relays that appears when nodes adhere to the principles of CSMA. Indeed, the farther away the destination from the source, the less the area where cooperators offer the maximum improvement is protected by the carrier sense mechanism.

It is important to observe that this performance degradation stems from two key factors: on the one hand CSMA reduces the number of cooperative phases that can take place, and on the other hand it affects their effectiveness. This inference is supported by the curves with black markers in Fig. 5.16, whose values are to be referred to the y axis on the right side of the plot. The former detrimental effect is shown by the square-marked line, that reports the fraction of times in which Coop-CSI has to refrain from relaying because a candidate that would offer performance gain over the direct link is not allowed to access the channel. Even when $\boldsymbol{S}$ and $\boldsymbol{D}$ are close to each other, more than 20% of the cooperative attempts have to be aborted, with such percentage increasing up to 30% for larger values of $\delta_{s,d}$. The circle-marked curve, in turn, depicts the ratio of the average duration of an actually performed cooperative transmission when no medium contention and uniform interference are considered to the same quantity evaluated when CSMA is applied, showing an efficiency

loss of up to almost 30%. It is then apparent that even when cooperation takes place carrier sense hampers its potential by enforcing the usage of relays that are located at suboptimal positions.

In conclusion, the analysis carried out in this section has pointed out some fundamental repercussions induced by medium access control based on carrier sense on cooperative mechanisms when CSI is available. While these aspects are already noticeable in simple topologies, their impact may be further stressed when relaying is applied to more realistic and large networking scenarios, where additional issues related to the interaction of several nodes come into play. With a view to understanding these problems, the next section will present the results of extensive simulation campaigns that we have performed.

5.8 Simulation Results for Proactive Cooperation

The proactive cooperative schemes proposed in this chapter have been tested by means of Omnet++ simulations in the 35-node scenarios described in Section 5.4, with main network and protocol parameters tuned as in the absence of CSI (see Tab. 5.1). In addition, the margin coefficient ε' (see Section 5.6), which intrinsically depends on the variability of the wireless environment and induces a critical tradeoff between transmission time and decoding probability, has been set after dedicated studies so as to maximize the aggregate throughput.[12]

As a general observation, our simulations have shown how the remarkable performance gains pointed out for relaying in simple topologies by the analytical framework of Section 5.7 (see Fig. 5.16) plummet when the protocols are implemented in large-scale ad hoc networks. Such a behavior is evident in Fig. 5.17, which reports the aggregate network throughput against the nominal load per node λ, highlighting how the improvements achieved by Coop-CSI over CSMA-CSI are curbed to 10% at saturation in the considered scenarios. Let us remark that this result is of broad applicability and of particular relevance, since, as discussed, the proposed strategies represent a bound for the class of protocols that take advantage of channel state information to realize proactive cooperation.

A first important insight on the issues that thwart the effectiveness of relaying is provided in Fig. 5.18. The plot highlights that Coop-CSI resorts to a two-hop link in less than

[12]Incidentally, notice from Tab. 5.1 that ε' is larger that ε, i.e., the margin coefficient used for DHARQ. When reactive cooperation takes place, the NACK sent by the destination prevents nodes in its surroundings from accessing the channel and thus reduces the dynamic of interference that affects a relayed transmission. Such a protection is not available for Coop-CSI, and a higher margin has to be provided.

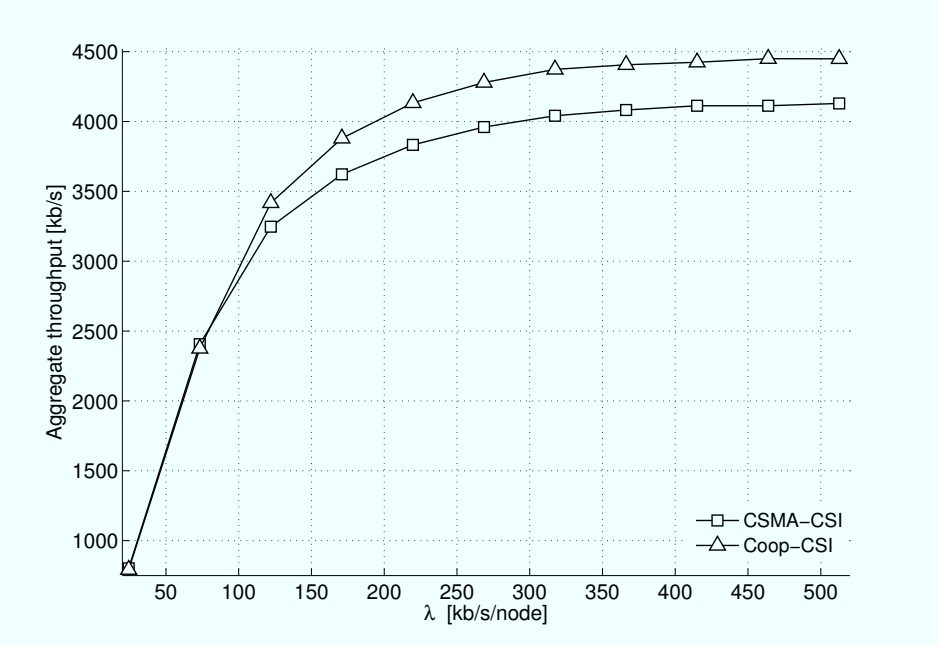


Figure 5.17. *Aggregate throughput for proactive cooperation vs nominal load expressed in kb/s/node.*

15% of the initiated communications, whereas non-cooperative transmissions take place in almost 90% of the cases. Recalling the discussion of Section 5.6, packets are directly delivered over the source-destination channel if either i) no node is available for relaying, i.e., not involved in other ongoing activities and sensing an idle medium, or ii) some candidates are present but none of them is able to reduce the overall duration of the data transfer by cooperating. The impact of these factors is reported by the white-marked lines in the figure, which sum up to the *non-coop phases* curve. In the first place, as expected, the higher the traffic injected in the network, the lower the probability of finding terminals which are idle and allowed to access the channel due to the harsher interference level (*no avail. relays*). Fig. 5.19 delves into the reasons that lead to such a relay unavailability, prompting how in the vast majority of cases cooperators are blocked by the power level sensed on the medium ($\sim$ 80%) and, less prominently, by the hidden terminal problem ($\sim$ 20%), while the impact of virtual carrier sense and of terminals that are already transmitting or receiving another packet is negligible. This result further supports the intuition that the low efficiency of reactive cooperation is not related specifically to the protocol that we propose, but rather arises from the intrinsic nature of carrier sense multiple access.

Going back to the discussion of Fig. 5.18, the plot suggests that the predominant obstacle to the establishment of two-hop links is by far the lack of nodes that, while sensing a power level on the medium below the carrier sense threshold, are also capable of shorten-

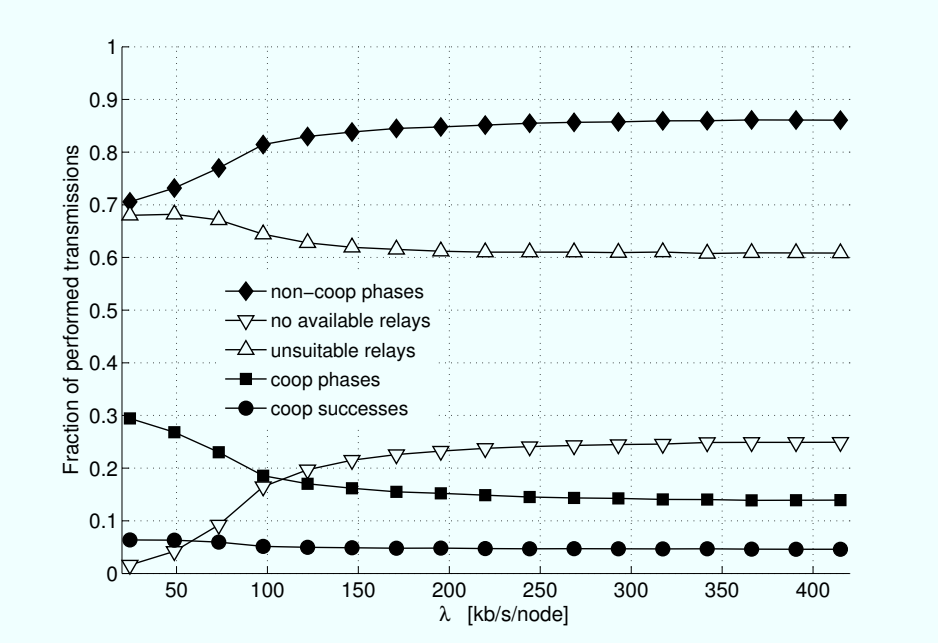


Figure 5.18. *Fraction of performed proactive cooperative phases over all the transmissions vs nominal load expressed in kb/s/node. Curves with filled markers indicate the reasons for not performing a cooperative transmission.*

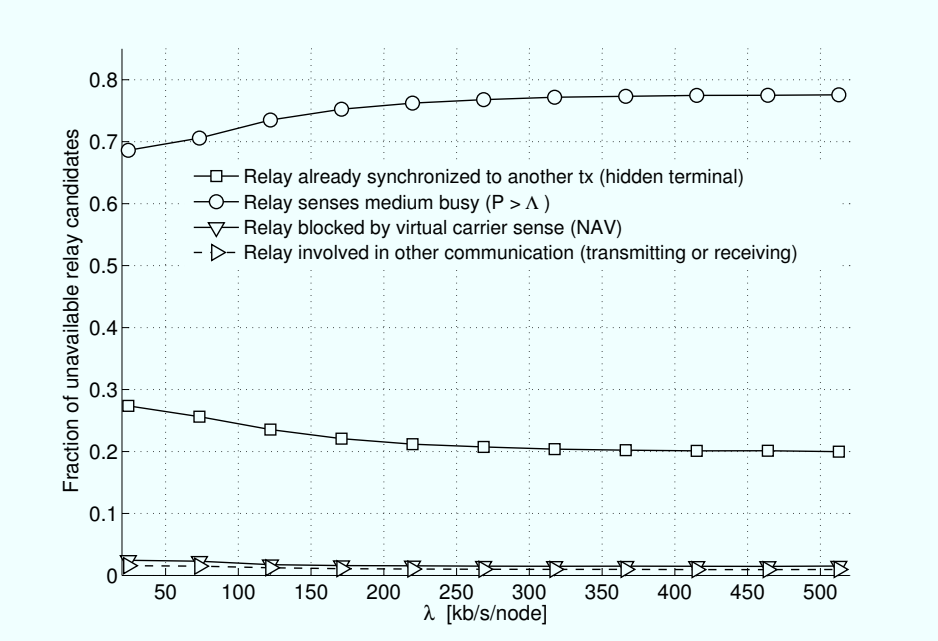


Figure 5.19. *Impact of the reasons that may lead a relay candidate not to be available for cooperation regardless of the channel quality it exhibits towards the destination, depicted against the nominal load expressed in kb/s/node.*

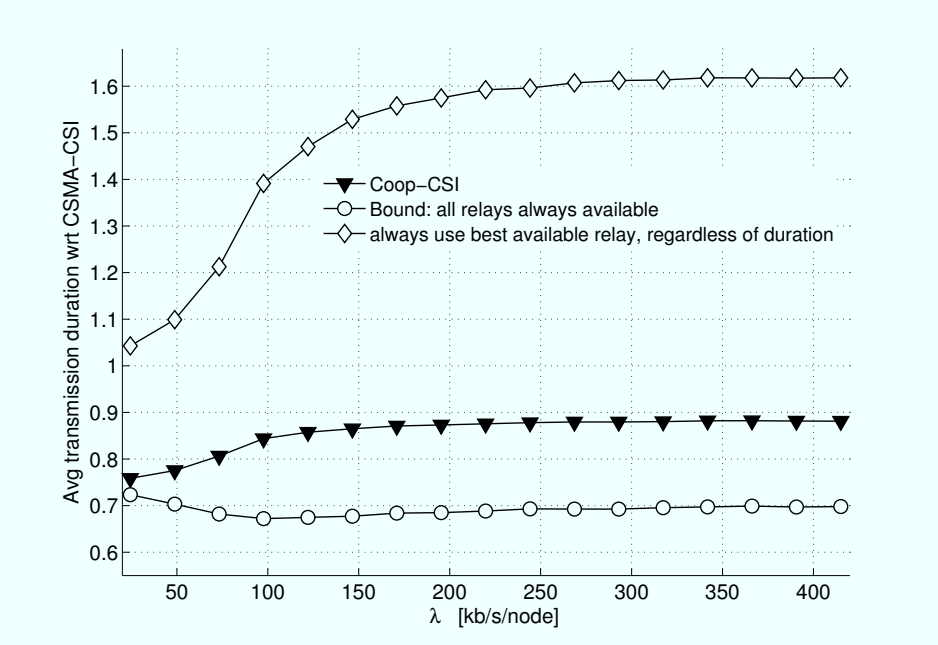


Figure 5.20. *Average duration of a transmission for different proactive cooperative schemes normalized to the average transmission duration for CSMA-CSI.*

ing transmission times (*unsuitable relays*), responsible for more than 70% of the non-relayed phases at saturation. It is thus apparent that a first and major limitation to the effectiveness of Coop-CSI stems from the reduced share of cooperative links it triggers. On the other hand, Fig. 5.18 also emphasizes that, even when performed, relayed transmissions are not efficient, with a success rate lower than 50% (*coop successes*).

The former issue is further investigated in Fig. 5.20. Here, the solid line describes the ratio of the average duration for a complete communication that characterizes Coop-CSI (either over a direct or a split link) to the same quantity achieved with the plain CSMA-CSI strategy, and points out again the limited improvements discussed earlier in terms of aggregate throughput. Conversely, the dashed curve shows the transmission duration gain that would be secured if all the neighbors of a source node were always available for cooperating, i.e., if the protocol could always count also on terminals that are actually prevented from relaying because of the power level sensed on the medium. The reported trend confirms the inferences prompted by our analytical discussion, e.g., Figs. 5.12 and 5.16, highlighting how even in large-scale distributed networks proactive cooperation may lead to important performance gains (up to 30%) if relays do not have to obey to CSMA. Such a potential, however, is intrinsically encumbered by the medium access policy, as nodes in the most favorable positions for cooperating, i.e., those that offer some advancement to the destina-

tion, are typically located at the periphery of the spatial region reserved by the carrier sense mechanism for the source's transmission and thus also experience non-negligible chances of sensing a busy channel. This fundamental discrepancy is further supported by the dash-dotted curve in Fig. 5.20, that plots the average transmission duration for Coop-CSI if the protocol always resorted to cooperation by selecting the best available relay candidate without considering the possibility of employing a direct link to the destination, i.e., applying (5.14) without the $T_{s,d}$ term, normalized to the duration of a communication in CSMA-CSI. Once more it is manifest that terminals allowed by the medium access control to transmit trigger two-hop links with much poorer performance compared to the basic solution. The result, then, clearly explains the large share of untapped cooperative phases that characterizes Coop-CSI as discussed in Fig. 5.18. In this perspective it is also important to remark how these issues would not be mitigated by more aggressive medium access approaches, i.e., allowing relays to access the medium even if the power level they perceive is above Λ, as they inherently depend on the relationship between CSMA and cooperation. Dedicated simulations, whose outcome is not reported here due to space constraints, have shown that setting a less stringent carrier sense threshold Λ_{rel} for cooperating terminals, i.e., $\Lambda_{rel} > \Lambda$, does not lead to an overall performance improvement. Indeed, the higher number of cooperative phases triggered by larger values of Λ_{rel} is more than counterbalanced by their reduced effectiveness resulting from the increased aggregate interference level they generate, as discussed in Section 5.4.

Fig. 5.21 focuses on the reasons that may cause a cooperative link to fail once it has been triggered. In order for a two-hop communication involving terminals $\boldsymbol{S}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ to be successful, three conditions have to be met: i) the data unit has to be retrieved at $\boldsymbol{C}$; ii) the destination has to decode the header of the packet sent over the $\boldsymbol{S}$-$\boldsymbol{C}$ link, so as to cache a corrupted version of the payload for later combining; and iii) the redundancy sent by the relay has to enable a data decoding at $\boldsymbol{D}$. By virtue of the rate adaptation policy of Coop-CSI as well as of the protection offered by the margin coefficient ε', the impact of factors ii) and iii) is rather limited (*data loss at relay* and *data loss over C-D link* curves). Conversely, the plot shows that the vast majority of cooperative phases fail since the destination does not retrieve the header of the payload sent by the source. In turn, such header losses may occur either because $\boldsymbol{D}$ does not synchronize to the incoming packet, or because the decoding does not succeed due to channel impairments, with the former reason possibly induced by the hidden terminal problem or by too low a received power level.[13] Recalling the analysis

[13] If the power level of an incoming packet is below a synchronization threshold, see Tab. 5.1, a node regards it as noise and does not even try to synchronize to it.

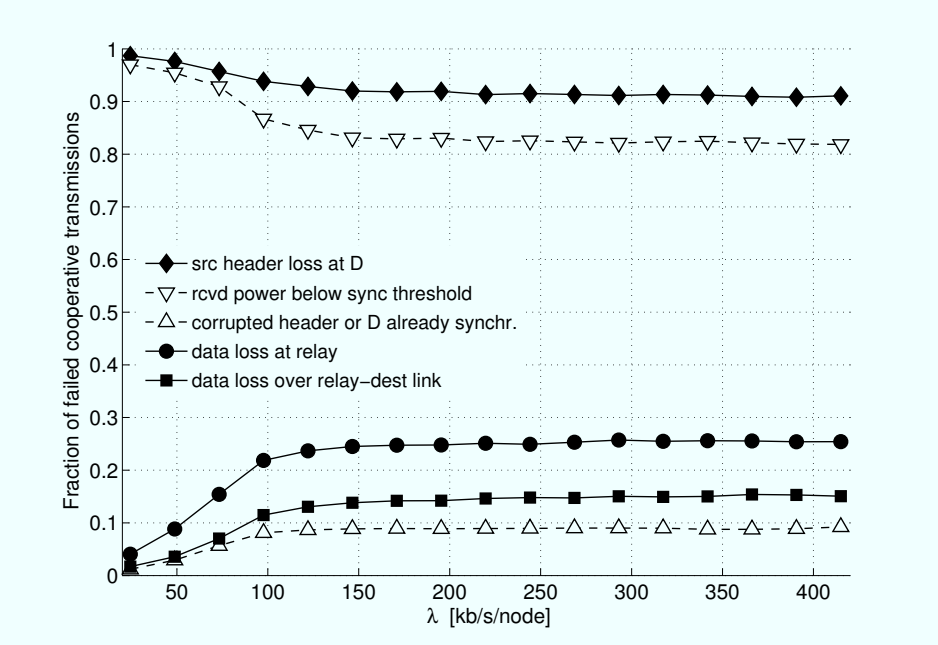


Figure 5.21. *Reasons that lead to the failure of proactive cooperative transmissions vs nominal load expressed in kb/s/node.*

of Section 5.4, the foremost reason for the destination to miss a header addressed to it when reactive cooperation is implemented is that the node is already synchronized to another ongoing transmission. Instead, Fig. 5.21 points out an insufficient received power as the key factor for cooperation not to be successful (dashed white-marked curves) in Coop-CSI. This difference can be explained in light of the spatial distribution for relay nodes induced by CSMA. As discussed, terminals available for proactive cooperation tend to be located with high probability in positions that do not allow particularly fast two-hop transmissions. Therefore, when a split link is actually triggered by Coop-CSI, the direct connection between source and destination is likely to offer extremely poor performance, in particular because affected by a deep fade. Such a condition, however, also weakens the useful power perceived at the addressee, hampering the header decoding and thus the effectiveness of relayed phases.

Following this argument, one could think of letting cooperation take place only when a minimum channel quality over the $\boldsymbol{S}$-$\boldsymbol{D}$ link is guaranteed. Nonetheless, while this approach would certainly increase the success probability of two-hop links, it may also reduce the already small share of cooperative communications that are performed. This tradeoff is apparent in Fig. 5.22, which depicts the behavior of the protocol when such a mechanism is implemented. According to this modified version of Coop-CSI, a source node checks for

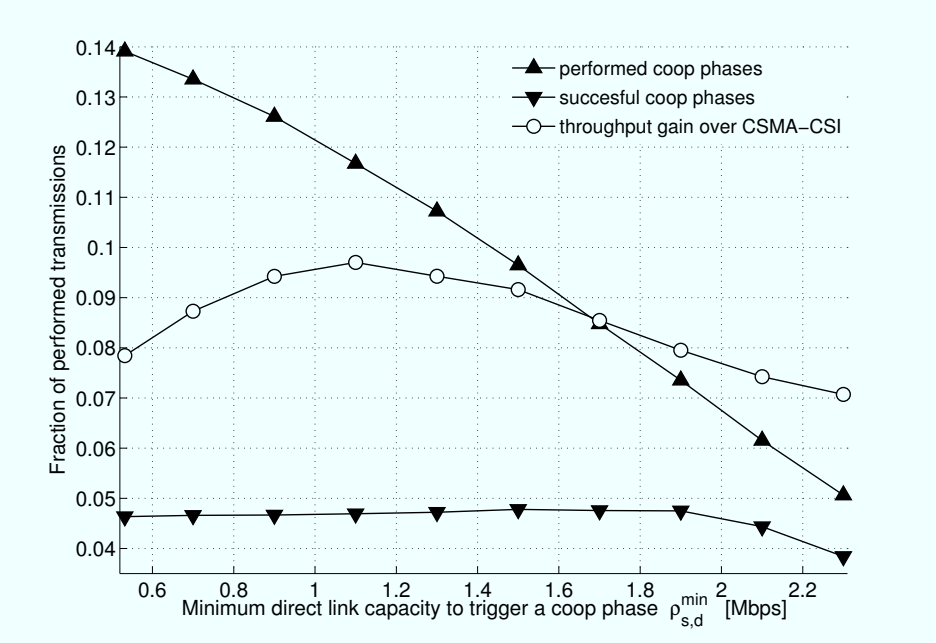


Figure 5.22. *Impact of cooperative phases and aggregate throughput gain with respect to CSMA-CSI when different thresholds of the minimum source-dest link capacity are considered for triggering cooperation.*

potential cooperators only if the minimum sustainable rate $\rho_{s,d}$ with its destination is above a threshold $\rho_{s,d}^{min}$, reported on the x axis of the plot. If so, the procedures described in Section 5.6.2 take place as usual. Otherwise, $\boldsymbol{S}$ resorts to a direct link with $\boldsymbol{D}$ or just drops the packet if $\rho_{s,d} \geq \rho_{min}$.[14] The figure highlights how an improvement in terms of reliability for proactive cooperation, i.e., higher values of $\rho_{s,d}^{min}$, comes at the expense of a drastic fall of the number of relaying phases. The outcome of such a compromise at a system level is reported by the white-marked line. The throughput gain over CSMA-CSI does not vary significantly, proving once more how the poor performance of proactive relaying in the large-scale ad hoc networks considered here is intrinsically related to the effects of the carrier sense-based medium access policy and not a consequence of the specific protocol implementations proposed in this work.

[14]Notice that the basic version of the protocol simply sets $\rho_{s,d}^{min} = \rho_{min} = 0.9$ Mbps, as per Tab. 5.1.

Chapter 6

Conclusions

This thesis has focused on the efficient design of cooperative strategies in wireless networks. While such solutions have been extensively shown in the literature to enable remarkable gains over their non-cooperative counterparts, they also require a higher level of coordination among nodes in order to be effective. From this viewpoint, for example, it is of paramount importance to properly detect when cooperation can actually be useful, so as to avoid unnecessary efforts, as well as to carefully identify who shall take part in the collaborative process. These tasks become even more critical and challenging in ad hoc networks, which are characterized by a completely distributed and self-organizing nature. In view of this discussion, not only have we devoted this work to the definition of novel cooperative schemes, but also we have considered the suitability of different link layer policies to support such approaches, especially when large and congested scenarios come into play.

A first instance of collaborative behavior has been introduced in Chapter 2 for ad hoc networks where directional transmissions and receptions are enabled by the use of multi-antenna systems. In this context, the directional exchange of data and control packets tends to spoil the perception nodes have of the activity going on in their surroundings, disrupting link layer coordination and thus severely bounding the potential of the network. Starting from this remark, our strategy proposes terminals to cooperate by sharing information on the current channel usage so as to better harmonize medium access, and has been shown by means of extensive simulations to effectively tackle such issues, leading to a significant increase in the sustainable spatial multiplexing and on the overall performance in terms of throughput, energy consumption and delivery latency.

We have then moved our focus to systems with single-antenna nodes that communicate omnidirectionally and we have concentrated on decode-and-forward cooperative tech-

niques. According to these schemes, terminals may act as relays for a source-destination pair by sending redundancy to the addressee of a packet on behalf of the data originator in case of a decoding failure over the direct link, with the aim of speeding up and improving the reliability of basic ARQ policies. The first step of our study has been the proposal of a novel paradigm, called hybrid cooperative-network coded ARQ, which, by smartly merging the principles of decode-and-forward and network coding, allows relays to exploit cooperative phases to serve their own traffic as well. In particular, a node that helps neighbors in need can transmit a linear combination of a packet taken from its own queue and of the payload the destination could not decode, in place of just a copy of the latter. The advantage offered by this solution is twofold. In the first place, our approach encourages terminals to take part in the collaborative process by offering them a tangible benefit for the support they provide, going thus beyond an intrinsic limitation of the basic relaying paradigm, that asks potential cooperators to behave in a completely selfless way. Secondly, the capability of coding together multiple data units at no additional cost in terms of bandwidth has the potential of triggering important gains as far as throughput and energy consumption are concerned.

Besides introducing the idea of coded retransmissions, we have also presented two link layers that effectively support it in two completely different environments, namely distributed systems, epitomized by CSMA-based networks, and centralized systems, characterized by a TDMA-based medium access control. The behavior of such schemes has been thoroughly tested by means of simulations in a variety of scenarios, including single- and multi-hop networks, and several observations of interest have been drawn. On the one hand, we have proven that hybrid cooperative-network coded ARQ boosts the overall performance in all the considered cases, doubling the gains achievable over plain ARQ with respect to the well-known and widespread decode-and-forward approach. On the other hand, it has emerged that the same collaborative paradigm may lead to completely different performance when implemented obeying to distinct spectrum sharing policies, so that while for instance a 40% gain can be achieved by our solution with TDMA in terms of aggregate throughput, no more than a 15% improvement is attainable with CSMA.

Chapters 4 and 5 have been devoted to the investigation of the reasons that lead to such a different behavior, and have highlighted how it is not due to the specific protocol implementations that we propose, but rather stems from the intrinsic capability of distinct medium access policies to support cooperation. In particular, the completely distributed nature of CSMA has been shown to severely stymie relaying, as it both often prevents nodes from caching packets for later combining and hinders medium access for cooperators. The

impact of such issues, instead, is significantly contained when even a loosely centralized system like TDMA is implemented. Furthermore, by means of both analysis and simulations we have proven that carrier sensing biases the geographic distribution of terminals available for cooperation, locating them in unfavorable positions and hampering the effectiveness even of schemes that go beyond the decode-and-forward solution, such as coded cooperation or proactive cooperation with rate adaptation.

In conclusion, this thesis has buttressed the idea that the inventive step of defining novel cooperative schemes has always to be accompanied by careful considerations on how effectively such principles can be supported by the underlying link layers, especially when large environments fall within the scope of applicability. Moreover, our work, by thoroughly investigating the impact of several medium access control policies, has provided important hints on how to properly design collaborative approaches and has highlighted many issues that, while usually neglected in the literature, are of critical importance for the efficiency of cooperation among wireless nodes.

Appendix A

Energy-Efficient Routing in Wireless Sensor Networks

In this appendix we group the outcome of some research activities, carried out during the Ph.D., whose focus falls out of the main topic of the thesis. In particular, we propose some interesting studies on energy-efficient routing protocols for Wireless Sensor Networks, spanning the issues of cross-layer MAC-NET design, sink mobility, and information security. All the results presented here have been obtained in collaboration with the *Energy Management and Sensors Systems* group of the IBM Zurich Research Laboratory (Rüschlikon, Switzerland).

A.1 Introduction

Wireless Sensor Networks (WSNs) are attracting an ever growing interest in the research community, thanks to their potential of enabling a wide range of new applications in both the military and civilian domain, such as building and battlefield surveillance, environmental monitoring, disaster prevention, and homeland security. These networks are typically composed of a large number of battery powered and computationally limited wireless nodes that are capable of offering some sensing functionalities, e.g., temperature, humidity, light, vibrations or potentially harmful gas. Such devices may be scattered over a wide geographical area and have to operate without external intervention for periods of time possibly in the order of months or years. The ultimate goal of a WSN is generally to collect environmental information and to deliver it over the wireless medium to one or more central units with less stringent energetic and computational constraints, referred to as *sinks*, which, in

turn, decide how to take advantage of the gathered data, e.g., by storing them or by taking proper actions. The topological structure that often characterizes sensor networks, together with the reduced transmission power suggested by battery saving purposes, enforces nodes to form an ad hoc network to route data from information sources to sinks in a multihop fashion. Therefore, in recent years several routing schemes specifically tailored to WSNs have been proposed, which can be partitioned in *static* and *dynamic* ones.

In *static* routing, a route discovery procedure is exploited to identify a single path from a source to a destination, which is later used to forward data packets. One of the best known examples is represented by the Ad hoc On demand Distance Vector (AODV) scheme [77]. With AODV, a route is established only when a node has data to deliver to a destination $\boldsymbol{D}$ and a path to it is not already known. In this case, the source $\boldsymbol{S}$ broadcasts a Route REQuest (RREQ) packet containing the address of the sink. Each node that receives this message checks its routing table to determine whether it can provide a path to $\boldsymbol{D}$. If not, the RREQ is rebroadcast. Otherwise, the terminal, potentially $\boldsymbol{D}$, initiates a return procedure by sending a unicast Route REPly (RREP) packet to the node it has received the RREQ message from. Upon receiving a RREP, a node updates its routing table, as it becomes aware of whom it has to forward packets addressed to $\boldsymbol{D}$ to, and iterates the procedure. Once the RREP message reaches $\boldsymbol{S}$, the path is built and can be used to deliver data. AODV does not need any geographic information to work properly, since the destination is located by flooding messages through the network. On the other hand, the algorithm provides a single path to $\boldsymbol{D}$. Therefore, as soon as one of the nodes along the discovered path becomes unavailable due to a weak radio link, battery depletion, or security attacks, the whole route collapses, and a new construction procedure is needed to establish an alternative connection. Let us also remark that these issues are not specific to AODV, but they are also valid for all single- or multi-path protocols that see a route as a predetermined sequence of forwarding nodes [78,79], and may lead to a severe performance degradation in WSNs.

On the other hand, *dynamic* routing does not separate path construction and data forwarding. Instead, these protocols route information contents by locally selecting the best-suited next forwarding node according to a metric, without knowing in advance the complete path that will be followed to reach the destination. Up to now, dynamic routing algorithms have mostly been proposed for networks where each node knows its own geographic location (geographic routing). In this context, Geographic Random Forwarding (GeRaF) [80] is a remarkable example. According to this scheme, a source that has data to deliver to a destination $\boldsymbol{D}$ transmits a Request-To-Send (RTS) packet with its and $\boldsymbol{D}$'s geographical co-

ordinates. Terminals that decode this message evaluate the geographic advancement they offer towards $\boldsymbol{D}$, compute a metric that is inversely proportional to this value, and set a backoff timer proportional to the metric. The node closest to the destination replies first with a Clear-To-Send (CTS) message, proposing itself as data forwarder. On the contrary, all other terminals overhear the CTS and give up the contention. Once the procedure has been accomplished, the source sends the data packet to the identified forwarding node, which, in turn, iterates the algorithm. The capability of dynamically choosing the next forwarding node among a set of potential candidates instead of relying on a static path strengthens dynamic routing approaches with respect to node unavailability due to weak radio link, battery depletion, or security attacks. Moreover, since the next-hop selection only requires a local information exchange between a node a its neighbors, these algorithm scale well for networks with a large number of wireless devices [81]. Such a result, instead, cannot be achieved by static routing protocols. Nevertheless, the aforementioned advantages are often achieved by assuming geographical knowledge at sensor nodes, and their applicability is then restricted, e.g., to scenarios where GPS is available.

In light of this discussion, in this appendix we propose novel routing protocols that, while relying on the dynamic paradigm, try to overcome some of the problems that affect it. We start in Section A.2 by discussing a scheme that extends geographic routing relying on a modified metric for the choice of the next-hop that considers not only advancement but also radio channel conditions and remaining battery energy, with the aim of prolonging the network lifetime. Then, in Section A.3, we tackle the issue of sink mobility, which severely hinders the effectiveness of multi-hop schemes that exploit location information to deliver packets. Finally, Section A.4 focuses on the design of a routing solution that goes beyond the need of geographical information to opportunistically identify the best suited next-hop. In this context, not only is the proposed solution able to extend the applicability of the dynamic paradigm to several scenarios, but also it offers important features in terms of network security against the presence of malicious nodes that may behave improperly and hamper the reliability of WSNs in critical scenarios.

A.2 A Class of Cross-Layer Optimized Protocols for Geographic Routing in WSNs

As discussed in Section A.1, routing algorithms play a role of pivotal importance in WSNs. In this perspective, *scalability*, due to the large number of nodes that may compose

the network, and *energy efficiency*, due to the battery-constrained nature of the involved devices, are two of the primary concerns for an effective design of these schemes. While the former can be achieved by resorting to solutions based on geographic routing, capable of forwarding data on the basis of only a local information exchange between a sensor and its neighbors [81], a road to energy efficiency is represented by the design of the network control algorithms with a cross-layer approach [82], jointly defining protocol functions that are assigned to separate layers in the classical communication model. From this viewpoint, a remarkable example of cross-layer optimized routing is represented by [83], where the authors introduce a scheme that incorporates the medium access control concepts of decode and forward cooperative relaying and leapfrogging at the network layer in order to improve energy efficiency by taking advantage of the broadcast nature of the wireless medium.

Starting from these observations, in this section we present and discuss a class of cross-layer optimized protocols for routing in energy-constrained WSNs, namely Distributed Routing, Cooperative Distributed Routing and Distributed Routing with Cooperative Relaying and Leapfrogging, together with a reference protocol for non-geographic schemes called Shortest Path Routing. All these algorithms assume carrier sense with collision avoidance (CSMA/CA) as underlying medium access control policy, and location-aware sensor nodes, being thus able to apply the geographic routing paradigm to the aim of achieving scalability. Moreover, the proposed solutions stress the cross-layer approach by both considering information on the instantaneous state of the radio links and residual battery energy level of sensor nodes for the path selection and by incorporating into the routing procedure some link layer concepts.

A.2.1 Protocols Description

A.2.1.1 Shortest Path Routing (SPR)

Shortest Path Routing follows a greedy approach for data forwarding. The next hop is always the closest node to the destination among the set of neighbors. No distributed route construction procedure is applied, and the path from a source to a sink is uniquely identified by the network topology.

A.2.1.2 Distributed Routing (RoDi)

Geographic routing algorithms typically consider advancement towards the destination as the only criterion to identify the next hop. However, this approach has some important drawbacks. First of all, a minimization of the route length does not necessarily induce a

reduction of the overall number of transmissions required to deliver data to the destination. In fact, the best positioned node may experience poor channel conditions with the source, and a data transmission addressed to it may fail, triggering a retransmission attempt. Secondly, the lifetime of sensor nodes should be taken into account. If a greedy approach is used, nodes that are closer to the destination are exploited more often for data forwarding, and thus incur a higher energy consumption. This likely results in early battery depletion and may induce holes in the topology with negative effects on the overall packet delivery capabilities.

In light of this, we propose a geographic routing protocol called Distributed Routing that introduces an enhanced version of the metric used for the distributed selection of the next hop. When a source $\boldsymbol{S}$ has data to deliver to a destination $\boldsymbol{D}$, it transmits an RTS message with its own location and the location of the destination. A node x that decodes the packet computes a metric m_x that takes into account its distance to the destination, $d_{x,D}$, the instantaneous channel gain $|h_{S,x}|$ with $\boldsymbol{S}$, and its remaining battery energy E_x:

$$m_x = E_x \cdot \frac{|h_{S,x}|^2}{d_{x,D}^{\alpha}}\,, \tag{A.1}$$

where α is the path loss coefficient of the radio channel. This metric increases for nodes with high channel gain to the source (i.e., likely to decode a successive data packet), high remaining energy and which are close to the destination. Once the computation is performed, a backoff timer inversely proportional to m_x is set and a contention procedure is started. During the backoff period, the node listens to the medium. If the sensed power exceeds a given threshold, the node gives up the contention and goes back to the idle state. On the contrary, if the timer expires, the node transmits a CTS frame with its ID, proposing itself as next hop. Notice that this procedure manages to select a relay and to negotiate the channel according to CSMA/CA at once.

A.2.1.3 Cooperative Distributed Routing (RoCoDi)

Let us assume that a node x has been selected as next hop by a geographic routing algorithm and suppose that the subsequent data packet cannot be decoded. In this case, a protocol like RoDi follows the CSMA/CA approach and requires the preceding node to perform another attempt addressed to x. However, such a retransmission is not likely to succeed before the channel conditions that induced the failure change favorably. On the other hand, thanks to the broadcast nature of the wireless medium, other nodes may have decoded the data packet even though it was not intended for them. The decode-and-forward paradigm

(see, e.g., Chapter 3) proposes that one of these nodes immediately retransmits the frame in place of the preceding node. In this way, two copies of the same packet received over independent channels are available at x, that can perform Chase combining, significantly increasing the probability of a successful decoding by taking advantage of spatial diversity [30].

In this work, we present a protocol called RoCoDi that extends RoDi in order to exploit cooperative relaying. Let $\mathcal{R}$ be the set of nodes that decode a data packet sent by the preceding node. If the reception at the next hop x fails, the node transmits a NACK frame asking for a retransmission. A node $r \in \mathcal{R}$ that receives this packet computes a metric m_r and sets up a backoff timer inversely proportional to it. The metric stems from the same principles that lead the next hop selection, and is defined as:

$$m_r = E_r \cdot \frac{|h_{r,x}|^2}{d_{r,x}^{\alpha}}\,, \tag{A.2}$$

where the channel gain between r and x is estimated from the reception of the NACK frame. During the backoff phase, carrier sensing is performed. If the power on the medium exceeds a given threshold, the node assumes that someone else wins the contention and goes back to idle state. On the contrary, if the timer expires, the node acts as relay by transmitting cached data. At the end of this phase, x either sends out an acknowledgment and proceeds with the forwarding or transmits a further NACK (if data were not decoded or no relay was present), passing the token back to $\boldsymbol{S}$ (that may perform another attempt or trigger a link failure).

A.2.1.4 Distributed Routing with Cooperative Relaying and Leapfrogging (RoCoDiLe)

The broadcast properties of the wireless medium offer, besides the discussed advantages of cooperation, the possibility of opportunistically shorten routes. Let us assume that a cooperative relaying phase has taken place with a node r having performed a data retransmission addressed to x on the behalf of $\boldsymbol{S}$. Nodes that are closer to the destination than x may happen to decode the relayed data packet even though it was not addressed to them. In this case, if one of such nodes takes over the task of forwarding data in place of x, it is possible to further approach the sink without undergoing additional hops. This mechanism is called *leapfrogging* and aims at reducing the number of transmissions required to deliver the payload, inducing potential gains in terms of delay and energy consumption. These benefits come at the expense of an increased protocol complexity required to coordinate the nodes that take part in the data exchange.

In this work, we discuss a protocol called RoCoDiLe, proposed in [83], that extends

RoCoDi by including a distributed algorithm based on carrier sensing to exploit leapfrogging. Let $\mathcal{L}$ be the set of nodes that successfully decode a data packet sent to x by a relay. A node in $\mathcal{L}$ compares its distance from the destination to the distance of x[1] from the destination. If the node, say l, is closer than the current next-hop to D, a carrier sensing contention procedure resembling the ones described for RoDi and RoCoDi is started by setting up a backoff timer inversely proportional to a metric m_l, computed as:

$$m_l = \frac{E_l}{d_{l,D}^{\alpha}} \,. \tag{A.3}$$

The node that wins the contention transmits a leapfrog request (LPFREQ) message addressed to x. If this packet is correctly received, x replies with an acknowledgment frame (LPFACK) and does not proceed with data forwarding. On the other hand, the leapfrog node takes the role of next hop only upon the reception of an LPFACK in order to avoid flow duplications in the network.

A.2.2 Simulation Environment

The protocols described in Section A.2.1 have been evaluated by means of Omnet++ network simulations. The networks were composed by 25 sensor nodes spread inside a 200m $\times$ 200m area with one source and one destination. Two types of network topologies have been considered: grid and random ones. In the former case, the nodes are disposed to form a 5 $\times$ 5 grid with a distance of 25 meters between any two vertically or horizontally aligned nodes. The source is located on the upper-left corner and the sink on the lower-right corner. In the latter case, the positions of the source and the destination are kept unchanged, but all the other nodes are uniformly and independently distributed over the area. In our simulations, we have only considered random topologies which are connected, where at least one route from the source to the sink can be formed with neighboring nodes.[2]

The wireless environment is subject to path loss with exponent $\alpha = 3.5$ and correlated Rayleigh fading. Transmissions are performed at an information rate $R = 2$ bit/s/Hz. Letting SINR(t) be the signal to interference and noise ratio perceived at the receiver at time t and defining the instantaneous channel capacity as $C(t) = \log_2[1 + \text{SINR(t)}]$, an outage event occurs if $C(t) < R$. In such a condition, the packet reception is assumed erroneous, otherwise it is correct.

[1] The position of the current next-hop could be included in the data packet sent by the relay.

[2] We consider as neighbor nodes whose relative distance is at most equal to the length of one hop over the main diagonal of the grid topology.

All the nodes in the network have an initial battery energy equal to 30 J. The power consumption in reception mode has been set to 30 mW, whereas the transmission power has been varied in order to obtain a target SNR.

The source injects traffic at a low rate of 0.4 packets per second. Signaling packets are 16 bytes long, while data packets are composed of 128 bytes. The behavior of the protocols has been assessed by averaging the results obtained performing 20 simulations that lasted 1000 seconds each. In the random topology scenario, each run corresponded to a different placement for the nodes.

A.2.3 Simulation Results

A.2.3.1 Grid Topology

Fig. A.1 shows the metric Packet Delivery Ratio (PDR) as a function of average SNR for the considered protocols. The PDR is defined as the ratio of the number of packets successfully received at the destination to the number of packets injected in the network by the source. The average SNR is experienced at a receiver whose distance from the transmitter equals the length of one hop over the main diagonal.

For high values of SNR, the PDR tends to one for all the protocols as expected. In the low-SNR regime, SPR performs very poorly because it relies on a single static path and thus the probability of reaching the destination is very low. Geographic routing significantly mitigates this problem. With this strategy, if the link with the neighbor which is closest to the destination is not good enough to ensure a successful data transfer, other forwarding opportunities are considered and exploited, enhancing the chances of delivering data to the destination. If RoDi is applied, a PDR gain up to 40% for SNR values between 10 and 15 dB is achieved. Further improvements are obtained if cooperative relaying is employed as can be seen by comparing the curves for RoDi and RoCoDi. For low-SNR values, the two protocols perform similarly. In these conditions, a link between two nodes is likely to fail during the contention phase and therefore cooperation is seldom used. As transmission power is increased, relaying starts to show its influence and gains up to 10% are achieved for mid-SNRs. The cooperative advantage is twofold: not only packet error rates are reduced because of spatial diversity, but also hybrid ARQ phases are triggered more often as the retransmission request is more likely to be correctly decoded, being addressed to a set of nodes and not only to the original data transmitter. As far as RoCoDiLe is concerned, some limited improvements over RoCoDi are obtained. This result is not surprising, as leapfrogging (unlike cooperation) is not meant to increase the robustness of established links but

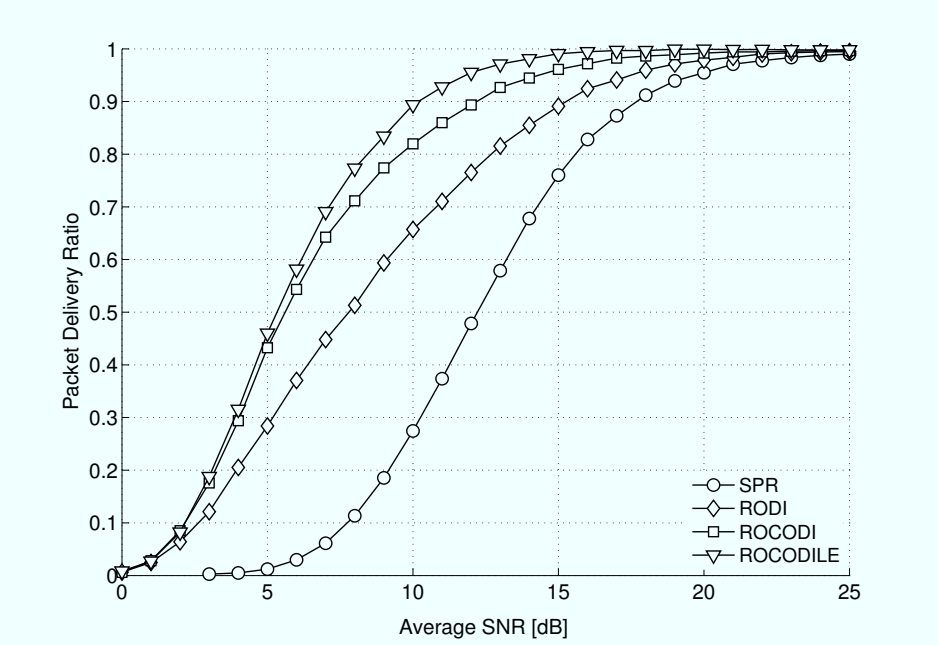


Figure A.1. *Packet Delivery Ratio vs SNR, grid topology.*

rather to decrease the latency. The slight increase in PDR is due to the reduction of the average route length offered by RoCoDiLe (refer to the discussion of Fig. A.2), as shorter paths are statistically less likely to induce packet losses.

Fig. A.2 presents the latency per packet delivery for the different protocols, defined as the average time needed to successfully deliver data from the source to the sink. The metric is depicted as a function of average SNR and has been normalized to the duration of a data packet transmission between two nodes. First of all, we notice that the average latency for SPR decreases as transmission power is increased, because of the reduced number of retransmissions required per successful packet delivery. In the low-SNR region, RoDi and RoCoDi are able to deliver to the sink many more packets than SPR (see Fig. A.1) but with a higher average latency. In these conditions, when a node sends out an RTS packet, the probability that some of its neighbors do not correctly decode it is not negligible. Therefore, the choice of the next hop is performed among a subset of the potential candidates, and the possibility of picking up routes different from the shortest one increases with a subsequent effect on latency. As SNR raises, this effect is mitigated and the average delay lowers as well. For low and medium values of transmission power, RoCoDi performs worse than RoDi. Cooperation increases the probability of delivering to the sink packets that incur a deep fade. However, such data flows experience a longer delay due to (successful) retransmissions and contribute to increase the average latency metric for RoCoDi. As far as RoCoDiLe is con-

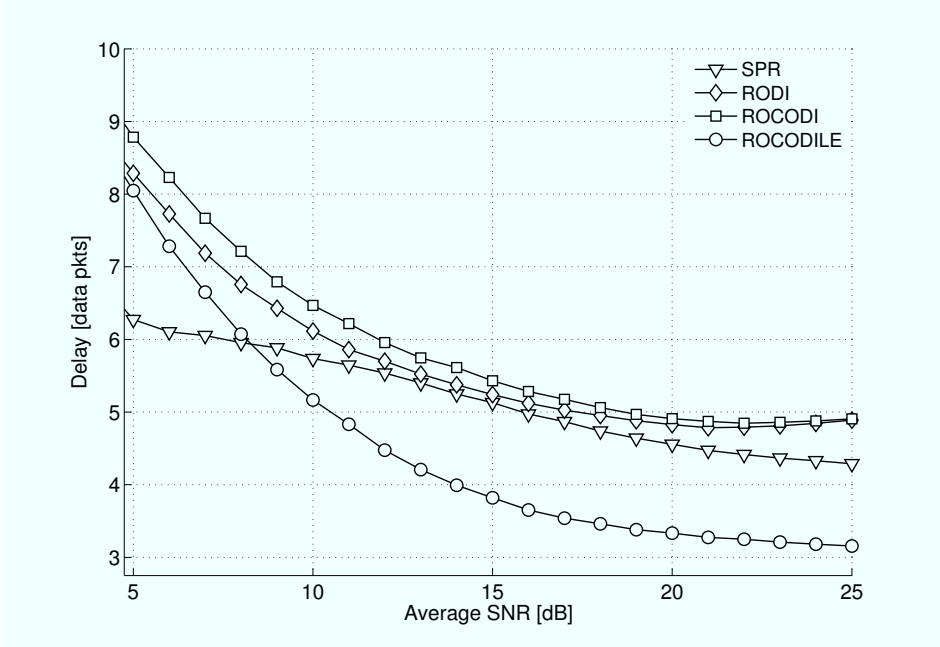


Figure A.2. *Delay vs SNR, grid topology.*

cerned, important advantages over the other geographic routing schemes are achieved in terms of average delivery time thanks to the opportunistic bypassing of some hops. The benefits induced by shorter paths overcome the latency introduced by the channel negotiation necessary to set up the leapfrogging procedure (i.e., LPFREQ and LPFACK) and RoCoDiLe is able to outperform even SPR for sufficiently high levels of transmission power. In the low-SNR region, the protocol performs slightly worse than SPR as leapfrogging is seldom triggered due to poor reception conditions.

Fig. A.3 depicts the transmission energy consumption in the network. We compare SPR and RoCoDiLe operating with the same transmission power (SNR = 15 dB). The x and y coordinates identify a node in the network topology, the source being located at $\{1,1\}$ and the destination being at $\{5,5\}$. The z coordinate represents battery consumption due to transmissions and, for the sake of clarity, it has been normalized to the energy required to perform a single transmission of a data packet. First of all, the plot highlights that the source experiences the highest consumption regardless of the protocol implemented. This is due to the fact that all the packets are generated there and therefore the node is involved in each communication. On the contrary, the sink shows a low power consumption as it only transmits control packets. Moreover, energy usage degrades along the main diagonal: the closer a node is to the sink, the lower the probability that it is reached by a frame being forwarded, as more hops need to be successfully performed. SPR only consumes the battery of nodes

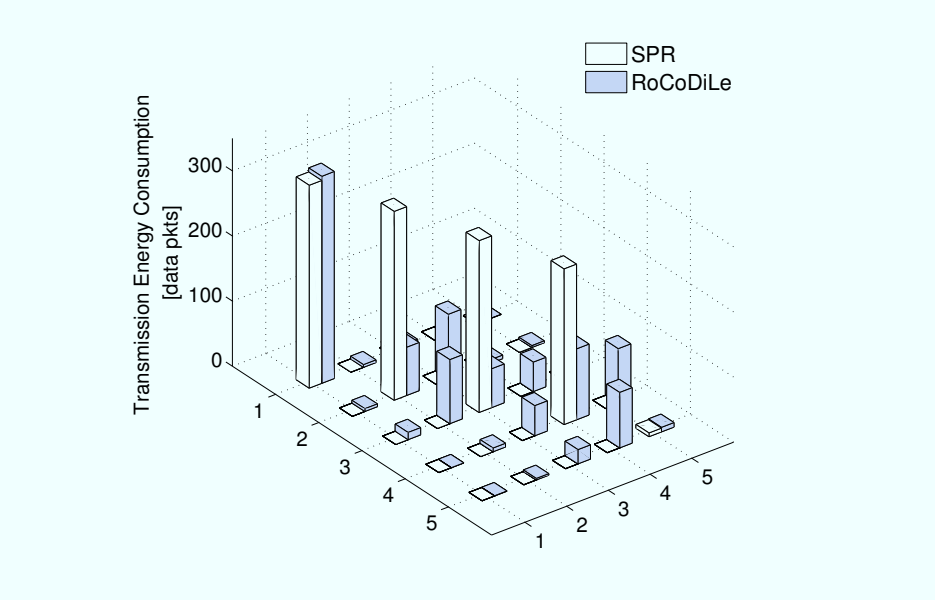


Figure A.3. *Transmission energy consumption, grid topology. Source node is located at {1,1} while destination node is at {5,5}.*

along the shortest path. On the other hand, the usage of geographic routing together with cooperative relaying and leapfrogging involves many more nodes in data flows. In this way energy consumption is much more distributed over the network and the average battery level is significantly higher for RoCoDiLe. In conclusion, not only the class of protocols that we propose is able to increase the efficiency of routing in terms of PDR and delay, but also network lifetime is significantly improved.

Finally, we consider the average overhead per data transmission induced by the protocols, depicted against average SNR in Fig. A.4. The metric is defined as the ratio of the total number of signaling packets sent in the network to the number of sent data packets (either as first transmissions, ARQ phases or relaying). The overhead for protocols that do not employ leapfrogging asymptotically tends to 3, as only the CSMA/CA negotiation packets are required (RTS, CTS, ACK). For low SNR, SPR presents a significant increase of the metric due to the higher number of ARQ phases undergone. The same reasoning explains why RoDi performs slightly worse than its cooperative version (RoCoDi). The higher overhead that characterizes RoCoDiLe is required in order to coordinate nodes that take part in the leapfrogging procedure and to avoid flow duplications and congestion in the network. For low transmission powers this mechanism is not often used and thus the additional signaling is limited. Leapfrogging is exploited more often in the mid-SNR region, as confirmed by both Fig. A.4 and Fig. A.1. Finally, for high values of transmission power, the overhead

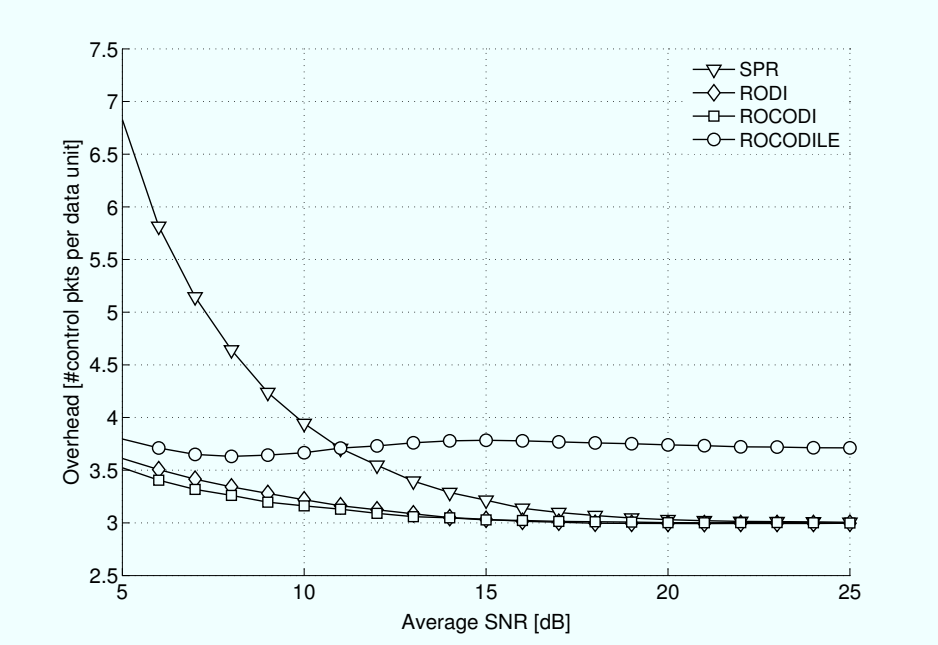


Figure A.4. *Average overhead vs SNR, grid topology.*

introduced by RoCoDiLe slightly lowers as it is counterbalanced by the reduction in the number of required retransmissions (resembling the trends for RoDi and RoCoDi).

A.2.3.2 Random Topologies

Fig. A.5 depicts the PDR vs average SNR for the considered protocols in random topology scenarios. Generally speaking, the trends presented for the grid topology are confirmed. However, for high values of the transmission power, the metric does not converge to 1. For SPR, this is due the greedy approach followed to select the next hop. As discussed, in our simulations we have considered only topologies where at least one path to the sink exists. However, this assumption does not guarantee that one such path can be found using maximum geographic advancement as the only criterion. As a result, SPR may not be able to deliver packets in some topologies, regardless of the quality of the link, and therefore the PDR is lowered. Geographic routing protocols significantly mitigate this effect by exploiting the distributed selection of the next hop, but still not all the packets are delivered successfully to the sink as some deadlock situations may occur (i.e., the dynamic selection of the next hop may pick up a node that does not have any connection to the final destination).

It is also interesting to observe that the curve of SPR decreases for extremely high values of SNR. This trend is due to the high transmission energy consumption that characterizes this activity region and that may lead some nodes to run out of battery within the simulation

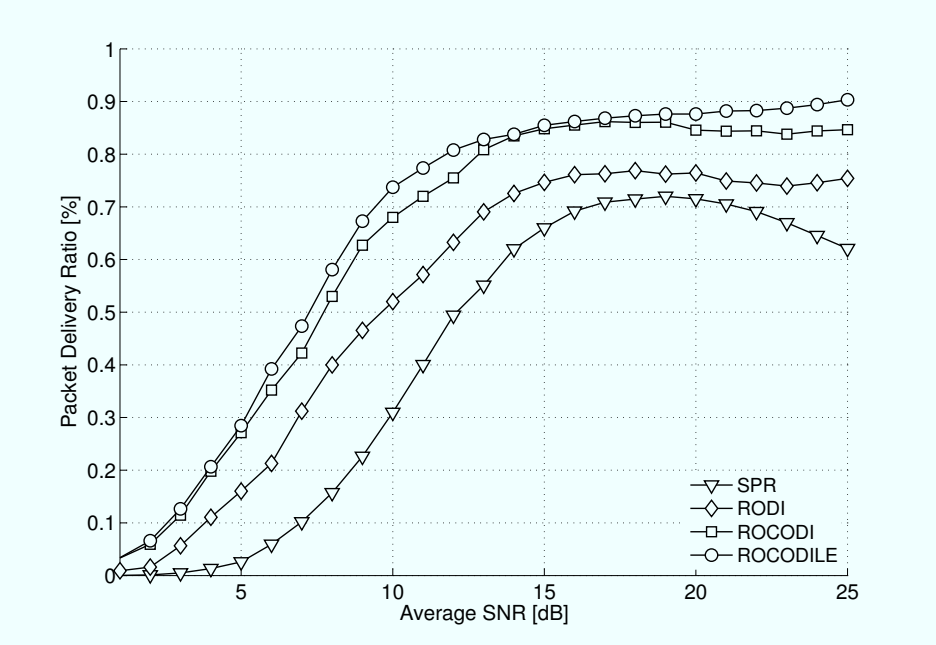


Figure A.5. *Packet Delivery Ratio vs SNR, random topology.*

interval. When such conditions occur, SPR is not able to reach the sink anymore, as only one route is known and exploited and the PDR drops. Such a trend does not affect the other protocols, the reason being twofold. On one hand, the better distribution of energy consumption in the network introduced by these schemes reduces the probability of node deaths. On the other hand, even if a node runs out of battery, the distributed selection of the next hop is able to dynamically switch to alternative paths, thus preserving high values for the delivery ratio.

Finally, we analyze the connectivity properties offered by the different protocols. We say that a network topology is connected using a specific routing scheme if at least one packet is delivered to the sink within the simulation interval. Fig. A.6 depicts the percentage of connected topologies in the low-SNR region. The metric is defined as the ratio of the number of connected scenarios using a given routing algorithm to the number of simulated scenarios.

Let us first consider SPR. For sufficiently large values of transmission power, the percentage of connected scenarios stabilizes to 80%. This shows that a greedy approach that considers geographic advancement as the only criterion for the next hop selection fails in determining a route in about one-fifth of the topologies, regardless of the link quality. On the other hand, for low-SNRs the number of non-connected scenarios significantly increases due to channel impairments that make it more difficult to reach the sink even if one greedy

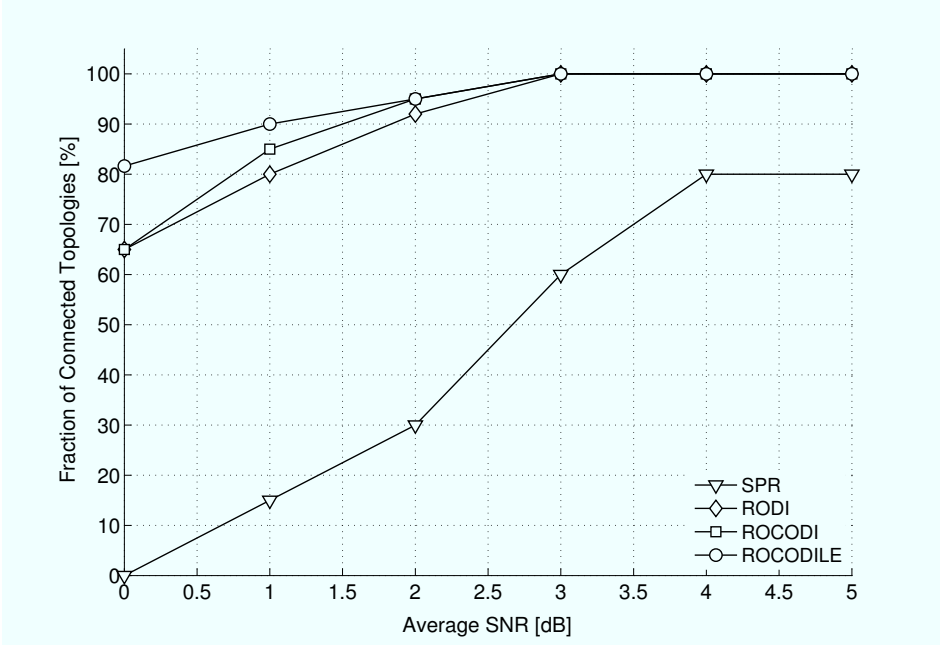


Figure A.6. *Connectivity, 25-node networks.*

path exists. As far as geographic routing protocols are concerned, for SNRs larger than 3 dB all the scenarios are connected. This shows that the distributed choice of the next hop significantly increases the reliability of routing by taking into account paths that differ from the shortest one. The robustness to channel impairments offered by cooperative relaying and leapfrogging further stresses the gains over SPR in the very low-SNR region, with improvements up to 60%.

A.3 Energy-Efficient Routing in WSNs with Mobile Sinks

In recent years, novel location-based services have created a growing interest in the design of wireless sensor networks with mobile sinks, i.e. mobile WSNs. In [44], for example, the authors consider a scenario in which multiple users equipped with mobile phones move through a sensor field and interact with a WSN by querying information of interest. In [84], a similar application has been described for a "store of the future", where customers receive local advertising and product information on their shopping-cart tablet via a WSN from a central server. Another example of a mobile WSN has been described for an intelligent transportation system in [85]. In this case, mobile sinks are represented by cars that collect updates from static sensors about traffic conditions and potential environmental dangers.

However, the presence of mobile sinks in a WSN may introduce frequent changes of the

network topology, which makes it difficult to establish and maintain a stable and reliable routing path in case of using static protocols. Whenever a route is broken due to sink mobility, route recovery messages have to be exchanged to build a new path. This approach significantly increases the overhead and thus degrades the network performance in terms of latency and energy consumption. These disadvantages, as discussed in the introduction of this appendix, can be alleviated by using a dynamic protocol such as geographic routing. Nonetheless, the capability of such schemes to avoid route construction comes at the expense of tracking the destination. To unleash the potential of the dynamic paradigm, then, it is fundamental to design an efficient strategy to continuously monitor the geographic position of the sink and distribute this information in the network.

Starting from these remarks, we have designed a novel geographic routing protocol for efficiently and reliably forwarding data packets from a static information source to a mobile sink through a multi-hop wireless sensor network. While the source and all sensor nodes are located at fixed and known positions, the mobile sink estimates and tracks its state (i.e. its location, velocity, and acceleration) from noisy measurements with a Kalman filter. To reliably and timely route data packets, the source predicts the location of the mobile sink. The state of the predictor is updated by receiving *STATE-UPDATE* messages from the mobile sink, which contain its current state. These messages are not periodically sent, but only if the Euclidean norm of the error between the predicted state and the state estimated by the Kalman filter exceeds a pre-defined threshold. The *STATE-UPDATE* messages as well as the *DATA* packets are forwarded in a multi-hop fashion through the wireless sensor network using geographic routing. Our simulation results demonstrate that the proposed scheme outperforms conventional routing protocols in terms of packet transfer reliability, latency, and energy efficiency.

We start our presentation by discussing some related works in the field of routing in mobile WSN in Section A.3.1 and by introducing the key ideas of our approach in Section A.3.2. Then, Section A.3.3 describes the method for estimating and tracking the state of the mobile sink, and its prediction at the source. Details on the proposed geographic protocol are given in Section A.3.4, while Section A.3.5 discusses the results of simulation campaigns performed to evaluate the effectiveness of the novel scheme against some benchmarks of interest.

A.3.1 Related Work

Several solutions for fast and reliable data delivery to mobile sinks in a WSN have been proposed in the literature. Most of them incorporate routing protocols that have originally been designed for static ad hoc networks and exploit the concepts of clustering and mobility prediction.

In [86], a two-tier data dissemination model for large-scale WSNs has been proposed. The main idea of this approach is to partition the WSN into clusters, each of them with a clusterhead that knows a route back to the source. Whenever a mobile sink enters the sensor field, a local flooding procedure is triggered to inform the clusterhead of its presence. This node then forwards the query of the sink to the source along the known route. Once the source has received the query, data packets can be transmitted to the mobile sink following again a two-tier approach: packets follow known and static routes up to the clusterhead that has in turn the role to forward them to the sink coping with its mobility. The advantage of this approach over classical AODV and DSR [78] routing is that the flooding of messages is kept local and only quick adjustments to the path have to be made as long as the sink moves within the cluster. However, it should be noted that flooding is not completely avoided and still has a severe impact on the network performance if the sink moves from one cluster to another.

In [85], a simple yet effective routing solution has been proposed that enhances the two-tier dissemination protocol of [86] by incorporating mobility prediction. The authors consider a WSN deployed along a street to gather environmental data and route the collected information to cars passing by. The cars (i.e. the mobile sinks) are equipped with a GPS-based positioning system and are thus aware of their location, speed, and mobility pattern. When a car enters the WSN, it sends out an information query to a sensor next to the road, containing its speed and trajectory, and indicating the region from where it wants to receive the requested environmental information. When a sensor receives the query, the request is forwarded by means of a geographic routing protocol to the sensor node that is closest to the center of the indicated region. This node takes on the role of a clusterhead and initiates gathering of local environmental information. Once the procedure is accomplished, the clusterhead aggregates the collected data and forwards the generated packet to the car at the predicted location. The proposed scheme is tailored to a specific application that is relevant for the automobile industry. However, its improvements cannot easily be achieved in an indoor environment because GPS cannot be applied.

In [87], the authors propose to enhance the performance of a vehicular ad-hoc net-

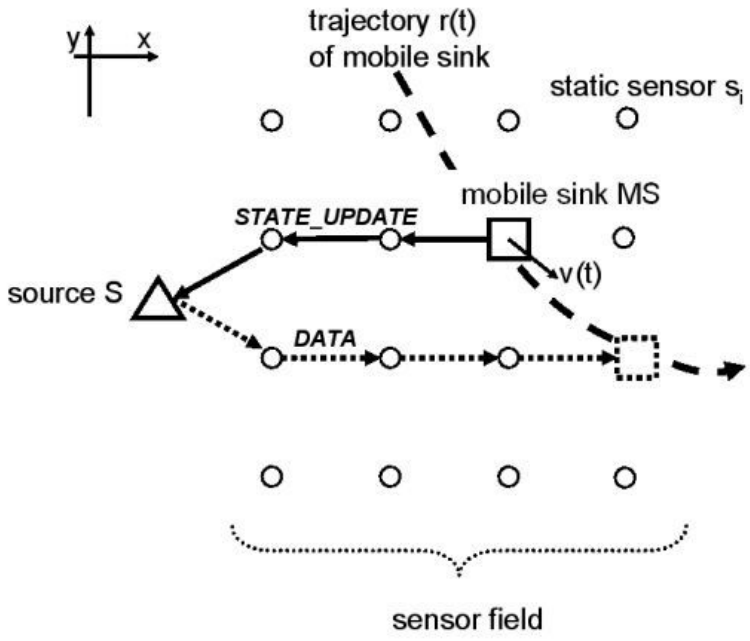


Figure A.7. *Mobile sink moving in a static sensor field*

work by embedding an adaptive beaconing strategy into the geographic routing protocol GPSR [88]. To reduce the overhead in the network, the frequency of beacon transmissions containing the geographic location of the mobile sinks is adapted to the node mobility and the network traffic. In [87], however, no method is described how routing can be efficiently combined with a location tracking scheme. On the contrary, in [84] a method is described for efficiently sensing and tracking slowly moving objects in an indoor environment but no information is provided on how data can be routed to them.

A.3.2 Geographic Routing in Mobile WSNs

In order to introduce the key ideas of our proposal, let us consider the scenario depicted in Fig. A.7, composed by a set of static wireless sensor nodes $\{s_i\}$ deployed at known geographic positions, a mobile sink MS that moves with a time-varying speed $v(t)$ along an unknown trajectory $r(t)$ through the sensor field, and a source $\boldsymbol{S}$ that is located at a known position. The source has one or several *DATA* packets to transmit to the MS. In general, such frames cannot be directly transmitted to the destination over a single radio link because the MS is often too far away; therefore, multi-hop transmissions have to be applied.

An efficient solution for forwarding a *DATA* packet hop-by-hop from the source to the mobile sink can be obtained by applying geographic routing. As discussed in Section A.1, this approach does not establish a static path from $\boldsymbol{S}$ to MS before transmitting data, but

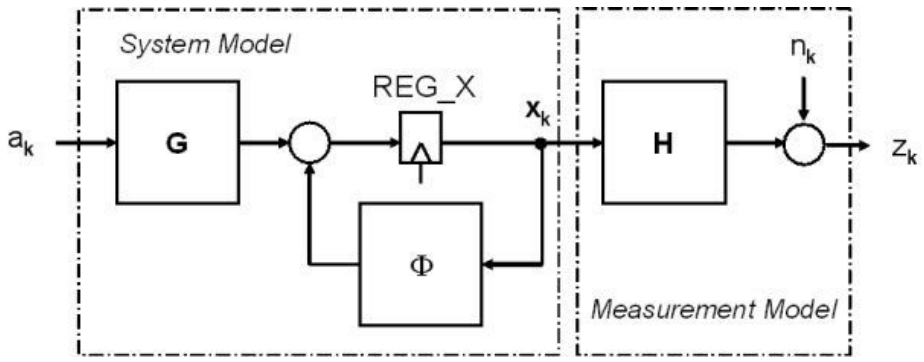


Figure A.8. *Discrete-time system and measurement model*

it rather dynamically establishes a route while forwarding the information. In particular, when a data packet for the sink is generated, the source selects the node among its neighbors which offers the maximum geographic advancement towards the MS. Once the next hop is chosen, the *DATA* packet is forwarded to the selected node that, in turn, iterates the procedure until the sink is reached.

Geographic routing can only be applied if all sensor nodes are aware of their geographic location as well as the position of the sink. However, while in the scenario under analysis the former condition can easily be met, the mobility of the sink creates the need for a mechanism to inform $\boldsymbol{S}$ about the current position of the destination. In order to achieve this, two tasks have to be accomplished: on the one hand, a mobile node has to continuously monitor its position, while on the other hand this information has to be distributed in the network. To achieve the latter goal, the MS has to transmit *STATE-UPDATE* messages to the source that contain information on its location. In this work, we propose a novel protocol that minimizes the number of *STATE-UPDATE* transmissions and exploits mobility prediction.

The next two sections describe in further details our approach to effectively fulfill the two goals.

A.3.3 State Estimation and Mobility Prediction Algorithms

The mobile sink estimates its geographic location with a Kalman filter that monitors noisy position measurements obtained with sensor devices such as accelerometers, gyroscopes, compass, etc. [84]. The design of the Kalman filter is based on a state-space model that takes into account the generation process of the acceleration driving the MS, its velocity, and position. In principle, the generation process can be modeled in different ways considering various mobility patterns of the MS and characteristics of the measurement device. To illustrate the basic idea of the proposed method, we define a simple state-space model that

describes a random movement of the MS along one axis of the coordinate system and ignores imperfections of the measurement device. The extension to a more complex, but more accurate model and to a multi-dimensional coordinate system is straightforward.

Let us assume that the mobile sink moves along a trajectory $r(t)$ that is a straight line parallel to one axis of the coordinate system. Initially, the MS is stationary at position $r(0) = 0$. For time $t \geq 0$, the MS is driven by a random acceleration $a(t)$ that is modeled by a first-order Markov process

$$\dot{a}(t) = -\alpha a(t) + w(t) \qquad \alpha \geq 0, \tag{A.4}$$

where $w(t)$ represents white noise with variance $\sigma_w^2 = 2\alpha\sigma_a^2$ and α is reciprocal to the time maneuver constant of the mobile sink. The movement of the MS can thus be modeled by setting the parameters α and σ_a^2. Typical values can be obtained from the Singer acceleration model [89]. The position $r(t)$ of the MS relates to its acceleration $a(t)$ according to the Newton law

$$\ddot{r}(t) = \dot{v}(t) = a(t) \ , \tag{A.5}$$

where $v(t)$ denotes the MS velocity.

The linear differential equations (A.4) and (A.5) represent the continuous-time *system model* of the investigated system and can be written in the state-space form

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}w(t) \ , \tag{A.6}$$

where the state vector $\mathbf{x}(t)$ and the matrices $\mathbf{F}$ and $\mathbf{G}$ are defined as:

$$\mathbf{x}(t) = \begin{bmatrix} r(t) \\ v(t) \\ a(t) \end{bmatrix}, \ \mathbf{F} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{A.7}$$

The corresponding discrete-time *system model* is given by

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{u}_k \, , \tag{A.8}$$

as shown in Fig. A.8, where the elements of the state vector $\mathbf{x}_k^T = [r_k \ v_k \ a_k]$ represent samples of the position, velocity, and acceleration of the MS taken at time kT. The matrix $\mathbf{\Phi}$ defining the transition between two successive states is given by

$$\mathbf{\Phi} = \begin{bmatrix} 1 & T & \frac{1}{\alpha^2}[-1 + \alpha T + e^{-\alpha T}] \\ 0 & 1 & \frac{1}{\alpha}[1 - e^{-\alpha T}] \\ 0 & 0 & e^{-\alpha T} \end{bmatrix}, \tag{A.9}$$

while $\mathbf{u}_k$ is a discrete-time white noise vector with covariance matrix

$$\mathbf{Q} = E\{\mathbf{u}_k {\mathbf{u}_k}^T\} = 2\alpha\sigma_a^2 \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix}, \tag{A.10}$$

whose coefficients are reported at the bottom of the page.

Moreover, we assume that the location position r_k embedded in additive white noise can be measured at the output of the system. The discrete-time *measurement model* is thus given by

$$z_k = \mathbf{H}\mathbf{x}_k + n_k \tag{A.11}$$

as illustrated in Fig. A.8, where the matrix is given by $\mathbf{H} = [1\ 0\ 0]$ and n_k is additive white noise with variance σ_n^2.

Based on the discrete-time state-space model given by (A.8) and (A.11), the optimal Kalman filter can be derived [90]. As shown in Fig. A.9, this filter monitors the noisy measurements z_k at the MS and computes in real-time a state estimate $\hat{\mathbf{x}}_k^T = [\hat{r}_k\ \hat{v}_k\ \hat{a}_k]$ of the MS that comprises the position estimate $\hat{r}_k$, velocity estimate $\hat{v}_k$, and acceleration estimate $\hat{a}_k$. An iterative computation of $\hat{\mathbf{x}}_k^T$ is achieved by firstly predicting the state of the MS based on the equation

$$\hat{\mathbf{x}}_k^- = \mathbf{\Phi}\,\hat{\mathbf{x}}_{k-1}. \tag{A.12}$$

The predicted state $\hat{\mathbf{x}}_k^-$ is thus given by linearly projecting the state estimate $\hat{\mathbf{x}}_{k-1}$ from time $(k-1)T$ one time interval ahead using the transition matrix $\mathbf{\Phi}$. Afterwards, the state estimate error covariance matrix $\mathbf{P}_k^-$ for the current predicted state is calculated using the updated state estimate error covariance matrix $\mathbf{P}_{k-1}$ of the previous state estimate $\hat{\mathbf{x}}_{k-1}$:

$$\mathbf{P}_k^- = \mathbf{\Phi}\mathbf{P}_{k-1}\mathbf{\Phi}^T + \mathbf{Q}. \tag{A.13}$$

On receiving a location position measurement z_k from the measurement unit at time kT, the Kalman filter updates its state estimate based on the equation

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K_k}(z_k - \mathbf{H}\,\hat{\mathbf{x}}_k^-). \tag{A.14}$$

$$q_{11} = \frac{1}{2\alpha^5}\left[1 - e^{-2\alpha T} + 2\alpha T + \frac{2\alpha^3 T^3}{3} - 2\alpha^2 T^2 - 4\alpha T e^{-\alpha T}\right] \quad q_{22} = \frac{1}{2\alpha^3}\left[4e^{-\alpha T} - 3 - e^{-2\alpha T} + 2\alpha T\right]$$

$$q_{12} = \frac{1}{2\alpha^4}\left[e^{-2\alpha T} + 1 - 2e^{-\alpha T} + 2\alpha T e^{-\alpha T} - 2\alpha T + \alpha^2 T^2\right] \quad q_{23} = \frac{1}{2\alpha^2}\left[e^{-2\alpha T} + 1 - 2e^{-\alpha T}\right]$$

$$q_{13} = \frac{1}{2\alpha^3}\left[1 - e^{-2\alpha T} - 2\alpha T e^{-\alpha T}\right] \quad q_{33} = \frac{1}{2\alpha}\left[1 - e^{-2\alpha T}\right].$$

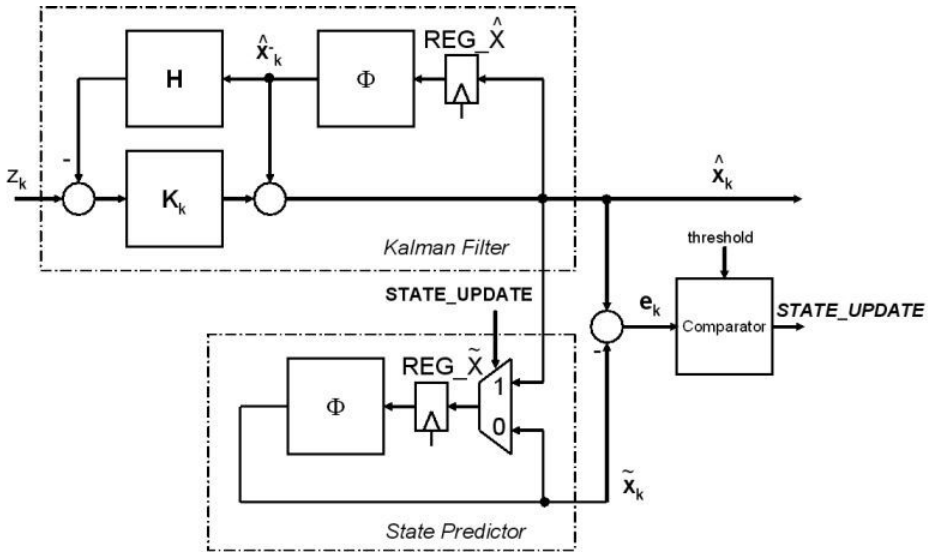


Figure A.9. *Kalman filter and state predictor at mobile sink*

The state estimate $\hat{\mathbf{x}}_k$ is thus calculated by adding to the predicted state $\hat{\mathbf{x}}_k^-$ the error between the location position measurement z_k and the reconstructed output parameter weighted by the Kalman gain vector

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \sigma_n^2)^{-1} . \tag{A.15}$$

To obtain the optimum Kalman filter gain settings, the covariance matrix of the state vector estimation error has to be computed according to

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}\mathbf{P}_k^- . \tag{A.16}$$

The state estimate $\hat{\mathbf{x}}_k$ provided by the Kalman filter is transmitted as parameter of the *STATE-UPDATE* message from the mobile sink via the sensor field to the source.

The source can thus predict the state $\tilde{\mathbf{x}}_k$ of the MS according to:

$$\tilde{\mathbf{x}}_k = \begin{cases} \hat{\mathbf{x}}_k & \text{if } \textit{STATE-UPDATE} \text{ message has} \\ & \text{been received at time } kT, \\ \boldsymbol{\Phi}\, \tilde{\mathbf{x}}_{k-1} & \text{otherwise.} \end{cases} \tag{A.17}$$

On receiving a *STATE-UPDATE* message at time kT, the source copies the state $\hat{\mathbf{x}}_k$ contained in the message into the register of the state predictor. When no message is received, the source predicts the state of the MS by linearly projecting the predicted state $\tilde{\mathbf{x}}_{k-1}$ one time interval T ahead using the system matrix $\boldsymbol{\Phi}$. Note that this prediction strategy differs from the one incorporated in the classical Kalman filter. While in the Kalman filter new estimates are obtained immediately after receiving a new measurement z_k, the state predictor only

takes into account new measurements after receiving an update request from the mobile sink.

To reduce the number of control message exchanges over the WSN, the mobile sink issues a *STATE-UPDATE* request only if the predicted value $\tilde{\mathbf{x}}_k$ deviates too far from the state estimate $\hat{\mathbf{x}}_k$ provided by the Kalman filter. To measure this deviation, a copy of the state predictor has also to be implemented at the MS as shown in Fig. A.9. The error vector $\mathbf{e}_k$ between the state estimate $\hat{\mathbf{x}}_k$ provided at the output of the Kalman filter and the predicted estimate $\tilde{\mathbf{x}}_k$ provided at the output of the state predictor is computed. A comparator calculates the Euclidean norm of the error vector and compares it to a threshold ε_{Th}. If the norm exceeds ε_{Th}, the comparator triggers an update of the state of the predictors at the mobile sink and at the source. At the mobile sink, the state estimate $\hat{\mathbf{x}}_k$ is directly loaded via the multiplexer into the register of the state predictor. The state of the predictor at the source is updated by transmitting the protocol message *STATE-UPDATE* with the state estimate $\hat{\mathbf{x}}_k$ as parameter.

A.3.4 Mobility Prediction Routing Protocol

To effectively implement the routing approach described in Sections A.3.2 and A.3.3, some coordination among nodes is required. In particular, methods to bi-directionally transfer information between a mobile sink and a static source through a WSN have to be provided. To this aim, we present a novel carrier-sense based cross-layer MAC and routing protocol called Mobility Prediction Routing (MPR).

Let us refer to the scenario of Fig. A.7 and let us start by considering the situation of a sink that has a notification to send to the source $\boldsymbol{S}$, either because it has just joined the network or because the deviation between predicted and Kalman-filter estimated state has exceeded a threshold. In this case, the mobile station transmits a *STATE-UPDATE* message containing the current estimate of its state following the standard 802.11 CSMA mechanism based on Binary Exponential Backoff (BEB) [5]. Nodes in $\{s_i\}$ that successfully decode this packet enter a distributed contention procedure to select the node that has to forward the information. To this aim, each candidate evaluates the advancement it offers towards the source[3] and picks an integer m inversely proportional to the computed value in the set $\{1, 2, \ldots, \text{CW}\}$, where CW is a protocol parameter that specifies the contention window for the selection of a next hop. The node, then, starts a backoff of duration m slots and senses

[3]We remark that since the source is assumed to be static and location aware, its geographical position can be easily distributed to other nodes in the WSN, e.g., by means of a broadcast procedure, after network deployment.

the aggregate power level on the medium $\bar{P}$. If $\bar{P}$ exceeds the carrier sense threshold Λ_{CS}, the node assumes that another terminal has won the contention and goes back to its activity. On the contrary, if the backoff interval expires and $\bar{P}$ remains below Λ_{CS}, the candidate is selected as data forwarder and replies to the sink with an acknowledgment (ACK) message.[4]

Once this first hop has been accomplished, basic geographic routing can be performed within the static WSN in order to deliver data to the source. The terminal in charge to forward the *STATE-UPDATE* message, after a carrier sense based BEB, transmits the packet to its neighbors. Nodes that decode the frame perform the distributed contention described earlier to identify the next hop which is closest to the source, which in turn replies with an ACK and iterates the procedure until the the final addressee of the message is reached.

Finally, upon receiving a *STATE-UPDATE* message from a sink, the source can start processing packets addressed to it. In particular, when a data frame has to be delivered to a mobile node, $\boldsymbol{S}$ applies the algorithm described in Section A.3.3 to predict the current location of the addressee. This information is then included in the packet and the aforementioned geographic routing procedures take place to deliver data to the sink through the static WSN.

A.3.5 Simulation Results

In order to test the effectiveness of the proposed solution, extensive Omnet++ simulations have been performed. In our studies, we have considered wireless sensor networks composed by 36 static nodes spread in a 6×6 regular grid over a 100×100 m^2 area and by one or more sinks moving within the region. The motion of the nodes has been modeled as discussed in Section A.3.3. In particular, the reciprocal of the maneuver time, α, has been set to 30 Hz. Such a value is suitable to describe the movement of both pedestrians and human-driven vehicles, and is thus of interest for many potential applications. Poisson traffic addressed to the mobile sinks with intensity $\lambda = 5$ pk/s/sink is injected in the network by a single source, located at the upper-left corner of the grid.

The Mobility Prediction Routing protocol has been compared to two benchmark schemes: AODV and basic geographic routing (Geo). In the latter approach, mobile sinks periodically transmit beacons containing an estimate of their current location. Let us remark that while *STATE-UPDATE* messages in MPR contain the Kalman filter estimate $\hat{\mathbf{x}}_k$, in Geo the raw noisy measurement of the current position z_k are included in beacons. Such frames are for-

[4] If the source is able to decode the *STATE-UPDATE*, it immediately replies with an ACK, i.e., $n = 0$, in order to avoid routing procedures that would be useless and expensive in terms of energy consumption.

warded towards the source according to the procedure described in Section A.3.4. In turn, whenever a new data packet has to be delivered, the source triggers the plain geographic routing algorithm assuming the addressee to be located at the position indicated in the last received beacon, i.e., no prediction is performed.

All the schemes rely on a carrier sense based medium access policy with BEB. In order to have a fair comparison, we have tuned the Short Retry Limit (SRL) so as to achieve a similar reliability for the different protocols under static conditions (i.e., when sinks are not moving). In particular, the SRL has been set to 5 for AODV, whereas 4 attempts were sufficient for basic geographic routing and MPR, due to the shorter average length of the routes packets have to undergo to reach a sink.

The wireless environment is subject to correlated Rayleigh fading and path loss with coefficient $\beta = 3$. The standard set of network and protocol parameters used in our studies are reported in Tab. A.1.

MPR and its competitors have been studied in two different scenarios. In the former, a single sink was present in the network, and the behavior of the protocols has been analyzed under different mobility conditions. This was accomplished by varying the parameter σ_a (see Section A.3.3), which represents the standard deviation of the acceleration of a sink, and hence gives an indication of how quickly a node can vary its speed and thus its trajectory. This investigation is of particular interest since it sheds light on the robustness of the considered routing policies to user mobility and tests their adequacy to different applications. Furthermore, the protocols have been assessed in networks with multiple sinks sharing the same mobility parameters. Such a scenario is apt to evaluate how the schemes respond to higher traffic loads and how they behave under harsh channel contention conditions.

All the results presented in this paper have been obtained by averaging the outcome of 50 independent simulations, each 1000s long, which provided the desired statistical confidence.

A.3.5.1 Networks with a Single Mobile Sink

Let us start our analysis by considering the response of the protocols to different mobility conditions when a single sink is present. The first performance criterion that we assess is the end-to-end Packet Delivery Ratio (PDR), defined as the ratio of the number of data packets injected in the network to the number of frames successfully received at the intended addressee. The metric is depicted against the standard deviation of the sink acceleration σ_a in Fig. A.10. As expected, when the mobile user is static or quasi-static (e.g., for low accelerations), all the considered schemes provide a good reliability of about 95%. On the

Table A.1. *Parameters used in our simulations*

Transmission power	3 dBm
Noise floor	-90 dBm
Detection threshold	-81 dBm
Path loss exponent, β	3
Carrier frequency	2.4 GHz
Bitrate	250 kbps
Reciprocal of maneuver time, α	30 Hz
Measurement noise variance, σ_n^2	2 m^2/s^2
Error detection threshold, ε_{Th}	5
Nominal load per sink, λ	5 pk/s/sink
MAC buffer size	16
BEB Initial contention window	32 slots
Short Retry Limit - AODV, Geo, MPR	5, 4, 4
Slot, DIFS, SIFS duration	20, 128, 28 μ s
Carrier Sense threshold, Λ_{CS}	-86 dBm
Contention Window for next hop selection, CW	16 slots
Payload	512 bit
Header, ACK, Control Packets	42, 96, 112 bit

contrary, higher values of σ_a induce a performance loss for the protocols. As for AODV, this can be explained considering the static routing approach it relies on. According to this, a path to the destination is composed by a set of forwarding nodes identified by means of a broadcast-based route discovery procedure. However, as soon as the sink moves out of the communication range of the last node in the identified sequence, the whole route collapses, and data cannot be delivered until a new path to the destination is established through route recovery. Such a condition occurs more often as the value of σ_a increases, with detrimental effects on the overall reliability (up to 10% loss for high acceleration). This drawback can be partially mitigated by applying geographic routing. Indeed, such a strategy can be robust to user mobility provided that sinks send out beacons with position updates with sufficient high frequency.

In particular, Fig. A.10 highlights that, for the topologies under study, a beacon interval of 1 second leads to a constant reliability for pedestrian or vehicle movement up to moder-

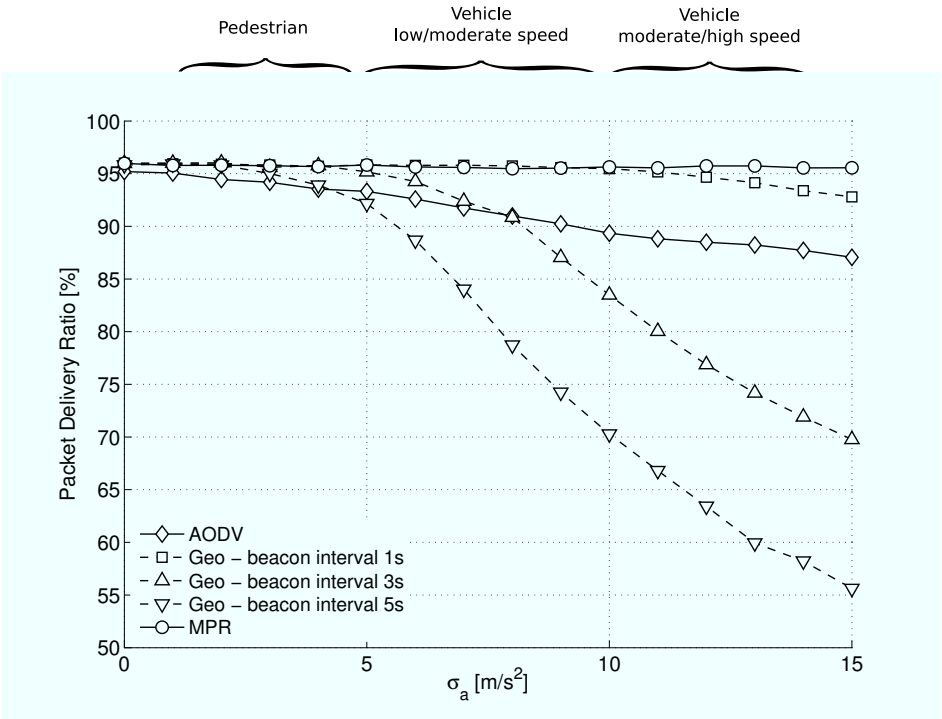


Figure A.10. *Packet delivery ratio as a function of the standard deviation of the sink acceleration σ_a.*

ate speeds, whereas the performance starts to slightly degrade only for higher values of σ_a. The improvement offered by geographic routing stems from the capability of the beaconing mechanism to provide the source with an estimate of the current position of the sink. This knowledge is exploited to track the destination and, unlike in AODV, to pre-emptively adapt the path frames have to undergo, thus reducing the number of packet losses. However, such a strategy is extremely sensitive to the accuracy of the location estimate available at the source. In fact, as soon as this information does not reflect the actual position of the addressee, either due to the loss of some updates or because the beaconing frequency is not high enough with respect to the mobility pattern of the sink, data deliveries fail, since packets are forwarded using an inappropriate route. This is clearly shown in Fig. A.10, where the curves for geographic routing with beaconing interval of 3 and 5 seconds plummet as σ_a increases. From these observations we can infer that the robustness against sink mobility that can be achieved by means of basic geographic routing comes at the expense of a high beaconing frequency, which in turn increases overhead and energy consumption in the network, as will be discussed later. Let us now focus, instead, on the MPR protocol. The plot highlights that our scheme offers high PDR for all the considered values of σ_a and is thus able to overcome the issues that affect basic geographic routing. This capability stems from two aspects. First of all, as discussed in Section A.3.3, MPR adapts the beaconing in-

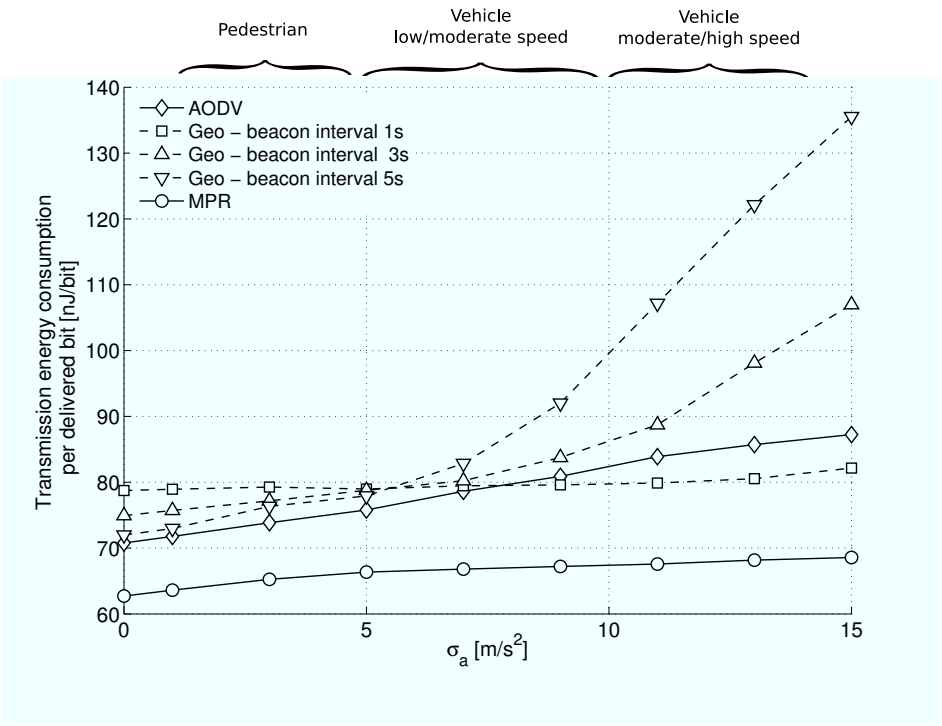


Figure A.11. *Transmission energy consumption per successfully delivered data bit as a function of the standard deviation of the sink acceleration, σ_a.*

terval to the mobility pattern of the sink. Therefore, the source always has at its disposal an up-to-date estimate of the destination's position and can reliably transfer data even for high values of σ_a. Secondly, with the proposed strategy, packets for the sink are routed towards the expected current location of the addressee obtained by means of the prediction algorithm rather than towards the position contained in the last received *STATE-UPDATE*, thus handling potential movements of the sink between successive beacons.

The second performance criterion that we consider is transmission energy consumption, which is of key importance in battery constrained WSNs. The metric is defined as the ratio of the transmission energy consumed by all the nodes in the network to the number of successfully received information bits at the sink, and is depicted against the standard deviation of the acceleration in Fig. A.11. Let us start by focusing on the behavior of the different protocols for low values of σ_a, i.e., when all the schemes offer the same level of reliability, as shown in Fig. A.10. As a first remark, we notice that AODV consumes less energy than basic geographic routing. In such conditions, sink mobility seldom induces failures in the paths established by AODV, and therefore the overhead related to route construction and route recovery is limited. On the contrary, the periodic beaconing approach followed by the geographic routing solution enforces the network to process control messages even

when not needed because the sink is not moving, and thus yields a higher energy consumption. Remarkable energy savings over both the strategies can be achieved by using MPR: as proven by Fig. A.11, our protocol offers improvements of about 15% over AODV and of up to 30% with respect to basic geographic routing. Not only is the the proposed solution able to avoid the overhead due to path construction and recovery procedures that affects static routing algorithms, but it also saves energy by triggering *STATE-UPDATE* messages only when needed in order to preserve reliability. The beneficial effects of MPR become ever more pronounced for higher values of σ_a. Indeed, faster and more dynamic mobility patterns of the sink negatively affect both AODV and geographic routing: the former approach has to face more frequent route failures with a subsequent increase in the control traffic overhead, while the latter schemes undergo a steady drop in reliability and thus tend to spend more energy to successfully deliver a packet to the sink. Instead, MPR exhibits just a slight raise in energy consumption even in such harsh conditions for routing protocols. This stems once again from the capability of the proposed solution to minimize the traffic required to reliably track the sink by matching the frequency of *STATE-UPDATE* transmissions to the actual movement pattern of the mobile station.

In conclusion, by jointly considering the results of Figs. A.10 and A.11, we can infer that the MPR protocol is able to offer both higher reliability and significantly reduce energy consumption, thus prolonging network lifetime, with respect to other conventional routing solutions for mobile WSNs, and thus appears to be a suitable solution for a large number of application scenarios.

A.3.5.2 Networks with Multiple Mobile Sinks

We have also tested the behavior of MPR and its competitors in networks with multiple sinks sharing the same mobility parameters. In this study, the standard deviation σ_a of the acceleration has been set to 4 m/s^2. Such a value describes a pedestrian movement, and is thus of interest for many potential applications, e.g., [84].

The first metric that we consider is the average end-to-end packet delivery ratio, depicted against the number of mobile sinks in the network in Fig. A.12. As a general remark, we notice that all the schemes incur a performance loss when more destinations have to be served. Indeed, not only does a larger number of sinks involve a higher network load in terms of data packets that have to be delivered, but it also augments the control traffic needed to establish and keep a connection with the mobile nodes. In turn, this traffic growth leads to a harsher channel contention and to an increased interference level, thus

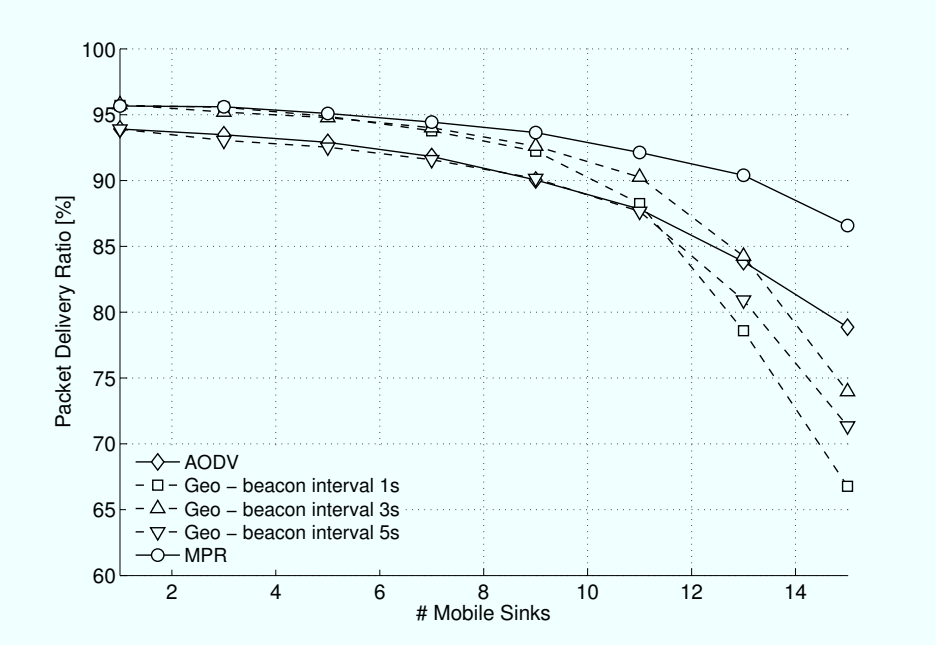


Figure A.12. *Packet delivery ratio as a function of the number of mobile sinks in the network. σ_a is set to 4 m/s^2 (pedestrian movement) for all the sinks.*

worsening the link reliability and the end-to-end PDR. It is also interesting to observe how this effect is more pronounced for basic geographic routing than for AODV. This can be explained considering the low sink mobility in our simulations. As discussed earlier, in these conditions AODV does not experience frequent route failures. Therefore, once a path towards a destination has been established (via a broadcast-based route construction procedure), further control traffic is not likely to be triggered often. On the contrary, the beaconing approach followed by geographic routing continuously pours control messages in the network even when not necessary, e.g., because a sink is standing still, thus leading to a quicker traffic congestion. Clearly, the higher the beacon frequency, the more detrimental this effect. As opposed to its competitors, MPR exhibits a more graceful degradation of the network performance, as shown in Fig. A.12, where our protocol achieves a PDR higher than 90% even when more than 10 sinks are present and outperforms AODV by almost 10% and basic geographic routing by as much as 20%. This remarkable result is once again due to the ability of the proposed solution to both avoid route construction procedures and minimize the number of *STATE-UPDATE* messages generated by each mobile node, thus significantly reducing the congestion in the network.

Another metric of interest is represented by the average end-to-end latency per successfully delivered data packet, depicted as a function of the number of mobile sinks in the

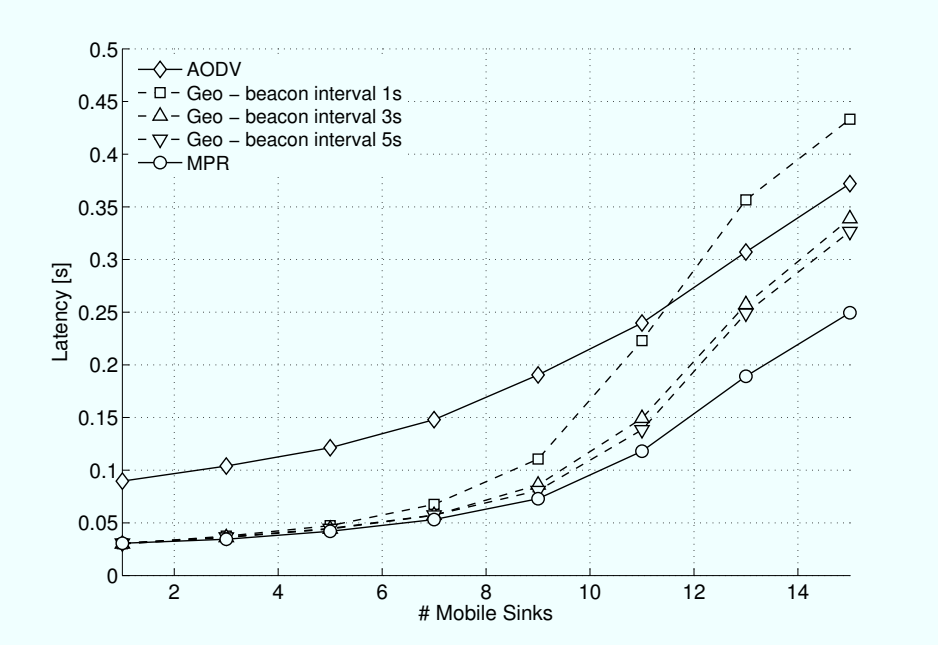


Figure A.13. *Average delivery latency as a function of the number of mobile sinks in the network. σ_a is set to 4 m/s^2 (pedestrian movement) for all the sinks.*

network in Fig. A.13. When a single destination has to be served, geographic routing based approaches, i.e., Geo and MPR, are able to almost halve the latency with respect to AODV. This stems from the capability of these schemes to exploit location information to forward data without resorting to the broadcast-based and time consuming route construction procedures that are instead needed by static routing protocols. On the other hand, the higher level of traffic triggered by the presence of multiple sinks in the network increases the delivery delay for all the schemes, due to both the longer average backoffs required to access the channel and to longer queuing time spent by packets at each hop. Nevertheless, Fig. A.13 emphasizes that MPR is able to contain this performance loss, outperforming AODV by more than 50% (avoidance of route construction procedures) and basic geographic routing by up to 80% when 15 destinations join the network (minimization of beaconing traffic).

In conclusion, the MPR protocol does not only offer high reliability even when the network has to support a high level of traffic, but also it is able to deliver information to multiple mobile sinks much quicker than its competitors, which makes it particularly suitable for all the applications that have tight constraints in terms of latency.

A.4 Dynamic Tunnel Routing

Many applications of WSNs, such as those in the military or surveillance domain, require nodes to be spread and operate in harsh or even hostile environments. In this context, reliable and resilient packet forwarding techniques are of capital importance in order to securely deliver the sensed data hop-by-hop from a remote sensor to the intended destination. In this perspective, indeed, not only do routing schemes have to cope with the impairments that characterize the wireless medium, but also they have to face threats posed by nodes behaving maliciously. For instance, adversaries that aim at stealing sensible information or at hindering proper data delivery may try to overhear data exchanges that take place among sensor nodes or to introduce in the network fake or altered messages (*outsider attacks*). Moreover, in a more subtle way, malicious wireless devices may join the routing path so as to drop or selectively forward the packets they are supposed to propagate (*selective forwarding attack*).[5]

In light of this discussion, recent research activity has focused on providing both static and dynamic routing for WSNs (see Section A.1) with security countermeasures. More specifically, static routing solutions can be protected against outsider security attacks by employing the Localized Encryption and Authentication Protocol (LEAP) [92]. This protocol has been designed to enable in-network processing, while restricting the impact of compromised nodes to their neighborhood. LEAP supports the establishment of several types of symmetric encryption keys on each sensor node, among them a cluster key and a pairwise link key. A cluster key is shared among neighbor nodes and is applied to encrypt the control messages exchanged during route establishment. On the other hand, during data forwarding a pair-wise link key protects information exchange between the nodes. While per-link encryption provides excellent protection against outsider attacks, the risk of breaking the path to the destination by compromising a single node with a selective forwarding attack is high. Therefore, to defend against security attacks of this kind, the use of static multiple-path routing protocols has been proposed. These schemes establish several paths to a destination in order to provide redundant and thus more reliable data forwarding. For example, the algorithm proposed in [93] ensures that most of the data packets are forwarded through the most efficient paths to the destination, while a small fraction of the packets are spread out to increase the chance that at least one of them reaches the destination. However, robustness against data forwarding attacks comes at the expense of an increased energy consumption, and thus a reduced lifetime for WSNs.

[5]For a comprehensive discussion on security issues that may affect WSNs please refer to [91].

As far as dynamic routing is concerned, these protocols, e.g., geographic routing [80] or RoCoDiLe (Section A.2), can also be protected against outsider security attacks by using LEAP. However, in case of using RoCoDiLe, the cluster key shared among neighbor nodes has to be used not only during the exchange of control messages but also when transmitting data packets. Since per-link encryption cannot be provided, the data protection against outsider attacks is worse as in case of using static routing or GeRaF. On the other hand, it is interesting to notice that dynamic routing solutions are more robust against selective-forwarding attacks, as compromised nodes, once detected and identified, can be bypassed without the need to initiate a novel route construction procedure by simply ignoring them during the data transfer process.

The dynamic paradigm, then, appears in general more appealing even in the perspective of network security. Nonetheless, as pointed out several times in this appendix, such solutions are typically designed assuming location information to be available at sensor nodes, e.g., by means of a GPS system. This assumption, though, may not hold in practice and severely limits the applicability of, and therefore also the benefits entailed by, dynamic routing. Starting from these remarks, in this section we propose a novel strategy called *tunnel routing* that enables reliable and resilient data forwarding in WSNs without requiring geographic location information. Tunnel routing exploits a route construction procedure to identify a family of trusted sensor nodes that form a *tunnel* between a source and a destination and that later contend among each other to act as data forwarders. Moreover, terminals within the tunnel may act as *support nodes* by providing on-demand retransmissions of data packets to combat channel impairments (cooperative relaying), or by initiating multipath data forwarding towards the destination in case they detect a misbehavior caused by a selective forwarding attack. The proposed protocol offers the robustness of dynamic routing without requiring any additional effort in terms of location management. Moreover, our solution is resilient to packet dropping and selective forwarding attacks and degrades gracefully in the presence of attacks to several nodes.

The remainder of our discussion is organized as follows: Section A.4.1 presents the network model that we consider. Then, Section A.4.2 describes details of the proposed dynamic tunnel routing protocol, while Section A.4.3 provides the results of some simulations performed to investigate the performance of our solution.

A.4.1 Network Assumptions and Thread Model

In our work, we consider WSNs composed by a large number of sensor nodes that are scattered over a wide geographic area. In this context, data generated at a source may have to undergo several hops in order to reach the intended destination. Nodes that can be reached in one or two hops from a sensor $\boldsymbol{X}$ are referred to as the 1- or 2-hop neighbors of $\boldsymbol{X}$. We assume that the packet transmission between any two sensor nodes is modeled by the input-output relation: $y(t) = h(t)a(t) + z(t)$, where $y(t)$ and $a(t)$ are the complex baseband receive and transmit signal; $h(t)$ models the radio channel and takes into account path loss and fading effects, and the additive noise $z(t)$ is a zero-mean white complex Gaussian process. We apply LEAP at the initial deployment of each sensor node to establish symmetric shared keys between terminals for providing confidentiality and authentication in the WSN. In particular, this key management protocol discovers all 1-hop neighbors of the newly deployed node, and establishes on it one pair-wise link key for secure point-to-point communication to each 1-hop neighbor and one cluster key that is shared with all its 1-hop neighbors. We assume that each node encrypts its generated control messages and data packets with the cluster key in order to allow 1-hop neighbor nodes to overhear the information transmission. In addition, we assume that each node possesses not only a list of all trusted 1-hop neighbors, but also a verified list of all 2-hop neighbors. The latter list is obtained by executing an extended LEAP algorithm [94]. Several insider attacks can be launched to compromise one or several nodes of the WSN [91]. Our protocol, as discussed, is designed so that it can reliably defend against forwarding insider attacks, where malicious sensor nodes may refuse to forward certain data packets or simply drop all of them ensuring that they are not forwarded any further. The proposed tunnel routing scheme can also cope with blind-letter attacks if a list of all trusted 2-hop neighbors is available at the sensor node.

A.4.2 Tunnel Routing Protocol

Tunnel routing is a novel dynamic routing strategy for reliably and resiliently forwarding data from a source $\boldsymbol{S}$ to a destination $\boldsymbol{D}$ in a multi-hop WSN. The algorithm first performs a tunnel discovery procedure in order to locate the destination in the network, identify a set of potential data forwarders, and gather information later required to route data towards $\boldsymbol{D}$. Each node stores such information in an internal routing table, whose entries comprise the address src_{addr} of the node that initiated the tunnel discovery procedure, the destination address $\text{dest}_{\text{addr}}$, a sequence number seq_{ID} that uniquely identifies the tunnel

discovery request, the hop distances $n_{hop,S}$ and $n_{hop,S}$ of the node from $\boldsymbol{S}$ and $\boldsymbol{D}$, and a parameter TimeStamp that defines the tunnel lifetime. Once the tunnel has been constructed, data packets can be routed from $\boldsymbol{S}$ through to $\boldsymbol{D}$ by applying a dynamic data forwarding scheme. In this section, we first describe the tunnel discovery procedure. Then, we discuss the dynamic data forwarding scheme and finally we present functionalities that can be added to the protocol to cope with errors caused by the radio channel and selective-forwarding attacks.

A.4.2.1 Tunnel Discovery Procedure

When a node $\boldsymbol{S}$ has data to deliver to $\boldsymbol{D}$ and an active tunnel to $\boldsymbol{D}$ has not yet been established, a new entry in its routing table is created, with fields src_{addr} = S, $dest_{addr}$ = D, $n_{hop,S} = 0$, $n_{hop,D} = \infty$, and TimeStamp = 0. The field seq_{ID} is initialized to seq_old + 1, if an entry for $\boldsymbol{D}$ with seq_{ID} = seq_old already exists, or to 1, otherwise. Once the entry has been written, $\boldsymbol{S}$ broadcasts an encrypted Route REQuest (RREQ) message to all its 1-hop neighbors, comprising the parameters source address, destination address, sequence number seq = seq_{ID}, and hop count number $n_{hop} = n_{hop,S}$. A node that decodes this packet compares the received parameter n_{hop} to a threshold $n_{hop,max}$. If $n_{hop} \geq n_{hop,max}$, the message is discarded to avoid long and unreliable paths during data forwarding. Otherwise, the node checks its routing table. If an entry for the node pair $\boldsymbol{S}$-$\boldsymbol{D}$ already exists with the same or a smaller sequence number than the one received, the message is discarded. Otherwise, the node creates a routing table entry with fields src_{addr} = S, $dest_{addr}$ = D, seq_{ID} = seq, $n_{hop,S} = n_{hop}+1$, $n_{hop,D} = \infty$, and TimeStamp = 0. Any terminal that has the $n_{hop,S}$ field set to a non-zero value for a given node pair $\boldsymbol{S}$-$\boldsymbol{D}$ is said to have the *Forward Flag* switched on for that couple. Once the table entry has been stored, the node creates an encrypted RREQ message with an updated n_{hop} parameter and transmits it. The described procedure is iterated until $\boldsymbol{D}$ is reached. When the destination receives the RREQ packet, it creates a new entry in its routing table with fields src_{addr} = S, $dest_{addr}$ = D, $n_{hop,S} = n_{hop}$, $n_{hop,D} = 0$, seq_{ID} = seq, TimeStamp = CurrentTime. Then, $\boldsymbol{D}$ broadcasts an encrypted Route REPly (RREP) message to all its 1-hop neighbors, comprising the parameters src_{addr} = S, $dest_{addr}$ = D, seq = seq_{ID}, $n_{hop} = n_{hop,D}$, and Time = TimeStamp. Neighbors that decode this message perform the test on $n_{hop,max}$ and, if the requirement is fulfilled, check if they have switched on the *Forward Flag* for the pair $\boldsymbol{S}$-$\boldsymbol{D}$. If so, they further determine if the sequence number stored in the seq_{ID} field equals the one provided by the RREP message. If the tests fail, the node discards the received message. Otherwise, the $n_{hop,D}$ and TimeStamp fields of the corresponding

entry of the routing table are updated according to the parameters received with the RREP message as follows: $n_{hop,D} = n_{hop}+1$, TimeStamp = Time. Any node that has the $n_{hop,D}$ field set to a non-infinite value for $\boldsymbol{S}$-$\boldsymbol{D}$ is said to have the *Backward Flag* switched on for that couple. Nodes with both *Forward* and *Backward Flag* on for a given pair are said to be *active* for that couple with sequence number seq given by the corresponding routing table entry. Once the table entry has been written, the node creates a new encrypted RREP message with an updated n_{hop} parameter and broadcasts the message. At the end of the procedure, i.e., when $\boldsymbol{S}$ is reached, the protocol has identified a tunnel connecting $\boldsymbol{S}$ and $\boldsymbol{D}$, composed by the set of active nodes. Such a tunnel is kept active for a fixed time interval after its establishment (tunnel lifetime).

A.4.2.2 Dynamic Data Forwarding

After an active routing tunnel has been established, dynamic transfer of data packets can start. To select the next hop within the tunnel, we propose a CSMA-based distributed contention that exchanges RTS/CTS messages between neighbor nodes. As opposed to other dynamic routing strategies [80, 83], our approach does not rely on any geographic location information. Let $\boldsymbol{X}_n$ be the active node in the tunnel with hop distance n from $\boldsymbol{S}$ that has been selected to forward a DATA packet towards $\boldsymbol{D}$, and let $\mathcal{A}$ be the set of active 1-hop neighbors of $\boldsymbol{X}_n$. To identify the best-suited next-hop $\boldsymbol{X}_{n+1}$, $\boldsymbol{X}_n$ broadcasts an encrypted RTS message with parameters S, D, seq, and its hop count distance $n_{hop} = n_{hop,D}$. Any node $\boldsymbol{X} \in \mathcal{A}$ which successfully decodes this message and has a hop count $n_{hop,D}$ less than or equal to the one contained in the RTS, computes a routing metric $m_x(t)$ to assess its adequacy to serve as next-hop, defined as:

$$m_x(t) = (n_{hop,S} + n_{hop,D})^{-1} |h_x(t)|^2 \,. \tag{A.18}$$

The metric balances two cost criteria: the path length in terms of hop count from $\boldsymbol{S}$ and $\boldsymbol{D}$, and the quality $|h_X(t)|^2$ of the radio link between $\boldsymbol{X}_n$ and $\boldsymbol{X}$. $m_x(t)$ increases for nodes with short path length and experiencing good channel conditions to $\boldsymbol{X}_n$. Moreover, the inverse proportionality to $n_{hop,S}$+$n_{hop,D}$ guarantees an advancement towards the destination and avoids routing loops. After computing the metric, the nodes contend with each other to act as data forwarder $\boldsymbol{X}_{n+1}$. For this purpose, each terminal $\boldsymbol{X} \in \mathcal{A}$ sets a backoff timer TIMER_FORWARD, whose duration is inversely proportional to $m_x(t)$. During the backoff period, nodes listen to the medium and compare the received radio signal power to a given threshold value. If the timer expires for a terminal before the sensed power exceeds the

threshold, it proposes itself as next-hop by broadcasting an encrypted CTS message with parameters S, D, and seq to its 1-hop neighbors. Otherwise, the node stops the timer because a CTS message has very likely been sent from a neighbor that won the contention. After $\boldsymbol{X}_n$ has decoded the CTS message, it unicasts the DATA packet encrypted with its cluster key over the radio channel to the chosen forwarding node $\boldsymbol{X}_{n+1}$. Moreover, a timer TIMER_MONITOR is started that may trigger a repetition of the data forwarding procedure if a successful reception and forwarding of the DATA packet at node $\boldsymbol{X}_{n+1}$ cannot be monitored by node $\boldsymbol{X}_n$ within this time interval. If $\boldsymbol{X}_{n+1}$ successfully decodes data, it iterates the dynamic forwarding procedure until the destination is reached. The timer TIMER_MONITOR is reset after $\boldsymbol{X}_n$ is able to monitor the transmission of the DATA packet from $\boldsymbol{X}_{n+1}$ to $\boldsymbol{X}_{n+2}$.

A.4.2.3 Reliable and Resilient Data Forwarding

In order to cope with communication errors induced by harsh channel conditions or by misbehaving nodes, we propose a set of additional functionalities for the described protocol. Let $\mathcal{B}$ be the set of active nodes that are common 1-hop neighbors of the two forwarding nodes $\boldsymbol{X}_n$ and $\boldsymbol{X}_{n+1}$. Such terminals can overhear packet exchanges between $\boldsymbol{X}_n$ and $\boldsymbol{X}_{n+1}$ (RTS/CTS/DATA), as they possess the cluster key for both the nodes. This capability makes it possible for terminals $\boldsymbol{S} \in \mathcal{B}$ to give $\boldsymbol{X}_n$ and $\boldsymbol{X}_{n+1}$ support in reliably and resiliently forwarding data packets, hence we refer to them as *support nodes*. In particular, when a support node $\boldsymbol{S}$ overhears a data exchange from $\boldsymbol{X}_n$ to $\boldsymbol{X}_{n+1}$, it stores the received packet in a buffer and starts a timer TIMER_MONITOR. As long as the timer is active, $\boldsymbol{S}$ provides on-demand cooperative relaying support to the $\boldsymbol{X}_{n+1}$ and monitors the behavior of this node to detect potential misbehaviors.

Cooperative Relaying: If $\boldsymbol{X}_{n+1}$ fails to decode data from $\boldsymbol{X}_n$, it caches the corrupted version of the packet and sends an encrypted RETransmit REQest (RETREQ) message to its 1-hop neighbors. Support nodes in $\mathcal{B}$ that receive the RETREQ and that successfully decoded the packet sent by $\boldsymbol{X}_n$ can give cooperative relaying support by re-encoding and encrypting such data, and forwarding it to $\boldsymbol{X}_{n+1}$. To select the best-suited node, we resort again to a CSMA-based distributed contention. Each candidate $\boldsymbol{S}$ computes a metric $m_s(t)$ proportional to the channel gain with $\boldsymbol{X}_{n+1}$ estimated from the reception of the RETREQ, and sets a backoff timer TIMER_RELAY inversely proportional to $m_s(t)$. During the backoff phase, carrier sensing is performed. If the sensed power exceeds a given threshold, the node assumes that someone else wins the contention and goes back to its activity. Otherwise, if

the time expires, the terminal acts as relay by sending a copy of the cached data to $\boldsymbol{X}_{n+1}$ as per the decode-and-forward paradigm. If the reception succeeds, $\boldsymbol{X}_{n+1}$ proceeds with data forwarding to a next-hop node $\boldsymbol{X}_{n+2}$. Otherwise, the terminal sends a retransmit request to $\boldsymbol{X}_n$, that may perform one or more forwarding attempts before triggering a failure.

Local Monitoring: Support nodes can also defend the network against compromised terminals that may entirely or selectively drop packets. As discussed, a node $\boldsymbol{S} \in \mathcal{B}$ that successfully decode data sent by $\boldsymbol{X}_n$ starts a TIMER_MONITOR timer. If within this period $\boldsymbol{S}$ decodes a data exchange from $\boldsymbol{X}_{n+1}$ to a next-hop $\boldsymbol{X}_{n+2}$, it verifies in its neighbor list that $\boldsymbol{X}_{n+2}$ is a trusted 2-hop neighbor and compares the incoming data to the original packet received from $\boldsymbol{X}_n$. When the former condition is met and the two packets are identical, $\boldsymbol{S}$ concludes that the communication over the $\boldsymbol{X}_n$ - $\boldsymbol{X}_{n+1}$ link succeeded and that $\boldsymbol{X}_{n+1}$ is properly forwarding the payload. If so, the node resets its TIMER_MONITOR timer and goes back to its activity. On the contrary, if the overheard packet is sent to a non-trusted 2-hop neighbor or no data packet transmission is observed before timer expiration, the support node assumes that $\boldsymbol{X}_{n+1}$ maliciously drops packets or redirects them to untrusted nodes. In this case, $\boldsymbol{S}$ resets initiates an alternative path towards the destination through the tunnel by applying the dynamic forwarding algorithm discussed in Section A.4.2.2. Since several support nodes may simultaneously detect the malfunction of $\boldsymbol{X}_{n+1}$, a malicious behavior automatically triggers multi-path forwarding of the data. Duplicate copies of the same packet can easily be suppressed by the destination or by common forwarding nodes in the routing paths. The capability of relying on multi-path data forwarding through the tunnel strongly increases the robustness of the protocol against insider attacks, as will be discussed in Section A.4.3. We remark that, as opposed to static multi-path routing protocols [93], our solution introduces multiple flows (thus increasing energy consumption) only if one or several malicious nodes actually hamper the correct operation of the network.

A.4.3 Simulation Results

In order to test the effectiveness of the proposed solutions, extensive Omnet++ simulations have been performed. Three routing schemes have been compared: AODV; a basic version of tunnel routing (TR) that does not exploit support nodes (i.e., that implements only the features described in Section A.4.2.2); and the complete Tunnel Routing with Support approach (see Section A.4.2.3), named TRS. All the schemes rely on a carrier-sense based medium access control, modified according to Section A.4.2 for the protocols that implement tunnel routing. We have considered topologies composed by 100 nodes spread over a

Table A.2. *Parameters used in our simulations*

Tx power / Noise Floor / CS threshold	-90 dBm, 92 dBm
Data Rate	250 kbit/s
Short Retry Limit at MAC layer	3
Payload length / Control pkts	1024 bit, 112 bit
Route lifetime	50 s

100m$\times$50m area. Poisson traffic with intensity $\lambda = 0.1$ pk/s is injected in the network by a single source located at {30,25}, and addressed to a sink located at {70,25}, the coordinates being expressed in meters, whereas the other sensor nodes are randomly distributed in the considered region. With this settings, $\boldsymbol{S}$ and $\boldsymbol{D}$ are on average 4 hops away. To analyze the resiliency of the different routing protocols to security attacks, each node within the network can be *malicious* with probability p_{mal}. A malicious terminal participates as usual to route (or tunnel) construction procedures, yet it constantly drops data packets it is supposed to forward.

The wireless environment is subject to path loss with exponent $\alpha = 3.5$ and correlated Rayleigh fading. The standard set of parameters used in our simulations is reported in Tab. A.2. All the results presented here have been obtained by averaging the outcome of 100 independent simulations of duration 10000s, which provided the desired statistical confidence.

The first metric that we consider is the end to end packet delivery ratio (PDR), depicted against p_{mal} in Fig. A.14. When nodes in the network properly forward DATA packets (i.e., $p_{mal} = 0$), all the schemes provide high reliability. Moreover, under this hypothesis, protocols that rely on the tunnel routing approach are able to offer an improvement of almost 10% with respect to AODV. This stems from the higher resiliency to channel impairments achieved by the proposed schemes. As discussed, AODV relies on a single path to forward data. Thus, as soon as one of the links along the route incurs harsh channel conditions due to fading or interference, the whole data flow towards the destination collapses until a new path is established. On the contrary, our protocols are able to select a data forwarder among a set of potential candidates, and therefore they manage to avoid weak links in favor of more reliable connections. Fig. A.14, also shows that, in the absence of corrupted nodes, TRS slightly improves PDR with respect to basic TR thanks to cooperative relaying, which helps in recovering link failures that may happen despite the dynamic choice of the next

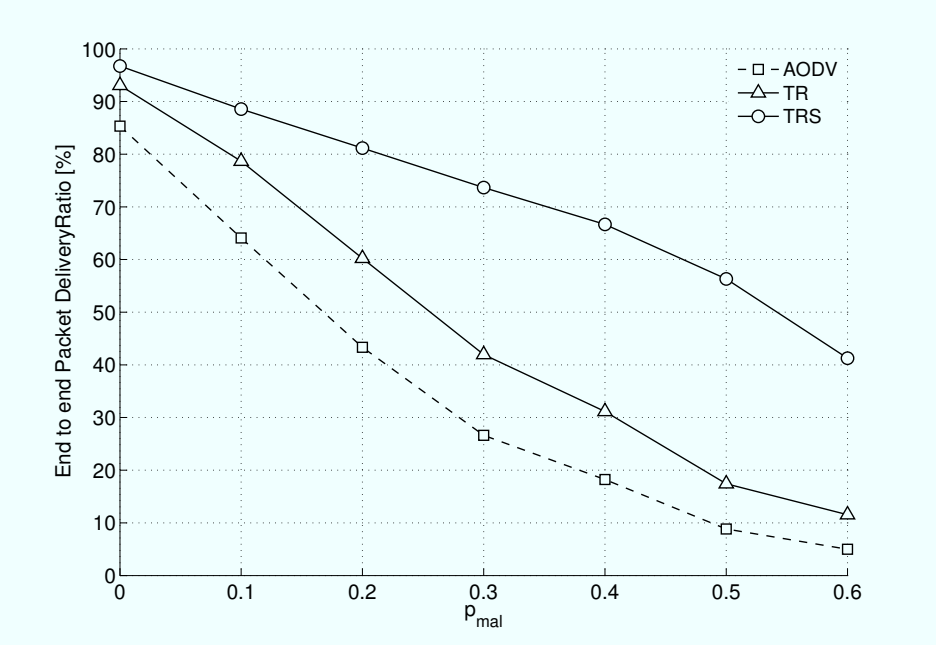


Figure A.14. *Dependence of the end-to-end Packet Delivery Ratio on the the fraction p_{mal} of compromised nodes in the network.*

hop. Incidentally, we notice that these gains are rather small. This stems from the low number of cooperative retransmissions that are triggered in the topologies under study due to the low level of aggregate interference. The impact of cooperative relaying would raise in networks characterized by multiple simultaneous data flows, whose analysis we regard as part of our future work. Let us now focus on the behavior of the different routing schemes when terminals may maliciously drop data packets. First of all, Fig. A.14 shows that the reliability offered by AODV plummets as the number of corrupted nodes in the network grows, with almost no packets delivered for values of p_{mal} higher than 0.5. As malicious terminals actively participate in the route construction procedure, the probability that at least one of such nodes is selected as part of a path identified by AODV increases with p_{mal}. In this condition, static routing is not able to deliver any data to the destination, since packets constantly get lost as soon as they reach the first compromised terminal along the path. The stall continues until a new route is built, either due to a link failure or to the expiration of the lifetime for the current path, leading to a poor reliability. This effect can be mitigated by using dynamic routing, as shown by Fig. A.14, where the TR protocol offers improvements up to 20% over AODV for values of p_{mal} up to 0.3. When TR is implemented, data can still be dropped if a malicious node within the tunnel is selected as a next hop. Nevertheless, since with our approach each frame undergoes a potentially different route to reach the des-

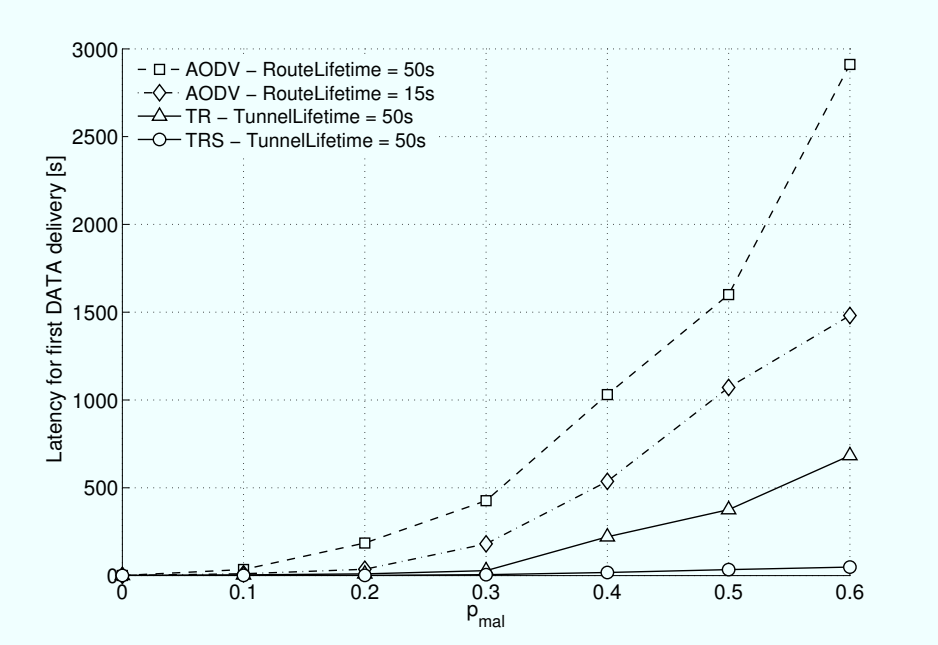


Figure A.15. *Average time required to deliver the first packet in the network as a function of the fraction p_{mal} of compromised nodes in the network.*

tination, the probability that subsequent packets are dropped is significantly lowered with respect to AODV. Finally, let us consider the performance of TRS. Fig. A.14 highlights that our protocol outperforms AODV by as much as 50% in terms of reliability for values of p_{mal} around 0.3. Moreover, the plot shows that even when half of the nodes in the networks are corrupted, TRS is able to successfully deliver 60% of the generated traffic to the sink. This remarkable result stems from the capability of the proposed solution to trigger alternative routes to the destination as soon as support nodes detect a misbehavior by a terminal which is supposed to forward data. In this way, even packets that are dropped along their path have the possibility to be recovered and to eventually reach the destination, thanks to the cooperative behavior of all the nodes within the tunnel.

Another aspect of interest for WSNs subject to insider attacks is the time needed for the first packet to reach the destination, as it describes the capability of a routing scheme to quickly establish a connection and to deliver information to a sink within an hostile environment. The metric is depicted against p_{mal} in Fig. A.15. The plot clearly highlights how the latency for the first delivery strongly increases for AODV as the number of compromised nodes in the network grows. This stems once again from the fact that no packet can be delivered to the destination until a route that does not comprise any malicious terminal is identified. On the contrary, the capability offered by schemes based on tunnel routing to ex-

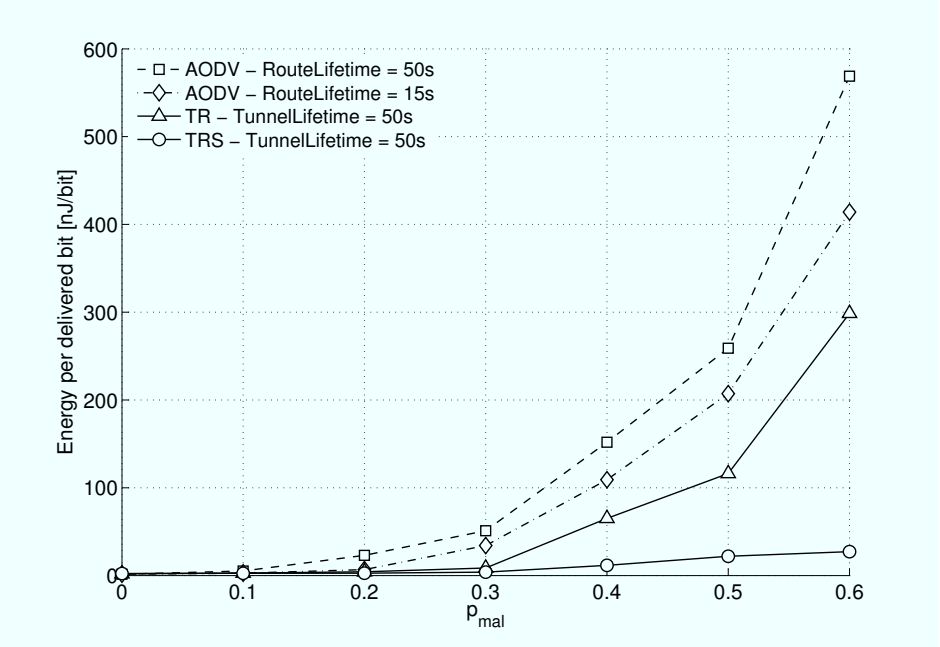


Figure A.16. *Energy required to deliver the first packet in the network as a function of the fraction p_{mal} of compromised nodes in the network.*

ploit distinct paths within the tunnel when forwarding different packets drastically reduces this latency. Incidentally, we remark that TRS outperforms basic TR thanks to its capability of resorting to multiple routes to recover frames that would otherwise be lost. Starting from this discussion, we observe that the performance in terms of latency for AODV can be improved by forcing the protocol to build routes more frequently (i.e., reducing the Route Lifetime parameter in Tab. A.2), so that the time required to find a route that does not include any compromised node is shortened. The dash-dotted curve in Fig. A.15 shows the results obtained following this approach. Two remarks can be made. First of all, although reduced, the latency undergone by AODV to deliver the first packet is still considerably higher that the ones obtained when using tunnel routing. Secondly, an increase in the frequency of route constructions induces a significant raise in terms of control overhead, and thus of energy depletion, which is a metric of capital importance in battery constrained WSNs, since a lower energy consumption turns into a longer network lifetime.

This trend is confirmed by Fig. A.16, , e.g., consider a comparison of the dashed and dash-dotted curves, which depicts exactly the transmission energy consumed in the whole network in order to deliver the first packet to the sink against p_{mal}. Moreover, this plot also highlights that the dynamic routing approaches outperform AODV, and shows how the quicker deliveries due to resiliency to insider attacks offered by TRS make it possible to

significantly reduce energy consumption with respect to basic TR.

In conclusion, we can infer that not only does TRS guarantee high end to end delivery reliability even when a large fraction of the network is compromised, but also it is able to connect source and destination with low latency and at low costs in terms of energy consumption with respect to existing and widely implemented static routing solutions.

Appendix B

Derivation of some probabilities of interest

B.1 Derivation of $\mathcal{O}^{t_i}(\boldsymbol{p}, t_i)$

With reference to the scenario described in Section 5.2, we shall compute the probability that the interfering terminal $\boldsymbol{I}$ induces an outage at the destination given that both $\boldsymbol{S}$ and $\boldsymbol{I}$ have been granted access to the medium, i.e., $\mathcal{O}^{t_i}(\boldsymbol{p}, t_i)$. Conditioning on the power received over the $\boldsymbol{S}$-$\boldsymbol{D}$ link, we get:

$$\mathcal{O}^{t_i}(\boldsymbol{p}, t_i) = \int_0^{+\infty} \Pr\{\mathcal{L}^{t_i}(\boldsymbol{\eta}, t_i) < L | \eta_{s,d}\} f(\eta_{s,d}, P\delta_{s,d}^{-\alpha}) d\eta_{s,d}\,, \tag{B.1}$$

where $\mathcal{L}^{t_i}(\boldsymbol{\eta}, t_i)$ is represented by (5.4), and $f(\eta, a) = e^{-\frac{\eta}{a}}/a$.

In order to solve (B.1), we identify three disjoint regions whose union spans the whole integration domain: $\mathcal{R}_1 = [0, \eta^*)$, $\mathcal{R}_2 = [\eta^*, \bar{\eta})$, and $\mathcal{R}_3 = [\bar{\eta}, +\infty)$, so that for $\eta_{s,d} \in \mathcal{R}_1$ an outage event is induced by the sole noise power, regardless of the values of t_i and $\eta_{i,d}$, whereas for $\eta_{s,d} \in \mathcal{R}_3$ the capacity of the $\boldsymbol{S}$-$\boldsymbol{D}$ channel is sufficient to support a successful communication during the interference-free interval of duration $T - |t_i|$. The values of η^* and $\bar{\eta}$ can be determined imposing the conditions $\mathcal{C}(\eta^*/N) = L$, and $|t_i| \cdot \mathcal{C}(\bar{\eta}/N) = L$, obtaining: $\eta^* = N \cdot (2^{\rho_{data}/B} - 1)$ and $\bar{\eta} = N \cdot (2^{L/(B \cdot |t_i|)} - 1)$, respectively. Taking advantage of this subdivision, it immediately follows that the integration over $\mathcal{R}_3$ returns zero, while the integrand over $\mathcal{R}_1$ simplifies to $f(\eta_{s,d}, P\delta_{s,d}^{-\alpha})$. Therefore, we get:

$$\mathcal{O}^{t_i}(\boldsymbol{p}, t_i) = 1 - e^{-\frac{\eta^*}{P\delta_{s,d}^{-\alpha}}} + \int_{\eta^*}^{\bar{\eta}} e^{-\frac{\xi(\eta_{s,d}, t_i)}{P\delta_{i,d}^{-\alpha}}} f(\eta_{s,d}, P\delta_{s,d}^{-\alpha})\, d\eta_{s,d}\,, \tag{B.2}$$

where

$$\xi(\eta_{s,d}, t_i) = \frac{\eta_{s,d}}{\left(\frac{2^{L/(B\cdot|t_i|)}}{1+\frac{\eta_{s,d}}{N}}\right)^{\frac{|t_i|}{T-|t_i|}} - 1} - N \tag{B.3}$$

is obtained by solving $\Pr\{\mathcal{L}^{t_i}(\boldsymbol{\eta}, t_i) < L\}$ with respect to $\eta_{i,d}$. The value of $\mathcal{O}^{t_i}(\boldsymbol{p}, t_i)$ for any topology $\boldsymbol{p}$ and for any value of t_i can be computed by numerical evaluation of (B.2).

B.2 Derivation of $\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$

Let us refer to the topology described in Section 5.2, and let $\boldsymbol{\eta}' = \{\eta_{s,c}, \eta_{i,c}\}$ be the vector of the powers received at $\boldsymbol{C}$ from $\boldsymbol{S}$ and $\boldsymbol{I}$, respectively. Furthermore, let $\mathcal{L}_c^{t_i}(\boldsymbol{\eta}', t_i)$ be the number of information bits sent by the source that are decoded at such node conditioned on the start time t_i of the interfering communication.[1] According to the discussion of Section 5.2, the probability that $\boldsymbol{C}$ is available for cooperation, under the assumption that $t_i > 0$, can be written as:

$$\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i) = \Pr\{\mathcal{L}_c^{t_i}(\boldsymbol{\eta}', t_i) > L \ , \ \eta_{i,c} + N < \Lambda\} \ . \tag{B.4}$$

Let now $\mathcal{H}_{+}^{p_i,t_i,\eta_{s,c}}(\boldsymbol{p}', p_i, t_i, \eta_{s,c})$ represent (B.4) conditioned on $\eta_{s,c}$, so that

$$\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i) = \int_0^{+\infty} \mathcal{H}_{+}^{p_i,t_i,\eta_{s,c}}(\boldsymbol{p}', p_i, t_i, \eta_{s,c}) f(\eta_{s,c}, P\delta_{s,c}^{-\alpha}) \, d\eta_{s,c}. \tag{B.5}$$

In order to solve this equation, we split the integration domain in three disjoint regions corresponding to those identified in Appendix B.1: $\mathcal{R}_1 = [0, \eta^*)$, $\mathcal{R}_2 = [\eta^*, \bar{\eta})$, and $\mathcal{R}_3 = [\bar{\eta}, +\infty)$. When $\eta_{s,c} \in \mathcal{R}_1$, the sole noise induces an outage at $\boldsymbol{C}$ so that the region does not contribute to the overall computation. On the other hand, when $\eta_{s,c} \in \mathcal{R}_3$, decoding is ensured regardless of the interfering power. It follows that in this case the integrand reduces to $\Pr\{\eta_{i,c} + N < \Lambda\} \cdot f(\eta_{s,c}, P\delta_{s,c}^{-\alpha})$, and (B.5) restricted to $\mathcal{R}_3$ returns:

$$\left(1 + e^{-\frac{\Lambda - N}{P\delta_{i,c}^{-\alpha}}}\right) \cdot e^{-\frac{\bar{\eta}}{P\delta_{s,c}^{-\alpha}}} \ . \tag{B.6}$$

Finally, if $\eta_{s,c} \in \mathcal{R}_2$, the level of interference determines both the decoding probability and the probability that $\boldsymbol{C}$ senses the medium idle. Therefore, solving $\mathcal{L}_c^{t_i}(\boldsymbol{\eta}', t_i) > L$ with respect to $\eta_{i,c}$, we can write the contribution of $\mathcal{R}_2$ to $\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$ as:

[1] $\mathcal{L}_c^{t_i}(\boldsymbol{\eta}', t_i)$ is promptly given by (5.1), replacing $\eta_{s,d}$ and $\eta_{i,d}$ with $\eta_{s,c}$ and $\eta_{i,c}$ respectively.

$$\int_{\eta^*}^{\bar{\eta}} \Pr\left\{\eta_{i,c} < \xi(\eta_{s,c}, t_i)\ ,\ \eta_{i,c} < \Lambda - N | \eta_{s,c}\right\} f(\eta_{s,c}, P\delta_{s,c}^{-\alpha})\ d\eta_{s,c}\ . \tag{B.7}$$

$\xi(\eta_{s,c}, t_i)$, defined in (B.3), has a zero for $\eta_{s,c} = \eta^*$, is monotonically increasing in the interval $[\eta^*, \bar{\eta})$, and exhibits a vertical asymptote at $\bar{\eta}$. Let us now define a value $\tilde{\eta}$ so that

$$\xi(\tilde{\eta}, t_i) = \Lambda - N\,. \tag{B.8}$$

From the aforementioned properties, we can infer that

$$\min\{\xi(\eta_{s,c}, t_i), \Lambda - N\} = \begin{cases} \xi(\eta_{s,c}, t_i) & \text{if } \eta_{s,c} \in [\eta^*, \tilde{\eta}) \\ \Lambda - N & \text{if } \eta_{s,c} \in [\tilde{\eta}, \bar{\eta}) \end{cases} \tag{B.9}$$

Therefore, (B.7) can be rewritten as:

$$\int_{\eta^*}^{\tilde{\eta}} \left(1 - e^{-\frac{\xi(\eta_{s,c}, t_i)}{P\delta_{i,c}^{-\alpha}}}\right) f(\eta_{s,c}, P\delta_{s,c}^{-\alpha})\ d\eta_{s,c} + \int_{\tilde{\eta}}^{\bar{\eta}} \left(1 - e^{-\frac{\Lambda - N}{P\delta_{i,c}^{-\alpha}}}\right) f(\eta_{s,c}, P\delta_{s,c}^{-\alpha})\ d\eta_{s,c}\ . \tag{B.10}$$

Finally, combining the results of (B.6) and (B.10), we obtain:

$$\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i) = e^{-\frac{\tilde{\eta}}{P\delta_{s,c}^{-\alpha}}} \cdot \left(1 - e^{-\frac{\Lambda - N}{P\delta_{i,c}^{-\alpha}}}\right) + \int_{\eta^*}^{\tilde{\eta}} \left(1 - e^{-\frac{\xi(\eta_{s,c}, t_i)}{P\delta_{i,c}^{-\alpha}}}\right) f(\eta_{s,c}, P\delta_{s,c}^{-\alpha})\ d\eta_{s,c}.$$

Given the complexity of the involved functions, it is not possible to find a closed form for the considered probabilities. However, $\tilde{\eta}$ and consequently $\mathcal{H}_{+}^{p_i,t_i}(\boldsymbol{p}', p_i, t_i)$ can be determined by means of numerical evaluation for any topology and for any value of t_i.

Selected List of Publications

Part of the work presented in this thesis has also appeared in the articles reported below.

Journal papers

[J1] A. Munari, F. Rossetto, M. Zorzi, "Cooperative Cross Layer MAC Protocols for Directional Antenna Ad Hoc Networks", *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 12, No. 2, pp. 12–30, Apr. 2008.

[J2] A. Munari, F. Rossetto, M. Zorzi, "Phoenix: Making Cooperation more Efficient through Network Coding in Wireless Networks", *IEEE Transactions on Wireless Communications*, Vol. 8, No. 10, pp. 5248 - 5258,Oct. 2009.

[J3] A. Munari, F. Rossetto, M. Zorzi, "Impact of Medium Access Control Strategies on the Effectiveness of Advanced Cooperative Hybrid ARQ Techniques", submitted to IEEE Transactions on Wireless Communications.

[J4] M. Levorato, A. Munari, M. Zorzi, "On the Impact of Carrier Sense Based Medium Access Control on Cooperative Strategies in Wireless Ad Hoc Networks - Part I", under submission to IEEE Transactions on Wireless Communications.

[J5] M. Levorato, A. Munari, M. Zorzi, "On the Impact of Carrier Sense Based Medium Access Control on Cooperative Strategies in Wireless Ad Hoc Networks - Part II", under submission to IEEE Transactions on Wireless Communications.

Conference papers

[C1] A. Munari, F. Rossetto, M. Zorzi, "A new Cooperative Strategy for Deafness Prevention in Directional Ad Hoc Networks," in *Proc. IEEE ICC'07*, Glasgow, Scotland, 24–28 Jun. 2007, pp. 3154–3160.

[C2] E. Fasolo, A. Munari, F. Rossetto, M. Zorzi, "Phoenix: A hybrid cooperative-Network Coding Protocol for Fast Failure Recovery in Ad Hoc Networks," in *Proc. IEEE SECON'08*, San Francisco, CA (US), 16–20 Jun. 2008, pp. 404–410.

[C3] A. Munari, W. Schott, "Performance Assessment of a Class of Cross-Layer Optimized Protocols for Geographic Routing in WSNs," in *Proc. IEEE PIMRC'08*, Cannes, France, 15–18 Sep. 2008.

[C4] A. Munari, F. Rossetto, M. Zorzi, "On the Viability of a Cooperative-Network Coding Protocol in Clustered Networks, " in *Proc. IEEE MilCom '08*, San Diego, CA (US), 17–19 Nov. 2008.

[C5] M. Levorato, A. Munari, M. Zorzi, "On the Effectiveness of Cooperation in Carrier Sense Based Ad Hoc Networks, " in *Proc. IEEE SECON '09*, Rome, Italy, 22–26 Jun. 2009.

[C6] A. Munari, W. Schott and Y.W. Law, "Dynamic Tunnel Routing for Reliable and Resilient Data Forwarding in Wireless Sensor Networks, " in *Proc. IEEE PIMRC '09*, Tokyo, Japan, 13–16 Sep. 2009.

[C7] A. Munari, W. Schott and S. Krishnan, "Energy-Efficient Routing in Mobile Wireless Sensor Networks Using Mobility Prediction, " in *Proc. IEEE LCN '09*, Zurich, Switzerland, 20–23 Oct. 2009.

[C8] A. Munari, F. Rossetto, M. Zorzi, "Hybrid Cooperative-Network Coding Medium Access Control for High-Rate Wireless Personal Area Networks", in *Proc. IEEE ICC'10*, Cape Town, South Africa, 23–27 May, 2010.

Bibliography

[1] IEEE Journal on Selected Areas in Communications, "Special Issue on Ad Hoc Networks," 1999.

[2] R. Ramanathan, "A Brief Overview of Ad Hoc Networks: Challenges and Directions," *IEEE Communications Magazine*, pp. 20–22, May 2002.

[3] S. Xu and T. Saadawi, "Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks?" *IEEE Communications Magazine*, pp. 130–137, Jun. 2001.

[4] C. A. Balanis, *Antenna theory, Analysis and Design*. New York (NY, USA): John Wiley, 2005.

[5] IEEE LAN MAN Standards, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications High-speed Physical Layer in the 5 GHz Band*, ANSI/IEEE Std., Sep. 1999.

[6] R. R. Choudhury and N. H. Vaidya, "Impact of Directional Antennas on Ad Hoc Routing," in *IFIP Personal and Wireless Communications*, Venice, Italy, Sep. 2003.

[7] R. Ramanathan, "On the Performance of Ad Hoc Networks with Beamforming Antennas," in *ACM MobiHoc*, Long Beach, CA, USA, 2001.

[8] M. Takai and J. Martin and A. Ren and R. Bagrodia, "Directional Virtual Carrier Sensing for Directional Antennas in Mobile Ad Hoc Networks," in *ACM MobiHoc*, Lausanne, Switzerland, Jun. 2002.

[9] S. Bandyopadhyay and K. Hasuike and S. Horisawa and S. Tawara, "An Adaptive MAC and Directional Routing Protocol for Ad Hoc Wireless Network Using ESPAR Antenna," in *ACM MobiHoc*, Long Beach, CA, US, Oct. 2001.

[10] R. R. Choudhury and X. Yang and R. Ramanathan and N. Vaidya, "Using Directional Antennas for Medium Access Control in Ad Hoc Networks," in *ACM MobiCom*, Atlanta, Georgia, US, Sep. 2002.

[11] Y. G. Ko and N. H. Vaidya, "Medium Access Control Protocols Using Directional Antennas in Ad Hoc Networks," in *IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.

[12] T. Korakis and G. Jakllari and L. Tassiulas, "A MAC Protocol for Full Exploitation of Directional Antennas in Ad-hoc Wireless Networks," in *ACM MobiHoc*, Annapolis, MA, US, Jun. 2003.

[13] G. Jakllari and I. Broustis and T. Korakis and S.V. Krishnamurthy and L. Tassiulas, "Handling Asymmetry in Gain in Directional Antenna Equipped Ad hoc Networks," in *IEEE PIMRC*, Berlin, Germany, Sep. 2005.

[14] F. Rossetto and M. Zorzi, "A Space-Time Based Approach to Solving the Gain Asymmetry in MIMO Ad Hoc Networks," in *IEEE VTC Spring*, Melbourne, Australia, May. 2006.

[15] R. R. Choudhury and N. H. Vaidya, "Deafness, a MAC Problem in Ad Hoc Networks when Using Directional Antennas," in *10th IEEE International Conference on Network Protocols (ICNP)*, Berlin, Germany, Oct. 2004.

[16] H. Gossain and C. Cordeiro and D. Cavalcanti and D. P. Agrawal, "The Deafness Problems and Solutions in Wireless Ad Hoc Networks using Directional Antennas," in *IEEE GLOBECOM*, Dallas, TX, US, Nov.–Dec. 2004.

[17] T. ElBatt and T. Anderson and B. Ryu, "Performance Evaluation of Multiple Access Protocols for Ad hoc Networks Using Directional Antennas," in *IEEE WCNC*, New Orleans, US, Mar. 2003.

[18] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad Hoc Networking With Directional Antennas: A Complete System Solution," *IEEE Journal on Selected Areas in Communications*, pp. 496–506, Mar. 2005.

[19] L.Bao and J.J.Garcia-Luna-Aceves, "Transmission Scheduling in Ad Hoc Networks with Directional Antennas," in *ACM MobiCom*, Lausanne, Switzerland, Jun. 2002.

[20] P. Mohapatra and S. V. Krishnamurthy, Eds., *Ad Hoc Networks: Technologies and Protocols*. Boston, MA, US: Springer, 2005.

[21] P. Ioannides and C. A. Balanis, "Uniform Circular Arrays for Smart Antennas," *IEEE Antennas and Propagation Magazine*, pp. 192–206, Aug. 2005.

[22] S. Verdù, *Multiuser Detection*. New York, NY, USA: Cambridge University Press, 1998.

[23] E. Royer and C-K.Toh, "A Review of Current Routing Protocols for Ad Hoc Wireless Networks," *IEEE Personal Communications Magazine*, pp. 46–55, Apr. 1999.

[24] H. L. V. Trees, *Optimum Array Processing*. New York, NY, US: John Wiley, 2002.

[25] S. Bellofiore, J. Foutz, R. Govindarajula, I. Bahceci, C. A. Balanis, A. S. Spanias, J. M. Capone, and T. M. Duman, "Smart Antenna System Analysis, Integration and Performance for Mobile Ad-Hoc Networks (MANETs)," *IEEE Transactions on Antennas and Propagation*, pp. 571–580, May 2002.

[26] A. Munari, *A New Cooperative MAC Protocol for Mobile Ad Hoc Networks Exploiting Directional Antennas (M.Sc. Thesis*. Padova, Italy: University of Padova, 2006.

[27] "Opnet Modeler," http://www.opnet.com.

[28] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity, Part I: System Description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, Nov. 2003.

[29] ——, "User Cooperation Diversity, Part II: Implementation Aspects and Performance Analysis," *IEEE Transactions on Communications*, vol. 51, pp. 1939–1948, Nov. 2003.

[30] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[31] L. Lai, K. Liu, and H. E. Gamal, "The Three-Node Wireless Network: Achievable Rates and Cooperation Strategies," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 805–820, Mar. 2006.

[32] T. E. Hunter and A. Nosratinia, "Cooperative Diversity Through Coding," in *IEEE ISIT*, Lausanne, Switzerland, Jun. 2002.

[33] T. E. Hunter, S. Sanayei, and A. Nosratinia, "Outage Analysis of Coded Cooperation," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 375–391, Feb. 2006.

[34] A. Stefanov and E. Erkip, "Cooperative Coding for Wireless Networks," *IEEE Transactions on Communications*, vol. 52, no. 9, pp. 1470–1476, Sep. 2004.

[35] S. Moh, C. Yu, S.-M. Park, H.-N. Kim, and J. Park, "CD-MAC: Cooperative Diversity MAC for Robust Communication in Wireless Ad Hoc Networks," in *IEEE ICC 2007*, Glasgow, UK, Jun. 2007.

[36] B. R. Hamilton and X. Ma, "Noncooperative Routing with Cooperative Diversity," in *IEEE ICC 2007*, Glasgow, UK, Jun. 2007.

[37] C. Cetinkaya and F. Orsun, "Cooperative Medium Access Protocol for Dense Wireless Networks," in *Third Annual Mediterranean Ad Hoc Networking Workshop*, Bodrum, Turkey, Jun. 2004.

[38] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. Panwar, "CoopMAC: A Cooperative MAC for Wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 340–354, Feb. 2007.

[39] H. Shan, W. Zhuang, and Z. Wang, "Distributed Cooperative MAC for Multihop Wireless Networks," *IEEE Communication Magazine*, pp. 126–133, Feb. 2009.

[40] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, Jul. 2000.

[41] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, Feb. 2003.

[42] P. A. Chou, T. Wu, and K. Jain, "Practical Network Coding," in *Allerton Conference on Communications*, Monticello, IL, US, Oct. 2003.

[43] J. Widmer and J.-Y. L. Boudec, "Network Coding for Efficient Communication in Extreme Networks," in *ACM SIGCOMM*, Philadelphia, PA, US, Aug. 2005.

[44] Y. Chen, S. Kishore, and J. Li, "Wireless Diversity through Network Coding," in *IEEE WCNC*, Vancouver, BC, Canada, Jun. 2006.

[45] S. Yang and R. Koetter, "Network Coding over a Noisy Relay: a Belief Propagation Approach," in *IEEE ISIT 2007*, Nice, France, Jan. 2007.

[46] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello Jr., "A Network Coding Approach to Cooperative Diversity," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3714–3722, Oct. 2007.

[47] X. Bao and J. Li, "Adaptive Network Coded Cooperation (ANCC) for Wireless Relay Networks: Matching Code-on-Graph with Network-on-Graph," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, pp. 574–583, Feb. 2008.

[48] E. Fasolo, F. Rossetto, and M. Zorzi, "Network Coding Meets MIMO," in *NetCod 2008*, Hong Kong, China, Jan. 2008.

[49] ——, "On Encoding and Rate Adaptation for MIMO_NC," in *IEEE 3rd International Symposium on Communications, Control and Signal Processing*, Malta, Mar. 2008.

[50] B. Nazer and M. Gastpar, "Computation over Multiple Access Channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[51] S. Katti, I. Maric, A. Goldsmith, D. Katabi, and M. Medard, "Joint Relaying and Network Coding in Wireless Networks," in *IEEE ISIT 07*, Nice, France, Jun. 2007.

[52] C. Peng, Q. Zhang, M. Zhao, and Y. Yao, "On the Performance Analysis of Network-Coded Cooperation in Wireless Networks," in *IEEE INFOCOM 2007*, Anchorage, AK, US, May 2007.

[53] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *ACM SIGCOMM 07*, Kyoto, Japan, 27-31 Aug. 2007.

[54] L. Zheng. and D. Tse, "Diversity and Multiplexing: a Fundamental Tradeoff in Multiple Antenna Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, Mar. 2003.

[55] H. Vikalo and B. Hassibi, "On Joint Detection and Decoding of Linear Block Codes on Gaussian Vector Channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3330–3342, Sep. 2006.

[56] E. S. Sousa and J. A. Silvester, "Optimum Transmission Ranges in Direct Sequence Spread Spectrum Multihop Packet Radio Network," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 5, pp. 762–770, Jun. 1990.

[57] H. S. Wang and N. Moayeri, "Finite State Markov Channel - a Useful Model for Radio Communication Cchannels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[58] F. Calì, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[59] M. Zorzi and R. R. Rao, "Error Control and Energy Consumption in Communications for Nomadic Computing," *IEEE Transactions on Computers*, vol. 46, no. 3, pp. 279–289, Mar. 1997.

[60] IEEE LAN MAN Standards, *Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, ANSI/IEEE Std., Apr. 2009.

[61] A. Varga, "Omnet++," http://www.omnetpp.org, 2001.

[62] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397–413, Aug. 1993.

[63] R. Jain, W. Hawe, and D. Chiu, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems", institution =," Tech. Rep.

[64] A. Munari, F. Rossetto, and M. Zorzi, "On the Viability of a Cooperative-Network Coding Protocol in Clustered Networks," in *IEEE MILCOM*, San Diego, CA, US, Nov. 2008.

[65] S. A. Mahmud, S. Khan, H. Al-Raweshidy, and K. Sivarajah, "Meshed High Data Rate Personal Area Networks," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 1, pp. 58–69, Jan. 2008.

[66] IEEE LAN MAN Standards, Part 15.3: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High-Rate Wireless Personal Area Networks (WPANs), *year = 2003*, ANSI/IEEE Std.

[67] J. Karaoguz, "High-Rate Wireless Personal Area Networks," *IEEE Transactions on Mobile Computing*, pp. 96–102, Dec. 2001.

[68] F. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II–The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1417–1433, Dec. 1975.

[69] A. C. Valera, W. K. G. Seah, and S. V. Rao, "Improving Protocol Robustness in Ad Hoc Networks Through Cooperative Packet Caching and Shortest Multipath Routing," *IEEE Transactions on Mobile Computing", volume =4, number = 5, month=*.

[70] B. Zhao and M. Valenti, "Practical Relay Networks: a Generalization of Hybrid-ARQ," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 7–18, Jan. 2005.

[71] M. Levorato, A. Munari, and M. Zorzi, "On the Effectiveness of Cooperation in Carrier Sense Based Ad Hoc Networks," in *IEEE SECON*, Rome, Italy.

[72] S. Tomasin, M. Levorato, and M. Zorzi, "Analysis of Outage Probability for Cooperative Networks with HARQ," in *IEEE ISIT*.

[73] A. Nosratinia and T. E. Hunter, "Grouping and Partner Selection in Cooperative Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 369–378, Feb. 2007.

[74] Y. Jing and H. Jafarkhani, "Single and Multiple Relay Selection Schemes and their Achievable Diversity Orders," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.

[75] Z. Lin, E. Erkip, and A. Stefanov.

[76] H. Lichte, S. Valentin, H. V. Malm, H. Karl, A. B. Sediq, and I. Aad, "Rate Per-Link Adaptation in Cooperative Wireless Networks with Multirate Combining," in *IEEE ICC*, Dresden, Germany, Jun. 2009.

[77] C. Perkins and E. Royer, "Ad Hoc On Demand Distance Vector Routing," in *IEEE WMCSA*, New Orleans, LA, US, Feb. 1999.

[78] D. Johnson, D. Maltz, and J. Broch, *DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks*, ser. Ad Hoc Networking. Addison-Wesley, 2001.

[79] S. Mueller, R. P. Tsang, and D. Ghosal, *Multipath Routing in Mobile Ad Hoc Networks: Issues and Challenges*, ser. Lecture Notes in Computer Science. Springer Berlin, 2004, pp. 209–234.

[80] M. Zorzi and R. R. Rao, "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 337–348, 2003.

[81] D. Estrin, R. Govindan, J. Heidmann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," in *IEEE/ACM Mobicom*, Seattle, WA, US, 1999, pp. 263–270.

[82] Z. Haas, "Design Methodology for Adaptive and Multimedia Networks," *IEEE Communication Magazine*, vol. 39, no. 11, pp. 106–107, Nov. 2001.

[83] P. Coronel, R. Doss, and W. Schott, "Geographic Routing with Cooperative Relaying and Leapfrogging in Wireless Sensor Networks," in *IEEE GlobeCom*, Washington DC, US, Nov. 26-30 2007.

[84] P. Coronel, S. Furrer, W. Schott, and B. Weiss, "Indoor Location Tracking Using Inertial Navigation Sensors and Radio Beacons," in *The Internet of Things (IOT 2008), LNCS 4952, Springer*, Zurich, Switzerland, Mar. 2008.

[85] D. Tacconi, I. Carreras, D. Miorandi, F. Chiti, and R. Fantacci, "Supporting the Sink Mobility: a Case Study for Wireless Sensor Networks," in *IEEE ICC*, Glasgow, UK, Jun. 2007.

[86] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang, "A Two Tier Data Dissemination Model for Large-Scale Wireless Sensor Networks," in *ACM MobiCom*, Cologne, Germany, Aug. 2005.

[87] Q. Chen, S. Kanhere, M. Hassan, and K. Lan, "Adaptive Position Update in Geographic Routing," in *IEEE ICC*, Istanbul, Turkey, Jun. 2006.

[88] B. Karp and H. Kung, "GPSR: Greedy Perimeter Stateless Routing for wireless networks," in *ACM Mobicom*, Boston, MA, US, Jun. 2000.

[89] R. Singer, "Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets," in *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 6, Jul. 1970.

[90] P. Maybeck, *Stochastic Models, Estimation and Control*, 2nd ed. Academic Press, New York, 1979.

[91] C. Karlof and D. Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures," in *IEEE GlobeCom*, San Francisco, CA, US.

[92] S. Zhu, S. Setia, and S. Jajodia, "LEAP: Efficient Security Mechanisms for Large-Scale Distributed Sensor Networks," in *ACM CCS*, Washington D.C, US.

[93] B. Deb, S. Bhatnagar, and B. Nath, "ReInForM: Reliable Information Forwarding using Multiple Paths in Sensor Networks," in *IEEE LCN*, Bonn, Germany.

[94] S. Lee and Y. Choi, "A Resilient Packet Forwarding Scheme Against Malicious Packet-Dropping nodes in Sensor Networks," in *SANS*, Orlando, FL, US, Feb. 2006.

Acknowledgments

I would like to thank in the first place my advisor, Prof. Michele Zorzi, for giving me the chance of pursuing my Ph.D., as well as for all the other career opportunities he has prompted me throughout these years. His suggestions, corrections and advices have played a key role as much in the completion of this thesis as in my professional growth.

A special thanks goes to Dr. Wolfgang Schott, whose tireless activity has gone way beyond a simple supervision of my work. All the technical and personal discussions we had during our collaboration have been priceless to me, and his precise attitude for work always supported and accompanied by a sincere passion for research will always stand as a model for me.

I would also like to thank Dr. Pierre Chevillat for the unique opportunity he offered me to join the IBM Zurich Research Laboratory, and all the other members of the Energy Management and Sensors System group, Linh Troung, Beat Weiss, Walter Hirt, together with Urs Hunkeler and Tomas Tuma, for the great moments we shared in having fun as well as in facing common challenges.

My gratitude goes to all the members of the SIGNET group at the University of Padova, whom I regard as friends more than colleagues. In particular, a warm thank you goes to Federico, Paolo and Angela: I could simply and sincerely not think of better office-mates.

Finally, a special thanks goes to Francesco Rossetto for a number of professional reasons but, most of all, for the great friend he has always been.

Ringraziamenti

Il mio ringraziamento va alla mia famiglia e a tutti gli amici che mi sono stati vicini nel corso di questi anni. La certezza di avere al mio fianco persone splendide su cui poter sempre contare ha reso infinitamente più semplice arrivare a questo traguardo, e ha ridimensionato tutte le difficoltà che si sono via via presentate. Non nominerò qui nessuno, ma sono certo che ciascuno di voi, leggendo queste poche righe, si sentirà chiamato in causa, e saprà bene i mille motivi per cui gli sono, e gli sarò sempre, grato.

Andrea