9

Scheduling, Routing, and Related Cross-Layer Management through Link Activation Procedures in Wireless Mesh Networks

Leonardo Badia[1], Alessandro Erta[1], Luciano Lenzini[2], and Michele Zorzi[3]

[1] IMT Lucca Institute for Advanced Studies, Piazza S. Ponziano 6, 55100 Lucca, Italy. {l.badia, a.erta}@imtlucca.it
[2] Dept. of Information Engineering, University of Pisa, via Diotisalvi 2, 56122 Pisa, Italy. l.lenzini@iet.unipi.it
[3] Dept. of Information Engineering, University of Padova, via Gradenigo 6B, 35131 Padova, Italy. zorzi@dei.unipd.it

9.1 Introduction

In a Wireless Mesh Network (WMN) [1] end users are provided with wireless broadband connectivity by means of a pre-defined system hierarchy. To describe this organization, several notations can be used. In the following, we adopt the terminology of [2]. The end terminals, also referred to as Mesh Clients (MCs), are connected to special nodes, denoted as Mesh Routers (MRs). These nodes do not generate traffic, since they are simply meant to relay the packets of their MCs. Additionally, some MRs, called Mesh Access Points (MAPs), can be provided with a wired connection, and can therefore act as gateways toward the Internet. The MAPs are also wirelessly interconnected to all the other MRs in a multi-hop fashion, without necessarily following pre-defined paths. Instead, an MC can interact only with the MR it is connected to. MRs form what is usually named as the *backbone* of the WMN, which can physically cover a large region in a wireless manner. This structure offers a good costs/benefits balance, since it almost entirely avoids cable set up. For this reason, it is deemed to be applicable in rural areas, where the deployment of wireline networks may be too expensive. WMNs can also be envisaged for dense residential or business areas, and in general anywhere the installation of cables is difficult because of physical obstacles.

There are several possibilities to specify the Medium Access Control (MAC) used by a WMN. These are often related to existing standards, especially IEEE 802.11 [3] and IEEE 802.16 [4], parts of which are dedicated to WMNs. Actually, the first hop from any MC to its related MR is often assumed to employ a radio access interface different from the one used in the backbone, and entirely *orthogonal* (i.e., perfectly non-interfering) to it, e.g., since it uses another frequency, and possibly another technology. Moreover, the first hop may adopt management strategies typical of cellular networks [5], and is therefore conceptually simpler. For this reason, we will not investigate this part of the WMN in greater detail. Conversely, realizing the interconnections among MRs poses many theoretical challenges, most of which are common to all kinds of multi-hop networks, such as Ad Hoc and Sensor networks. However, when revising them for WMNs, some important properties come into play. Usually, MCs can be portable devices, whereas MRs and MAPs are not mobile. Therefore, the backbone management does not suffer from most mobility issues, neither at the transport layer (i.e., paths do not need to be updated), nor at the physical layer (channel variability is relatively moderate). Moreover, communications in a WMN are usually to or from the Internet, thus all routes have either the source or the destination in a MAP. Finally, as MRs can be easily placed near to a power outlet, energy saving is not an issue. These properties considerably distinguish the backbone of WMNs from an Ad Hoc network (for what concerns pre-defined hierarchy and absence of mobility) or a Sensor network (lack of terminal battery limitations).

The issues which arise in the backbone management relate to different layers of the protocol stack. On the one hand, the creation of low-interference and high-rate paths to the MAPs is key to achieve good rates at each MR. This may also involve the exploitation of multiple channels as, for example, MRs can own several Network Interface Cards (NICs), which can simultaneously operate on different frequencies. On the other hand, the link layer needs to schedule packets over multiple links in order to achieve good transmission parallelism and possibly forward more data towards the MAPs at the same time.

The main problems which will be investigated by our analysis are:

- routing algorithms, i.e., network level procedures to discover efficient paths which connect the ordinary MRs (and therefore their MCs) to the MAPs. Note that routing strategies designed for Ad Hoc Networks usually admit also peer-to-peer communications, which are not common for WMNs. Moreover, the goal in WMNs is more often to obtain high system throughput rather than maximizing battery lifetime.
- link scheduling, which involves medium access level procedures to activate communication links. Its goal is to ensure network connectivity while at the same time satisfying physical constraints related to technology, interference and network management.
- cross-layer management, operating at an intermediate level with both network and link layer procedures, jointly addressing these problems.

The aforementioned issues involve other related topics, which are also worth discussing. In certain cases very broad subjects are involved, which will be discussed here only for what concerns their impact on the definition of routing and scheduling strategies. There are also other aspects of these matters which fall out of the scope of the present article and therefore will not be discussed here in detail. However, the reader will be addressed to external references to find further material on them. Some of the related problems which will be framed into our analysis are:

- channel assignment and node placement: in our analysis, these are considered to be aspects of network deployment, which means they have already been performed at the time routing or scheduling strategies are sought. However, it will be briefly outlined how it is possible to incorporate them into the same cross-layer framework used for routing and scheduling with a modular approach, thus with no need for significant modifications of the reasonings presented in the rest of the article.
- models of wireless interference; for this point, two important considerations must be made. First of all, we propose a detailed review and classification of the possible approaches to characterize interference. We try to resolve terminology ambiguity due to the use, in the literature, of different names for the same model or of the same name for distinct models. Moreover, we discuss the choice of the model itself, which is driven by two contrasting aspects. On the one hand, the interference model should be as accurate as possible. In this sense, the use of heavily simplified interference models may end up in poor algorithm performance when applied to realistic cases. On the other hand, a certain degree of approximation is unavoidable as related to the properties of the Medium Access Control (MAC) protocol. In fact, in a layered network management, algorithms operating on top of the link layer necessarily abstract some aspects of the physical layer, such as interference. For these reasons, we will first concentrate our analysis on general results which hold true regardless of the interference model, such as theoretical performance bounds. Then, we will discuss how these findings translate to practical cases, at which point different interference models need to be taken into account.

The rest of this paper is organized as follows. In Section 9.2 we give a brief overview of the problem studied and we clarify terminology and notations employed in the rest of the paper. In Section 9.3 we present a review of the papers which discussed related topics in a way applicable to WMNs. In Section 9.4 we mathematically formalize the problem, in particular identifying the constraints determined by capabilities of the terminals and wireless interference. This latter aspect, in particular, is discussed in depth in Section 9.5, proposing a classification of interference models, and also touching MAC protocol issues. In Section 9.6 we give both theoretical and practical evaluations of the performance of WMNs. Even though the problem is NP-complete and exact approaches are hard, we present some original analytical results which determine both upper and lower performance bounds, and we give quantitative insights by applying them to sample WMN topologies. Finally, we present the conclusions.

9.2 Preliminaries

We represent the backbone of a WMN as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. The *nodes* in set $\mathcal{N}$ are the MRs, which are in turn connected by the *edges* belonging to set $\mathcal{E} \subseteq \mathcal{N}^2$, thus representing the communication links of the backbone. This approach is commonly used for multi-hop wireless networks [6, 7], even though the graph is often considered bi-directional, i.e., with undirected edges. This is a limiting assumption, as will be discussed in Subsection 9.5.1. Similarly to [8, 9], we will assume instead that the edges, as actual wireless communication links, are uni-directional. Thus, the communication link where a sender node $i \in \mathcal{N}$ transmits to a receiver $j \in \mathcal{N}$ is represented by an element $e \in \mathcal{E}$ equal to the ordered pair (i, j). The inclusion of this link in $\mathcal{E}$ actually happens only if node j can receive a transmission from i in the absence of any other interference source.

In the following, we will denote with $\mathcal{R}_i$ and $\mathcal{S}_i$ the set of nodes which are possible receivers from and senders to node i, respectively. In other words, $\mathcal{R}_i$ and $\mathcal{S}_i$ contain the one-hop output and input neighbors of i. Formally:

$$\mathcal{R}_i = \{j \in \mathcal{N} : (i, j) \in \mathcal{E}\} \tag{9.1}$$

$$\mathcal{S}_i = \{j \in \mathcal{N} : (j, i) \in \mathcal{E}\} \tag{9.2}$$

We will also refer to other properties of the communication link represented by edge $(i, j) \in \mathcal{E}$. To quantify the *capacity* of the link we make use of variables r_{ij}, called *link rates* and collected into a matrix $\mathbf{R} = (r_{ij})$. Rate r_{ij} can be regarded as the number of bits which can be transmitted over the link represented by edge (i, j) in a given time unit. When required by physical specifications, we will also consider a parameter g_{ij} corresponding to the wireless link gain[1] over (i, j). A matrix $\mathbf{G} = (g_{ij})$ can be introduced collecting the g variables for all edges.

In our investigations, we consider an underlying Space and Time Division Multiple Access (STDMA) scheme [10]. For wireless multi-hop networks it is in fact crucial to exploit space and time parallelism in order to obtain an efficient transmission scheme.

Our mathematical representation of scheduling and routing over the WMN backbone is similar to the ones reported in [6, 8, 11]. A link represented by edge $(i, j) \in \mathcal{E}$ is said to be *active* if node i transmits to node j. Thus, for any edge $e \in \mathcal{E}$ of the graph, we define a binary variable $x_e(t)$, which varies over a discrete (slotted) time and indicates activation of the corresponding link at time t, i.e., $x_e(t) = 1$ if the link is active and $x_e(t) = 0$ otherwise. By varying t, the activation variables $x_e(t)$ determine a time-division scheduling for the WMN backbone according to what we refer to in the following as

[1] The wireless link gain is the ratio between received and transmitted power. It is well known that wireless channels are strongly time-varying. However, for simplicity, we will consider slowly varying scenarios where the g_{ij} parameters can be approximated as constants.

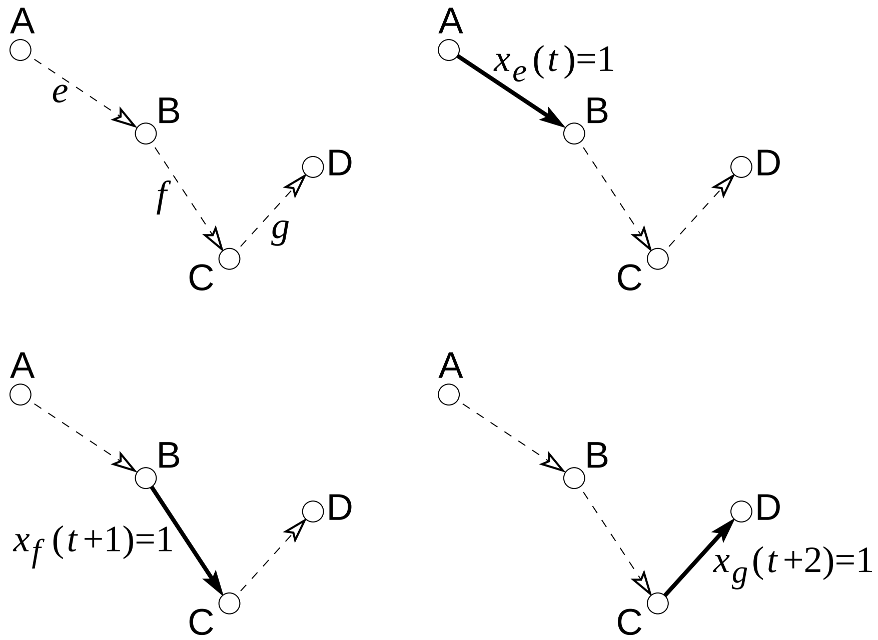

Fig. 9.1. Example of route obtained through link activation.

link activation pattern. Similarly to the analysis presented in [11, 12, 13], we remark that the derivation of the scheduling through a link activation pattern implicitly determines the routing as well. This is visible, for example, in Fig. 9.1, where a packet needs to be sent from A to D. Assume that nodes B and C do not have packets to send themselves and can act as relays. A route is created from node A to node D by subsequently allocating links e (from A to B), f (from B to C), and g (from C to D). Note that the entire route is actually realized by operating over three time slots.

To be efficient, such an STDMA link activation scheme needs to be aware of the network topology. This is a strong requirement in many kinds of wireless multi-hop networks, where nodes are mobile, but, as the backbone usually consists of fixed nodes, this is not much of an issue for WMNs. Moreover, finding an efficient STDMA link activation pattern has the drawback of being computationally expensive. However, this can be done by a centralized unit (e.g., located in one of the MAPs, which are usually the most computationally capable among the MRs), which determines a proper transmission schedule and communicates it to the other nodes. This can be realized by broadcast messages or by piggy-backing this information in other control messages.

In the following, we will specify modalities according to which the 0–1 decision variables corresponding to link activation can be determined. In particular, it is not restrictive to focus on an *uplink* problem, i.e., on how to deliver a given amount of packets, known a priori, *from* all MRs *to* any of the

MAPs in the shortest time. This problem can be generalized to a downlink problem (i.e., to activate links so as to deliver traffic from any of the gateways to all MRs), which is conceptually identical. In fact, the downlink problem can be solved by looking at an equivalent uplink problem with reversed delivery requirements (i.e., where packets are to be sent from nodes to gateways instead of the opposite). In the uplink problem, since we assume a directed graph, we should also reverse flow directions and link parameters (e.g., g_{ij} must be changed into g_{ji}). The link activation pattern found for the uplink problem can be flipped over time to obtain the solution to the downlink problem.

However, activation variables $x_e(t)$ can not take arbitrary values. The management of link activations should satisfy feasibility conditions related to the physical nature of the problem. Among the key points which will be discussed in the following, we highlight here the interference requirements, which forbid certain links from being simultaneously activated, since some of the resulting transmissions will not be successful. Considerations about interference are also often coupled with MAC protocol issues. In fact, as shown in many contributions (e.g., in [14]), in a centralized environment deterministic access provided by an STDMA scheme obtains better performance than random access schemes such as the Distributed Coordination Function (DCF) of the IEEE 802.11 MAC. However, our STDMA scheduling might simply be a deterministic link activation pattern superimposed onto an underlying MAC protocol, which is designed for distributed and random access. Indeed, several contributions [6, 12, 15, 16, 17], make (explicitly or implicitly) this assumption and for this reason combine interference and MAC protocol issues when determining the compatibility of simultaneous link activation.

The most widely used classification of interference models in the literature dates back to [18] and distinguishes between the so-called *physical* and *protocol* interference models. In the former, the feasibility of simultaneous link activations is determined by the Signal-to-Interference-Ratio (SIR) of all receivers being above a given threshold. The latter imposes instead simpler interference conditions modeled through graph neighborhood relationships. Actually, more than of a single model, we should speak of *protocol models*. In fact, the protocol model was originally intended to represent the IEEE 802.11 MAC protocol (hence the name), but in certain papers a slightly different implementation can be found, especially when IEEE 802.16 is used instead, even though the interference model is still called the same. These issues will be discussed in detail in the following, and we will present an original classification which also aims at solving some terminology inconsistencies.

In addition to these two classes there is another possible approach, i.e., to directly estimate the interference, e.g., by measuring it in the scenario of interest [19], or through higher layer statistics [20]. Since we take an a priori approach to interference characterization, we will not discuss this *measurement-based* interference model further. However, it is worth mentioning as the one which is, in a sense, adopted by some related contributions, especially those dealing with routing metrics, e.g., [20, 21].

Table 9.1. Taxonomy of related work.

Reference	Schd	Rout	ChAs	interf.	approach
Alicherry *et al.* [6]	✓	✓	✓	P	O,A
Tang *et al.* [7]		✓	✓	P	O
Cruz, Santhanam [8]	✓	✓		Φ	O
Brar *et al.* [10]	✓			Φ	A
Kodialam, Nandagopal (1) [12]	✓	✓			T,O,A
Kodialam, Nandagopal (2) [13]	✓	✓	✓	P	T,O,A
Jun, Sichitiu [15]	✓			P	T
Ben Salem, Hubaux [17]	✓			P	T,A
Draves *et al.* [20]		✓		M	O
Yang *et al.* [21]		✓	✓	M	T
Wu *et al.* [22]		✓	✓	M	A
Salonidis, Tassiulas [23]	✓			P	A
Djukic, Valaee [24]	✓			P	A
Jain *et al.* [25]	✓	✓		P	O
Cao *et al.* [26]	✓	✓			O
Wei *et al.* [27]	✓	✓		P	O,A
Subramanian *et al.* [28]	✓			Φ	A

Schd = scheduling, Rout = routing, ChAs = channel assignment
For "interf."(=interference): P=protocol model, Φ=physical model, M=measurement.
For "approach": O=optimization framework, A=practical algorithm, T=theoretical results.

9.3 State of the Art

There is a vast literature in the field of wireless networks. The increasing interest for WMNs has recently brought researchers to revise typical issues of wireless networks in the context of this emerging technology. Specifically, traditional research topics such as link scheduling, routing, channel assignment and topology control find in WMNs new challenges and applications, as WMNs raise challenges and problems which need new solutions as the existing ones do not apply directly.

In this section, we provide an exhaustive up-to-date review of the literature on routing, scheduling and related cross-layer approaches for WMNs. Scientific papers are classified according to the investigated research topics so as to guide the reader to the contributions of interest. In Table 9.1, we report a taxonomy of the reviewed papers. For each paper, the table indicates the research issues addressed, the assumption about the interference model and the proposed approach.

The rationale for TDMA scheduling over WMNs can be derived from the very general approach for multi-hop wireless networks presented in [11]. In [23], the authors propose a distributed implementation of such an approach for Ad Hoc Networks. However, the resulting rationale can be applied, with minor modifications, to WMNs also. In the paper, a fluid model is proposed to quantify link activations, and the resulting evaluations are used by the terminals so as to share the medium in a fair and entirely distributed manner. Wireless interference is characterized through the protocol model. Another related approach to address TDMA scheduling for a WMN is presented in [24], where again the protocol model is used.

This same interference model is also used in two different papers, [15] and [17], where wireless mesh scheduling is investigated from a theoretical point of view. Among the contributions presented in these papers, we highlight in particular that the former gives a lower bound on the length of the optimal WMN link activation pattern, whereas the latter determines an upper bound on the same value, and proposes a fair scheduling mechanism. In the following sections, we will revisit the theoretical results of these papers and extend them so that they can be applied in a more general way, i.e., with any interference model.

Another paper dealing with scheduling in WMNs is [26], where the specific case of IEEE 802.16 Mesh mode operating with centralized scheduling is addressed. Here, an optimization framework is presented to maximize system throughput under specific fairness constraints. However, no interference model is presented, since the link allocation is only limited by what will be referred to in the following as half-duplex constraint. A further point of interest of this paper is that the authors consider a Pareto dominance approach to compare scheduling solutions and find the optimal one.

The limitations imposed by oversimplified interference models heavily affect the scheduling, as shown in [10]. The main contribution of this paper is to show that assuming the protocol interference model may lead to inefficiencies in the scheduler implementation, whereas taking the physical interference model into account can achieve better network performance. To this end, a fast heuristic algorithm is proposed which assumes pre-determined traffic weights on each link (which, e.g., can come from a routing algorithm executed a priori).

Like scheduling, routing is also a challenging task in WMNs. In this scenario, several papers have investigated the task of properly defining metrics to be used in routing algorithms [20, 21, 22, 28]. In [20], the authors introduced a routing metric which is computed through the estimation of the interference of the links belonging to a path by means of delay probes. This approach has been extended in [21] and [22] to the multi-channel case by also including the channel assignment problem. The former paper uses a theoretical approach, whereas the latter presents a practical algorithm supported by experimental results. Finally, [28] includes interference awareness considerations in the computation of the routing metric, by utilizing the physical interference model.

Topology control considerations are included in the routing investigation performed in [7]. In this paper, optimality conditions to derive routing under QoS constraints are studied for a multiple channel network. The protocol interference model is used.

In general, standard solutions based on shortest-path algorithms are very likely not to be suitable for WMNs [21]. In fact, routing metrics based on the minimum hop count may have poor performance because they try to exploit wireless links between distant nodes. These long wireless links can be slow and lossy, leading to poor throughput. Furthermore, the objective of a traditional shortest-path routing algorithm is usually in contrast with that of link scheduling algorithms. Assuming a predefined path between a source and a destination implies that any link scheduling algorithm is forced to activate only the links belonging to that path. The link scheduling may thus result sub-optimal in the sense that any scheduling algorithm is prevented from optimizing the exploitation of the available network resources.

A pipelined approach which addresses both routing and scheduling is considered in some recent papers. In [27] scheduling and routing are performed. The scenario is specifically an IEEE 802.16 Mesh operating with centralized scheduling. Here, a two-step procedure is proposed. First, a route selection algorithm identifies low interference paths toward the destination. Then, scheduling is performed among the routes by considering compatible link activations according to the protocol interference model.

However, as shown in [25], scheduling and routing algorithms impact on each other and their optimality is strongly coupled. In particular, after a review of interference models, [25] addresses the question of combining optimal link scheduling with sub-optimal routing and vice versa. The main conclusion is that interference-awareness is also beneficial at the routing level. In a more general sense, this also implies that a joint optimization of routing and scheduling [11] is the most preferable solution.

An example of framework for joint scheduling and routing has been described in [12], where the authors introduce a heuristic technique to solve the joint routing/scheduling problem. Specifically, routing and scheduling are solved as optimization problems over an undirected graph. The authors consider communication links as compatible if they respect what we call duplex constraints, i.e., the number of transmissions and receptions that nodes can simultaneously perform are limited. No additional interference constraint is considered. Necessary and sufficient conditions are then derived to guarantee the link scheduling feasibility. Therefore, for a given pair of nodes the objective is to determine the maximum achievable flow rate under the duplex constraints and the link scheduling feasibility conditions. This can be formulated as a linear programming problem. The proposed solution ensures that the link scheduling is feasible as the scheduling constraints are considered when solving the routing problem. The scheduling of each flow is then performed by coloring the network graph with a known graph coloring algorithm [29]. In [13], the authors extended their model to multiple channels and the protocol interference model. Specifically, they derive both necessary and sufficient conditions for a feasible channel assignment and scheduling in a multi-radio network. Again, the channel assignment problem is modeled as a linear optimization problem. Additionally, a heuristic algorithm is proposed for solving the problem. A similar approach has been proposed in [6]. In this paper, the authors mathematically formulate a joint channel assignment and routing framework, taking into account the protocol model interference constraints, the number of channels in the network, and the number of radios available at each mesh router. Within this framework, they devise a heuristic to perform routing and channel assignment aimed at optimizing the network throughput performance.

Finally, in [8] a joint analysis of routing and scheduling for multi-hop networks is presented, which also includes power control. Another interesting aspect of this paper is that it addresses half-duplex limitations of the wireless medium, as well as directionality of links and the physical interference model. However, the paper does not directly investigate WMNs, but rather it mainly focuses on systems similar to Ad Hoc or Sensor Networks, since the objective of the optimization is the minimization of the power consumption, which is not an issue in WMNs.

To sum up, joint and cross-layer approaches have been proposed by several papers dealing with routing and scheduling for WMNs, but the formulation of an overall framework which encompasses all of these issues is still an open field of research. Existing approaches are often unsuitable for WMNs due to dissimilar optimization goals and/or oversimplified interference models. The formulation of a comprehensive framework for these issues, also addressing technological issues in a realistic manner and correctly taking into account link directionality, duplex constraints and different possibilities for the inter-link interference model, is a promising goal for future research. As a first step in this direction, we will give in the following some guidelines and analytical insights on the performance of joint routing and scheduling in WMNs.

9.4 Problem statement

To study routing and scheduling under the graph formulation reported in Section 9.2, we will use the language of constrained linear programming problems, as this is an approach commonly used to decide the assignment of x_e variables [6, 12]. We will therefore speak of *constraints* to describe any limitation imposed to the activation of links by MAC and physical layers.

These constraints can be of different natures, and we will describe them in separate subsections. First of all, edge activation implies node activation for transmission and reception, for which there are limitations on the transceiver capabilities of each node involved. As will be discussed in Subsection 9.4.2, the activation of edge (i, j), which employs i and j as transmitter and receiver, respectively, may not be feasible, if these nodes participate to other link activations.

Moreover, links which involve different nodes for what concerns both the transmitter and the receiver, might or might not activate simultaneously depending on the mutual electro-magnetic interference. For this reason, we need to define a compatibility relationship among the links in the network. Several models for this will be reviewed in Subsection 9.5.3. In most cases, they can be subdivided into the two main classes of *protocol* and *physical* model, already mentioned in Section 9.2. To better understand the "protocol model", in Subsection 9.5.3 we will also briefly discuss the underlying assumptions of IEEE 802.11 and IEEE 802.16 standards for what concerns access control.

Prior to investigating in detail these constraints, we give an overview of other related problems which can be framed within our approach, which is the goal of Subsection 9.4.1.

9.4.1 Channel assignment and node placement framed into the model

Some issues discussed previously in Section 9.3 can be incorporated in our framework. For instance, it is common to assume that the wireless medium has several channels available for transmission. From a simplified point of view, these channels are often considered orthogonal [12, 22, 30] and it is further assumed that the MRs own different NICs so that they can communicate on many channels in parallel. A relevant point in this case is whether the terminals can rapidly change the channels on which their NICs are active. With current state-of-the-art technology [31], the order of magnitude of channel switching time can be 0.1 s, which is likely to be much higher than one time-slot; thus, we need to assume that the assignment is not modified during the schedule, and every node can be active only on certain channels of choice.

For this reason, the study of channel assignment in this case mostly relates to routing, and corresponds to identifying low-interference paths whose parallelism is further improved by the presence of orthogonal channels. In fact, links which would interfere if scheduled jointly can be activated together if their transmitters and receivers are tuned to different channels. The issue of channel assignment in the orthogonal case is therefore often seen as a graph-coloring problem, where colors assigned to edges represent orthogonal wireless channels. From the perspective of our link activation framework, the orthogonal multiple-channel assignment can be incorporated following a similar rationale, e.g., by defining variables $x_e^{(c)}(t)$, where the additional color index c spans over a set of channels $\mathcal{C}$ and denotes the channel possibly used by link e. This imposes additional constraints, i.e., that the number of activated channels for a node is less than or equal to a given parameter, corresponding to the number of NICs it owns, and that a link can be activated only if transmitter and receiver share a common active channel.

In general, the additional challenges imposed by the presence of multiple orthogonal channels are not further considered here, since they are out of scope of our analysis. Note only that, from a purely mathematical point of view, if frequency is considered as a perfectly separable resource, differences between frequency-division and time-division multiplexing are limited and they can be translated into our framework. For this reason, most of the conclusions we will draw in the time domain also holds for orthogonal multiple channel assignment.

An important observation raised, e.g., in [32], stems from the observation that, in real network systems, contiguous channels are not perfectly separated at the physical level, but are instead partially overlapping. In general, this is regarded as an undesired effect and to deal with it channels are assumed to have guard bands that are not used for transmission. It is, for example, usual to limit the use of IEEE 802.11 MAC to channels 1, 6, and 11, which can be considered as orthogonal with a good degree of accuracy, leaving the remaining channels unused [16]. However, an entirely different approach is used in [32] and related papers. These contributions show that the existence of partially overlapping channels, instead of being a problem, may turn into an advantage for the network if properly exploited. In particular, it is possible to partially obtain transmission parallelism even by using a single NIC. Intuitively speaking, this happens as the intended transmitter and receiver do not need to be tuned to the same channel, but they can choose two different partially overlapping channels. The choice of the channel to which a node tunes is therefore a trade-off between maximizing the overlap for useful connections and minimizing it for the interference it causes to other links when transmitting.

Such an extension to multiple overlapping channels can be framed in our joint routing and scheduling framework, even though it would require a long analysis which can not be reported here for space reasons. However, we consider it as a possible interesting subject for future research.

Another possible related investigation is the evaluation of the network deployment, especially for what concerns node placement (MRs and MAPs). In most of the related work it is assumed that the nodes' positions are decided a priori. The reason for this is twofold: on the one hand, it is realistic to think of network deployment as realized in a different design phase than routing and scheduling; on the other hand, it is also difficult to allow for an entirely free node placement, due to physical and environmental constraints, as well as the not-in-my-backyard problem. Nevertheless, it is still possible to allow a certain degree of choice without violating realism. This can be done by following the approach presented in [33], and adapted to WMNs in [34]. The problem statement is slightly changed, so that nodes of the graph no longer represent terminals but are instead *candidate positions* where terminals can be placed (in [33] terminals are UMTS base stations, whereas in the WMN case, they are MRs and MAPs). An additional binary decision variable y_n is introduced for every $n \in \mathcal{N}$ to denote whether position n is actually occupied by a terminal or not. The rest of the analysis proceeds identically, with the only modification of requiring any edge activation variable $x_{(i,j)}$ to be less than or equal to both y_i and y_j, as a communication link can be actually activated only if both its ends correspond to physically deployed terminals.

9.4.2 Transceiver constraints

Our graph-based approach determines a joint scheduling and routing through link activation. As communication links are represented through edges of the graph, most of the constraints are edge-based, i.e., they must be respected by every active edge. However, the first important constraint we discuss is node-based, i.e., it has to be evaluated at every node, and relates to the fact that the node capabilities for transmission and reception are limited. In particular, we focus here on narrowband channels, where it is not possible to receive simultaneously from multiple sources. We remark that special techniques, such as Wideband Code-Division Multiple Access (WCDMA) [35] or Multiple Input Multiple Output (MIMO) [36] channels, can improve this condition. However, they are out of the scope of our investigations. In the following, we therefore assume that at most one signal can be decoded, and any other transmission the receiver is able to listen to can only be regarded as interference. The presence of interference at the receiver does not necessarily mean that the packet can not be correctly decoded. As will be shown in the next subsection, the interference model comes into play at this point. If the protocol model is used, any superposition of signals will result in a collision, i.e., no packet can be received. In the physical model, the strongest received signal may still be successfully decoded. However, regardless of the interference model, the maximum number of possible simultaneous successful receptions is *one*.

A similar situation happens for the transmitter. Even though on the wireless medium it is possible to operate in a multicast fashion, i.e., from one transmitter to many receivers, in this case the *same* transmission takes place for all of them. Note also that multicast transmission, which would require additional specifications, e.g., for duplicated packet control, does not correspond to the problem we consider, where the intended destination is only one. For these reasons, we will assume in the following that multiple transmissions from the same node are forbidden. However, we remark that the issue of exploiting the possibility for some relay nodes to listen to the communication, even when they are not the intended receivers, to improve the network connectivity by exploiting cooperation [37] or network coding [38] is a very promising subject for future research in wireless networks.

Finally, not only can simultaneous transmissions and receptions be at most one, but also the wireless communication medium is intrinsically *half-duplex*, i.e., a node can not listen on the same channel on which it is transmitting at the same time, or the transmitted signal will jam any packet reception [39]. Possible solutions to this problem, so as to realize a sort of full-duplex communication with simultaneous transmission and reception at a node, can be to utilize more than one NIC to exploit the possible presence of multiple channels [16, 22], or to use multiple directional antennas [40, 41]. However, these techniques do not entirely solve the problem, as they obtain full-duplex capability at the price of additional resources. Moreover, they decrease network connectivity, which in certain cases can be an undesirable effect, as the nodes should use compatible channels or antenna beams. Finally, we remark that in the multiple channel case any NIC is still utilized in a half-duplex fashion, i.e., no simultaneous transmission and reception is still possible on the same channel. For these reasons, we impose that the activation of links should satisfy the constraint of not activating more than one operation (i.e., either a transmission or a reception) for each node. Formally, this constraint translates into:

$$\forall i \in \mathcal{N}, \forall t : \qquad \sum_{j \in \mathcal{S}_i} x_{ji}(t) + \sum_{j \in \mathcal{R}_i} x_{ij}(t) \leq 1 \tag{9.3}$$

Apparently, the importance of including the half-duplex aspect in this constraint is often underestimated when modeling multi-hop wireless networks. In fact, the need for such a constraint is rarely mentioned. This may be due to the fact that, as already emphasized, most of the investigations use the protocol interference model which, as discussed in the following, prevents simultaneous transmission and reception at the same node from happening. However, we believe that it is important to distinguish the edge-based interference constraints from the node-based duplexing limitation. Indeed, the interference constraint does not necessarily translate into the protocol model, which can be replaced, e.g., by the physical model. Instead, the duplexing limitation holds irrespective of the interference model. For this reason, we will always impose the half-duplex constraint in any problem formulation. Note also that our assumption in this respect might seem different from [12], where the authors allow for the possibility of using both directions of the link at the same time in what they call *full-duplex* case. However, this case is used in conjunction with the protocol interference model. In general, any case where full-duplex nodes are mentioned does not refer in reality to the possibility of transmitting and receiving on the same frequency *at the same time instant.*

Apart from this very general constraint, other limitations to the simultaneous activation of edges make specific assumptions on the nature of radio interference and on the underlying MAC protocols. We will review both these aspects in the following section.

9.5 Interference Models and Relationships with MAC Protocols

In Section 9.2 we mentioned the need for using uni-directional edges in our network graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. First of all, this subsection aims at motivating this choice in more detail. After this explanation, we outline some aspects of well-known MAC protocols and finally we review which model of mutual interference among nodes can be used to determine if the activation allows the correct reception of all transmitted packets.

9.5.1 Link directionality

The choice of using the directed graph representation, which captures the *anisotropic* nature of wireless links is surely more realistic from the physical point of view. In fact, wireless links are characterized by strong asymmetry [42]. Due to environmental limitations and also different power levels, it is even possible that two nodes i and j belonging to $\mathcal{N}$ are linked only one-way, i.e., $(i, j) \in \mathcal{E}$ but $(j, i) \notin \mathcal{E}$.

However, our choice is not only motivated by the desire to better adhere to reality. In fact, the frequent assumption of bi-directional links is in most cases due to the application of the analysis to IEEE 802.11 scenarios. As the IEEE 802.11 standard is supposed to work on entirely reliable links only, edges in $\mathcal{E}$ need to be the bi-directional, This also relates to the choice, which will be discussed in the next subsection, of modeling interference with the protocol interference model, in its implementation more closely related to IEEE 802.11.

Yet, the decrease in the problem complexity gained with bi-directionality assumption is marginal (the number of edges is only decreased by a constant factor of 2), and implies an oversimplification in modeling interference conflicts [9], especially when focusing on a centralized STDMA scheme, if an underlying IEEE 802.11 MAC is not employed. Instead, not only is the problem version with directed edges of $\mathcal{E}$ more accurate, but it also includes the undirected graph as a special case.

9.5.2 Overview of MAC protocols

To realize distributed medium access with low cost technology, random MAC protocols are often used. In particular, the IEEE 802.11 standard has obtained a great success for what concerns its DCF-based version, operating with *four way handshake*, which implies that the transmission is initiated after a successful request-to-send (RTS) and clear-to-send (CTS) exchange, and after the data transmission an acknowledgement (ACK) is also to be sent from the receiver to the transmitter.

However, IEEE 802.11 is known to suffer from many problems, which are severely limiting for WMNs. In fact, to operate in a totally distributed manner, IEEE 802.11 requires the transmission of many control packets, whose overhead is often heavy. In a WMN most of them are not necessary since most of the control can be centralized. Moreover, its collision avoidance mechanism often imposes unnecessary constraints which limit network parallelism, especially because it does not properly capture wireless interference. Finally, the main advantage of the conceptual simplicity and ease of implementation of the IEEE 802.11 MAC is not strictly required in the WMN backbone, which is composed of more expensive and technologically advanced terminals. Anyway, note that IEEE 802.11 can still be used to interface a MR with its MCs; however, as already said, this part of the network is not investigated in our analysis.

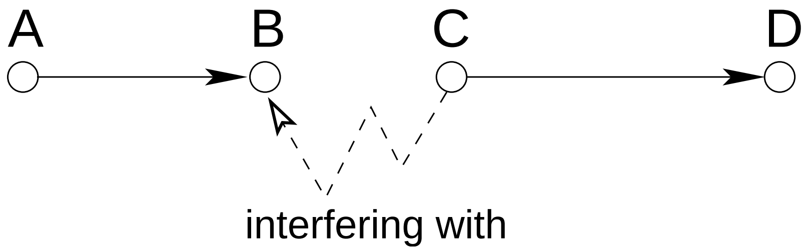

Fig. 9.2. A case of transmission showing hidden terminal problem.

Another standard which is envisioned to have applicability for WMN is IEEE 802.16 [4], which besides the point-to-multipoint (PMP) mode is also available in a Mesh mode. We focus in the following on the distributed scheduling version. IEEE 802.16 aims at partially solving some of the aforementioned problems as, unlike IEEE 802.11, it utilizes a random-based procedure in the control frame with a *three way handshake* procedure, where a Request is answered by a Grant, which is finally followed by a Confirm message from the transmitter. Part of the advantage of IEEE 802.16 stems from the additional requirement for topology awareness, which is exploited in the *distributed election* mechanism to guarantee that no collision arises in the control message exchange.

The reason for these protocols to include specific handshaking procedures, and possibly also further random decisions and exponential backoff algorithms, is to cope with the fact that nodes can operate in a distributed fashion. In fact, every random medium access protocol, therefore also IEEE 802.11 and IEEE 802.16, potentially suffer from problems due to uncoordinated transmissions. One well known inefficiency of random access protocols is the *hidden terminal problem*. This occurs when a node transmits, being unaware of other ongoing transmissions, which will cause a collision at some receivers.

An instance of this problem is shown in Fig. 9.2, where terminal A transmits to B and C transmits to D. Assume that A and C can both transmit to B but not to each other; hence, the reference to C being “hidden” to A and vice versa. In this case, A is unaware of C’s transmission, which can be harmful for the reception at terminal B. Conversely, also node C is not informed of A’s intention to transmit to B. Thus, a collision will occur at B, i.e., presence of strong interference, which is generally assumed to cause inability for the receiver (in this case, node B) to successfully decode the packet.

It is very easy to construct other similar examples of hidden terminal problems, see also [43, 44] where the interested reader can find further details. In general, the hidden terminal problem affects the transmission efficiency in the sense of causing possibly erroneous transmissions, which result in wasted bandwidth.

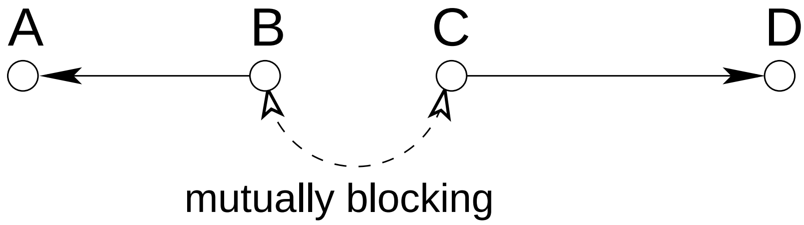

Fig. 9.3. A case of transmission showing exposed terminal problem.

At the same time, a similar issue with different consequences is the *exposed terminal problem*, which is exemplified in Fig. 9.3. Here, B and C intend to transmit to A and D, respectively. This time, the wireless medium is used inefficiently as both transmissions could be accomplished in parallel, but the senders are instead "exposed" to each other; thus, one of them transmits but the other refrains from sending packets as soon as it listens to the transmission of the other node, considering that it could cause collision. In this case, the medium access procedure is inefficient due to the low channel utilization not fully exploiting both possibilities of sending data over the channel.

Actually, both IEEE 802.11 and IEEE 802.16 MAC protocols aim at partially solving these problems. The *four-way handshake* mechanism of DCF tries to avoid the hidden terminal problem, since, e.g., a CTS sent by the intended receiver silences other potential transmitters which listen to it, thus blocking their transmissions. However, the exposed terminal problem is still unsolved, and is often considered as one of the main reasons of IEEE 802.11's inefficiencies [14]. The Mesh mode of the IEEE 802.16 standard operates similarly to avoid the hidden terminal problem, since the three way handshake can work in the same way. Moreover, the distributed election mechanism allows to alleviate the exposed terminal problem.

However, if applicable, a perfectly centralized medium access, which follows a pre-determined collision-free schedule, would work even better to prevent such situations from arising. A WMN would be theoretically able to apply a centralized STDMA schedule with the only constraints of the half-duplex limitation and the physical interference (i.e., without additional limitations to the transmission parallelism imposed by random MAC protocols), which clearly obtains better performance than more constrained cases. In general, centralized control can have disadvantages due to delay in collecting the information from the whole network, which causes the topology awareness to be inaccurate. However, in a WMN MRs are fixed and the network topology is therefore relatively stable, thus this solution is likely to be preferable to distributed algorithms. Still, centralized scheduling is also applicable if an underlying MAC protocol (in particular, either IEEE 802.11 or IEEE 802.16) is present, even though the performance will be suboptimal due to the additional protocol constraints.

As a side comment, note also that, in spite of the aforementioned techniques to solve them, the hidden/exposed terminal problems may be present in case of link asymmetry and/or time-varying channel. This happens because the rationale behind these protocols assumes that hidden transmitters are necessarily in the reception neighborhood of the potential receiver. However, this is true only if $g_{ij} = g_{ji}$. If this condition is not verified, a node can be unaware of hidden terminals even after a successful handshake exchange. Similarly, in the asymmetric channel condition, some nodes can become aware that they are exposed terminals only when certain interfering nodes, which are unable to listen to the packets of those nodes, start transmitting. Finally, due to erratic behavior of the wireless channel, it might happen that topology information collected at a single node is outdated or wrong. Also in this case, centralized control would help the network management in identifying and solving inconsistent information, whereas if the nodes operate in a distributed fashion, the effect of the hidden or exposed terminal problem may be stronger.

9.5.3 Characterizing Interference

The contribution in [18], besides having settled the basis for information-theoretic studies on the capacity of wireless networks, also introduced two useful models of interference among radio transmissions. Following their classification, we refer to them as *protocol* and *physical interference model*, respectively. Indeed, the literature reports several variations of these models, which we review below. For simplicity, we will avoid more complicated extensions which model transmission aspects such as directional antennas, capture effect (when modeled with a threshold) and so on. An overview about this can be found in [45].

Protocol interference models

The protocol interference model, in its original version, follows the rationale behind the IEEE 802.11 MAC. It models interference as causing *collision*, i.e., impossibility of correctly decoding a received packet, if other nodes in the network simultaneously exchange messages with sufficient power to disturb the ongoing transmission. The main advantage of an interference description through the protocol model is its conceptual simplicity, and the ease of mathematically formalizing the resulting interference conditions. We believe that this is, in fact, the main reason for the widespread use of the model.

The rules of the protocol interference model simply forbid that certain transmissions are simultaneously activated, when it is assumed that they will cause collision. It should be noted that, in spite of the node-based nature of the interference, this criterion is modeled through an *edge-based* constraint, i.e., to be verified for any active edge. As reported in [6, 12], a way to formalize this constraint is to define a conflicting set of edges $\mathcal{I}(e)$ associated to any edge $e \in \mathcal{E}$. According to the notation employed, the set $\mathcal{I}(e)$ may or may not also include e itself. In the following we will tacitly assume that e is included in $\mathcal{I}(e)$. The required condition is then that if edge e is active, its associated set $\mathcal{I}(e)$ must contain no more than one active edge (i.e., e itself). Formally,

$$\sum_{f \in \mathcal{I}(e)} x_f(t) \leq 1 \quad \text{if link } e \text{ is active at time } t \text{, i.e., } x_e(t) = 1. \tag{9.4}$$

In other formulations where e does not belong to $\mathcal{I}(e)$, the condition above can be promptly modified by imposing the sum of activity variables over $\mathcal{I}(e)$ to be 0 if e is an active link.

Sometimes, this relationship is translated into a *conflict graph* $\mathcal{G}_C = (\mathcal{E}, \mathcal{L}_C)$ where conflict relationships among edges are represented [16, 17]. In this formulation, the *nodes* of graph $\mathcal{G}_C$ are the *edges* of $\mathcal{G}$, whereas every edge of $\mathcal{L}_C$, which is a pair (e, f) with $e, f \in \mathcal{E}$, represents that $e \in \mathcal{I}(f)$. Though conceptually nice, this representation turns out to be very impractical in most cases, since $\mathcal{E}$ usually contains many more elements than $\mathcal{N}$ (in the worst case, $|\mathcal{E}| = |\mathcal{N}| \cdot (|\mathcal{N}| - 1)$, where $|\cdot|$ is the cardinality of the set). Also, considering the conflict graph does not solve the problem of the high computational complexity of graph operations (usually NP-complete), rather sometimes it worsen it, due to a larger graph size. Finally, we remark that the contributions which utilize the conflict graph representation consider the WMN to be a bi-directional graph, thus they utilize a bi-directional version of the conflict graph. However, this worsens the problems with respect to the link directionality issue. In fact, in this way it is impossible to describe that e causes a collision at f but not vice versa. For these reasons, the conflict graph description, introduced here for the sake of completeness, will not be mentioned further in our analysis, but the simpler approach based on the set of conflicting edges $\mathcal{I}(e)$ will be used.

The way to determine this set depends on which MAC protocol is used. In the literature, there are subtle differences among its definition, since some authors refer to the protocol model albeit they have in mind a different access strategy than IEEE 802.11 MAC or implicitly implement some protocol improvement. We refer to them as the *class of protocol interference models*, which actually encompasses several mathematical formulations. In the following we will speak of protocol model without any further specification only when describing general properties of the class. Otherwise, a specific version of the model will be mentioned.

Before describing other more complicated versions, we intentionally introduce a very simple model belonging to the protocol interference class. One straightforward possibility of defining $\mathcal{I}(e)$, though also an extreme one, is to consider $\mathcal{I}(e) = \mathcal{E}$ for all $e \in \mathcal{E}$, i.e., at most one edge can be activated at any given time throughout the whole network. In other words, either exactly one edge is active, or no edge is active at all. Due to this property, we refer to this version as the *01protocol* model. Even though it is quite oversimplified, it can be useful as a theoretical term of comparison. In fact, the 01protocol model is clearly the worst possible case of interference condition, where space diversity can not be exploited to obtain transmission parallelism.

Actually, this situation necessarily occurs on certain special topologies. For instance, in [45] this model is mentioned as used in [46] to derive the performance of DCF in an IEEE 802.11 hot-spot controlled by a single access point. Indeed, it is true that the 01protocol model holds here, but the reason is not electromagnetic interference, but rather that the topology is a star network (every node is connected only to the access-point). Therefore, the reason for having such a constraint of at most one link activation at any given time stems from the transceiver constraints, not from interference. We emphasize that this limitation should not be confused with interference constraints. Apart from their different motivation, already discussed in Subsection 9.4.2, the 01protocol is clearly a more restrictive condition than the transceiver constraint (i.e., the interference constraint described by the 01protocol is a sufficient condition for the duplex constraint). It may happen that the transmission parallelism is very difficult to obtain due to physical reasons. In certain cases, the propagation environment may exhibit extremely low attenuation from path loss so that interfering signals propagate for very long distances. If this is the situation, the 01protocol model can be appropriate to capture such weakness of the links even if the topology is loosely connected. On the other hand, as already mentioned the duplex constraint holds true for any single-channel network regardless of the interference model.

Apart from the simple 01protocol model, other versions need to rely on propagation aspects, though still simplified, to be formally described. It is common in the literature [16, 17, 44] to adopt a simple approach which makes use of geometric considerations, by implicitly assuming omni-directional propagation, isotropic environment and absence of fading. Actually, these assumptions are introduced only for the sake of presentation, as they are clearly unrealistic from the transmission physics point of view. Note however that it is possible to remove them without changing the rationale.

First of all, define the concepts of *coverage* and *disturbance* of a node.[2] Node i is said to *cover* node j if a transmission from i can be correctly received by j in the absence of any other transmission (i.e., the only factor degrading the signal quality is the thermal noise at the receiver). This means that an edge (i, j) exists in $\mathcal{E}$ and therefore $j \in \mathcal{R}_i$, or identically $i \in \mathcal{S}_j$. Similarly, node i is said to *disturb* node j if j can detect that i is transmitting, even though it may not be able to decode the message. The coverage relation is clearly a sufficient condition for disturbance, but not necessary. It is also common to find these relationships as translated into the definition of a *coverage area* and a *disturbance area*.

[2] The term which is most widely used [6, 14] for the latter is "interference." We use the term "disturbance" to avoid confusion for the reader, as the term "interference" is used in our analysis with a broader meaning, and does not necessarily refer to the protocol model.

Following the line of neglecting several propagation effects and considering only the distance-based path loss, a so-called *transmission range* can be defined. Note that in the literature on Ad Hoc networks, this range is often considered equal for all nodes. For WMNs this might be a strong approximation, since nodes may be considerably heterogeneous. Moreover, another weak point of this definition is that the distance up to which a communication link can be activated does not depend on the transmitter's characteristics (in particular, on its transmitted power) only, but also on the receiver's sensitivity. However, the transmission range assumption can be relaxed without changing the rationale, so we leave them only for presentation reasons. Thus, in the following we assume that coverage and disturbance areas are circular with radius equal to the transmission range and to a given constant ϑ (usually larger than 1) times the transmission range, respectively. Formally, if r_{TX} is the transmission range, the coverage and disturbance area for a node n are two-dimensional balls centered on n (i.e., on its location) with radii r_{TX} and $\vartheta \cdot r_{\mathrm{TX}}$, respectively.

In the original and more common version, which we call hereafter *11protocol model*, it is implicitly assumed that the IEEE 802.11 MAC is employed. For this reason, we make the assumption that links are bi-directional, as IEEE 802.11 is designed to work for bi-directional links only, and heavily relies on this hypothesis.

Following the IEEE 802.11 MAC, the 11protocol model dictates that a transmission on $(i, j) \in \mathcal{E}$ is interference free, and can therefore be activated, only if there are no transmitters, nor receivers, belonging to any active link, with either i or j in the disturbance area, apart from i and j themselves. Remember that, to enable the transmission, the IEEE 802.11 MAC protocol requires that both i and j are in the coverage area, and thus also in the disturbance area, of each other.

Note that the reason for requiring the absence of interferers in both *receiver's and transmitter's* disturbance area of both interfering *transmitters and receivers* is that the IEEE 802.11 standard forces the receiver to acknowledge RTS and data packet with CTS and ACK, respectively. In other words, due to the four way handshake, a logical receiver is also a physical *transmitter*, therefore it can cause disturbance to others. Similarly, the logical transmitter needs to perform *reception* (i.e., to receive CTS and ACK), for which it has to be collision-free. The four way handshake of IEEE 802.11 exactly aims at avoiding the hidden terminal problem on both forward and reverse link (the existence of which is specifically required by the protocol).

However, if IEEE 802.11 MAC is not used in the WMN backbone, there is no reason to impose such restrictive constraints, e.g., to mute a node which potentially disturbs the transmitter, but not the receiver. Note that to see this we need uni-directional edges, which we previously claimed to help in reducing unnecessary constraints on multiple transmissions, beyond being a better model *per se*.

In particular these conditions can be relaxed if the IEEE 802.16 MAC is used instead. There are differences, not discussed here since they are out of scope of the analysis, between the four-way and three-way handshake, which do not only involve the packets exchanged, but also the aforementioned relationships of disturbance among nodes. We can then formulate a *16protocol interference model*, which proceeds identically to the 11protocol model, with the notable exception that a collision is determined only when the designated *receiver* falls within the disturbance range of another *transmitter*. Any other combination (transmitter is under coverage of an interfering transmitter, or another receiver covers either the receiver or the transmitter) does not do any harm.

The 16protocol model solves not only the hidden terminal, but also the exposed terminal problem, and it better accounts for the directionality of the wireless links. In particular, note that the condition of interference of the 16protocol model refers to the intended receiver being under coverage of an interfering transmitter, not vice versa, since these conditions may not be equivalent.

A possible definition of $\mathcal{I}(e)$ in the 11protocol model is thus

$$\mathcal{I}(e) = \{f \in \mathcal{E} : \text{transmitter or receiver of } f \text{ disturbs transmitter or receiver of } e\}, \tag{9.5}$$

whereas in the 16protocol model it is

$$\mathcal{I}(e) = \{f \in \mathcal{E} : \text{transmitter of } f \text{ disturbs receiver of } e\}. \tag{9.6}$$

Note that in both definitions $\mathcal{I}(e)$ includes e itself.

In most papers dealing with WMN backbone management, the 11protocol model is what is meant when the protocol model is cited. However, if links are not bi-directional and the MAC does not follow the IEEE 802.11 standard, and especially if the IEEE 802.16 standard is used instead, there is no reason for using the 11protocol, and the 16protocol model would be more appropriate.

The general behavior of the model heavily depends on the ratio between the disturbance range and the coverage range. Apart from being in general hardware dependent, this value is also hard to quantify exactly, since the concepts of disturbance and coverage themselves have a vague physical meaning. In most cases, ϑ is arbitrarily chosen between 1 and 2, e.g., 1.6. This follows the approach commonly used, e.g., in Sensor Networks, where it is however conceptually more appropriate due to the fact that nodes are homogeneous (a condition which does not hold in WMNs).

If ϑ can be taken equal to 1, both 11protocol and 16protocol model can be translated to a simpler formulation connected with graph neighborhood relationships. In fact, in the case $\vartheta = 1$ the coverage range is equal to the disturbance range, and the coverage relationship (which is always necessary, but also sufficient for the disturbance if $\vartheta = 1$) is implicitly assumed in determining the existence of an edge in $\mathcal{E}$ between a transmitter and a covered receiver.

Thus, node i disturbs j if and only if they are neighbors. The exact kind of neighborhood depends on which version of the protocol model is considered.

For the 11protocol model,

$$\mathcal{I}((i,j)) = \{(k,\ell) \in \mathcal{E} : \{i,j\} \cap (\mathcal{R}_k \cup \mathcal{R}_\ell) \neq \emptyset\} \tag{9.7}$$

whereas for the 16protocol model

$$\mathcal{I}((i,j)) = \{(k,\ell) \in \mathcal{E} : j \in \mathcal{R}_k\}. \tag{9.8}$$

From this formulation, it is clear that the 16protocol model simplifies the 11protocol model as it considers the receiver j being in the coverage range of an interfering transmitter k as the situation where collision occurs. The 11protocol model instead considers four possible combinations as colliding, i.e., all cases where i or j is under coverage of either an interfering transmitter k or an interfering receiver ℓ.

This last formulation of the protocol model through neighborhood relationships is very common in the literature. We briefly remark that it can be extended to cases where the disturbance area is larger than the coverage area, i.e., $\vartheta > 1$. This happens by considering an extended graph with virtual edges $\mathcal{E}_I$, which can not be activated as useful communication links but simply describe the interference relationships. The one-hop output neighborhood $\mathcal{R}_i$ of a node i can then be replaced by a larger set $\mathcal{R}'_i$ defined similarly to what reported in (9.1) but replacing $\mathcal{E}$ with $\mathcal{E} \cup \mathcal{E}_I$.

To sum up, the protocol interference model is easy to implement, and it offers several possibilities both to describe MAC aspects, which have been classified in the three different versions (01protocol, 11protocol, 16protocol), and to employ the preferred mathematical model (coverage/disturbance range, conflict graph, neighborhood relationships). However, these practical advantages come at the price of some theoretical drawbacks. In fact, all versions of the protocol model are imperfect in capturing wireless interference. First of all, the notion of coverage range (or equivalently, conflict graph or node neighborhood) is not entirely realistic. If several power levels are adopted, it is not possible to define a single measure of coverage even from an abstract perspective. Differently from, e.g., motes of a Sensor Network, the MRs may be heterogeneous devices and therefore may have unequal characteristics in terms of transmit power, receiver sensitivity, installation site and so on. Thus, it is very hard to summarize all these physical layer effects under a single item, e.g., a single coverage range.

Moreover, a definite reason of criticism against the protocol model is that interference is not a binary relationship [10, 25]. It is true that the *outcome* of interference evaluations can be reasonably limited to two values, i.e., the activation of multiple links is either interfered or interference-free. However, the *number* of involved nodes and edges, especially in large topologies, is larger than 2, and the “disturbance” relationship defined above does not correspond to a well stated binary operation when other communication links are active in the network.

For example, strong interference, which leads to packet loss, may be present in case three specific edges are simultaneously activated, but not when any two of them are. Thus, no specific link alone causes interference, but the problem is the joint effect of all links. Seen from the point of view of a single edge e, it might happen that $f, g \in \mathcal{E}$ can individually coexist with e, but not jointly. In this case, it is doubtful whether f and g should be inserted in $\mathcal{I}(e)$. We remark that usually the conflict set $\mathcal{I}(e)$ is evaluated pair-wise, as defined above, which would lead to problems as the joint activation of f and g is not prohibited. On the other hand, the alternative approach where any edge possibly disturbing e (even if this happens only if other links are activated as well) is put into $\mathcal{I}(e)$, would be too conservative to be practically useful.

Physical interference model

These problems can be overcome by means of the physical interference model, whose rationale is as follows. The packet error rate (PER) at the receiver is a monotonically decreasing function of the Signal-to-Interference-and-Noise Ratio (SINR). It is often reasonable to simplify this relationship and consider a threshold approach, where it is assumed that a packet is correctly received with probability 1 if the SINR is above a given threshold. A way to formalize this is:

$$\frac{P_i g_{ij}}{\sum_{k \neq i} P_k g_{kj} + N_j} \geq \gamma_j \, , \tag{9.9}$$

where (i, j) is the link of interest, the index k in the lower sum denotes a possible interferer (i is in fact excluded from the sum, as it is the intended transmitter), P_x is the power emitted by node x, g_{xy} is the path gain from x to y and N_j is the noise at the receiver node j. The value γ_j, which defines the SINR threshold, can be in general a different value for every node j.

Hereafter we use these assumptions, which are made only for ease of exposition, but without loss of generality, as avoiding them would only lead to a more cumbersome (though conceptually identical) formulation. We take $\gamma_j = \gamma$ for all j. We also neglect the noise terms and we consider an equal power level P among all transmitting nodes. In particular, the last assumption is equivalent to assuming that the power level is simply fixed. If this is the case, the elements (g_{ij}) of the matrix $\mathbf{G}$ can be replaced by $g'_{ij} = P_i g_{ij}$ and the power term can be omitted. If the power level is instead not fixed, it would become necessary to also include Power Control in the analysis. However, this can be performed within a very similar framework, as shown in [8].

In the context of our framework which describes scheduling and routing through link activation patterns, the constraint can be formalized as follows:

$$\frac{x_{ij}(t) g_{ij}}{\sum\limits_{k \in \mathcal{S}_j \setminus \{i\}} g_{kj} \sum\limits_{\ell \in \mathcal{R}_k \setminus \{j\}} x_{k\ell}(t)} \geq \gamma \tag{9.10}$$

if link (i, j) is active at time t, i.e., $x_{ij}(t) = 1$.

The basic assumption of the model, i.e., the possibility of reducing the PER to a step function around the value γ, is indeed an approximation. However, it is much more accurate than those made under the protocol models. In fact, it better takes into account physical propagation, and allows for a correct packet reception even in the presence of (moderate) interference, differently from the collision assumption. Also, it properly accounts for the cumulative character of interference. Indeed, the choice of γ depends on the shape of the PER function, which in turn relates to the modulation scheme, and on the PER value which is considered as acceptable at the application level. However, none of these factors depend on MAC issues; thus, the physical model allows to operate between MAC and other layers in a more modular manner.

The drawback of this model is that it translates into more complex mathematical relationships than the protocol model. Moreover, if a specific MAC needs to be addressed, additional constraints are required. For example, in an IEEE 802.11 network, the physical model fails to describe certain constraints on link activation due to the RTS/CTS exchange, which are instead taken into account in the 11protocol model. Hence, which is the best model to use ultimately depends on the purpose of the analysis. From the point of view of theoretical analysis of WMNs, however, the physical model has a good point against the protocol model, as described, e.g., in [10, 15]. In dense topologies, where the number of incoming or outgoing links at a node is high, the protocol models are very restrictive and obtain lower network parallelism, due to their requirement of silencing allegedly colliding connections. As shown above, when the WMN topology is rich of edges, all protocol models approach the 01protocol model, which is the most restrictive case and implies that at most one edge is activated. This is a problem for WMNs, which, being meant to provide good network coverage and high data rates, usually have a dense topology. The better performance obtained in this sense by utilizing the physical model should also imply the need to re-think existing access protocols for WMN. Indeed, the Mesh versions of both IEEE 802.11 and IEEE 802.16 take these aspects into account. However, in our view the protocol design of improved interference-aware routing and scheduling strategies is still an open research challenge.

9.6 Performance evaluation

In this section, we focus on the problem of defining efficient link activation patterns which not only satisfy all the constraints but also deliver traffic to the MAPs acting as gateways for the WMN. We focus on the minimal time scheduling problem, i.e., to deliver a given amount of traffic from all the non gateway MRs to the MAPs (as we deal with the uplink case) in the shortest possible time. This problem is also closely related to the throughput maximization, i.e., to obtain the highest amount of traffic delivered to the gateways in an assigned time. Indeed, with minor modifications our framework can work to solve this problem as well.

In the following, we will refer to the backlog queue length at node i, assumed to be varying over time, as $q_i(t)$. Thus, all non-gateway MRs have, at time 0, a backlog of length $q_i(0)$ to be sent to any of the MAPs. The minimal time scheduling problem corresponds to finding the lowest length $T_{\min}$ of a feasible link activation pattern which delivers all traffic to the gateways. Denoting with $\mathcal{Y}$ the set of gateways, this means that

$$T_{\min} = \min\{t : q_i(t) = 0\ ,\ \ \forall i \in \mathcal{N} \setminus \mathcal{Y}\}. \tag{9.11}$$

For simplicity, we assume that the value of $q_i(0)$ is known a priori and no further packet arrivals take place after link activation has started. In this way, if the uplink problem can be solved over a specified finite time-horizon T, i.e., $T_{\min}$ is lower than or equal to T, its solution can also serve as the basis for a periodic schedule, where a link activation pattern of length T is indefinitely repeated. In other words, it is possible to see the uplink problem as a way to deliver a given amount of packets under loose delay guarantees (i.e., every packet is delivered within $2T$ slots, provided that the arrival rate to the MRs from MCs can be assumed constant). A further extension is possible to the cases of prioritized traffic with different priority classes or different required delay guarantees. Another option is to consider packet arrivals within the time frame. All these differences do not change most of the considerations we will present in the following, and can be investigated within a similar framework. We identify them as possible interesting directions for future research.

Finding the shortest-time link activation pattern for the uplink problem can be addressed in the context of an optimization problem, by adding proper *flow constraints* to the already mentioned duplex and interference constraints. Among these, the most important is the flow conservation property, i.e., the traffic transmitted over (i, j) in a given time slot t is upperbounded by the number of packets available at node i after the transmission occurred at time $t-1$. Note that this is true if, as assumed above, all traffic arrives at the MRs at the beginning of the schedule.

Other conditions may impose that an edge (i, j) can be active at time t only if $q_i(t) > 0$ or that the edges *from* an MAP are never activated. These constraints are not strictly necessary, but they eliminate from the feasible region parts of the search space which are guaranteed to contain only non-optimal solutions.

Several approaches have been proposed to formalize the problem of finding the optimal link activation pattern to minimize the time to deliver all traffic to MAPs [7, 26]. However, the resulting optimization problem is NP-complete [10]. For this reason, we focus our analysis on some theoretical results on the overall performance of WMNs for the minimal scheduling problem. Within this approach, not only is it possible to frame other existing results, but also we are able to draw interesting guidelines and conclusions about the performance of WMNs.

9.6.1 Theoretical performance bounds

Determining the value of $T_{\min}$ is interesting for both theoretical and practical reasons. In fact, the problem of delivering a given amount of traffic can also be seen from the information theoretical point of view as a capacity estimation, since the shorter the time to deliver a given amount of packets, the higher the throughput over a given time interval. Also, if $T_{\min}$ is sufficiently low, a centralized periodic scheduling can be implemented. However, the problem of determining $T_{\min}$ exactly is very complicated. Not only is it an NP-complete problem, but also it strongly depends on the network parameters, i.e., the graph topology, the edge rates and the initial backlog at each node.

Thus, solutions based on integer linear programming often introduce simplifications to make the problem more tractable. For instance, [13] employs a *fluidic approximation* to the link rates, i.e., the x_{ij} variables are relaxed to be time-invariant real numbers between 0 and 1 instead of being binary digits variable over time. In other words, the x_{ij} variables represent the average activity of link (i, j) over the time period. However, this approach has some drawbacks, for example it leads to rounding problems. If it is found that the optimal average link activity for link (i, j) is, say, 0.83, and T is found to be equal to 10, it is not clear whether (i, j) should be active on 8 or 9 time slots. Moreover, the practicality of the approach is decreased with respect to the initial integer problem, where the solution could be directly translated into a schedule simply by taking the resulting link activation pattern, which is no longer possible. Finally, the overall $T_{\min}$ to schedule all the traffic is underestimated with respect to the original integer case, as observed by the authors themselves.

Another possibility which is sometimes proposed [7] is to employ topology control to reduce the number of edges which can be activated. Even though this indeed decreases the complexity of the problem, we argue that this procedure can lead to a severe decrease of the transmission parallelism, therefore obtaining low throughput as a result. Thus, it is in general not recommended to prune edges to decrease the cardinality of $\mathcal{E}$. This is true even for the cases where topology control is claimed to be interference-aware: as discussed in previous sections, allocating non interfering simultaneous connections is a task to be performed at the MAC layer, i.e., through a scheduler (or in our case, through a joint routing-scheduling procedure), not with a routing algorithm. If interference awareness is introduced in the network by simply reducing the possible routes, the most significant result is a decrease of the overall performance.

Finally, the most natural way to deal with difficult problems, i.e., to introduce a heuristic solution method, is also common in the literature [6, 27]. Indeed, to identify novel and possibly topology-adaptive heuristics or metaheuristics is another possible direction for further research. Instead of proposing yet another heuristic, we present some theoretical results which hold in general for WMNs. Similar findings have been also presented in other contributions [13, 17, 15], which however heavily rely on the assumption of the protocol interference model. Instead, our analysis is *independent of the underlying interference model*, as it only relies on the half-duplex assumption, which, as discussed in Subsection 9.4.2, holds in any case. Under this hypothesis, we derive theoretical bounds for the performance of WMNs, in which the interference model of choice can be framed (obtaining different results, according to how restrictive it is).

Prior to describing the analytical formulation, note what follows about the notation of the following theoretical statements. As observed above, a TDMA scheduler operates on discrete time slots. Thus, the *number of slots* required to accomplish a transmission is necessarily integer. However, in the following we will refer to the *time to transmit* a given amount of traffic as a real number. Differently from [13], this does not imply that we are relaxing the constraints of $x_{ij}(t)$ to be integer, but simply that in all the cases where the time to transmit is fractionary, the number of slots corresponds to its rounded-up version. Thus, the results shown in the following can be refined to properly capture the fact that time slots are integer numbers by adding ceilings where necessary.

The first result is an upper bound on $T_{\min}$ which can be seen, to some extent, as introduced by [17]. The overall idea of this paper is to determine the minimal time scheduling by deriving maximal cliques of edges which can be compatibly allocated. This is just a different formulation of the problem, which does not solve in any way its NP-completeness. Besides, the whole analysis is based on the protocol interference model. However, an interesting point is given in the paper. No matter how inefficient the schedule is, at least one edge should be activated at a time. Thus, an upper bound for $T_{\min}$, denoted as $T^U_{\min}$, is implicitly obtained, exactly by taking this as a worst case assumption. Note that this upper bound corresponds to what, in Subsection 9.5.3, was referred to as 01protocol model. Therefore, this upper bound is also tight in the sense that there is an interference condition in which the shortest link activation pattern must necessarily be $T^U_{\min}$ slots long. This is exactly when the interference corresponds to the 01protocol model, which is the worst possible case.

To derive $T^U_{\min}$, we simply take a weighted shortest path to the gateways from any node, e.g., by using the well known Dijkstra algorithm, where the weights are the inverse of the link rates r_{ij}. In fact, it is easy to see that, if transmitting a backlog q over a link of rate r takes a time equal to qr^{-1}, the time to transmit it over the series of two links (activated one at a time) having rate r_1 and r_2, respectively, is $q(r_1^{-1} + r_2^{-1})$. Finally, the value of $T^U_{\min}$ is derived as

$$T^U_{\min} = \sum_{i \in \mathcal{N} \setminus \mathcal{Y}} q_i(0) \sum_{e \in \mathcal{P}_i} r_e^{-1} \tag{9.12}$$

where $\mathcal{P}_i$ is the shortest path in the sense mentioned above for node i.

The upper bound described by $T^U_{\min}$ corresponds to a very conservative case of protection against interference. The minimal time to deliver all the traffic is $T^U_{\min}$ only in the case of 01protocol model, or very high SIR target in the physical model. For this reason, we introduce also a lower bound on $T_{\min}$ for a *single gateway case*. This condition is likely to be present in most WMNs and has been first envisioned by [15] as a possible bottleneck for the network capacity of such systems. The authors of this paper argue that if a single gateway is used and the highest rate of all links entering in it is a, there is a lower bound on the time to deliver all packets, equal to Q/a, where Q is the sum of all backlogs in the network at time 0. This lower bound is trivial in most cases, but it might be interesting for certain sparse topologies. Especially in [15] it is shown that a chain topology behaves badly in this sense. Moreover, the authors further improve this result by giving some theoretical considerations based on the protocol model (more specifically, the 11protocol model version).

Though inspired by this result, we follow here another approach. We demonstrate that taking only the half duplex constraint into account is sufficient to significantly improve the aforementioned lower bound. Even though the problem of determining a tight lower bound on $T_{\min}$ would still be NP-complete, we remark that in practice our theoretical result gives a good estimate of $T_{\min}$ for the case of no interference (hence, the other extreme with respect to $T^U_{\min}$) in several cases. If more than one gateway is present, the gateway bottleneck is strongly mitigated; thus, an immediate conclusion of our analysis is that WMNs perform significantly better if two or more MAPs are available.

To derive the lower bound, referred to in the following as $T^L_{\min}$, we proceed as follows. Consider, as in [15], the edge entering the gateway with highest rate (equal to a). The transmitter node of this edge would be called in the following "MR number 1" and its backlog will be denoted as q. As above, let Q indicate the overall backlog in the network. Let s be the highest rate of all edges entering MR number 1, and let b be the highest rate among the edges entering the gateway, not counting the one from MR number 1 (hence, $a \geq b$). If multiple nodes can be chosen as MR number 1, as several edges to the gateway have equal rate, simply put $b = a$ and s will be consequently equal to the highest possible rate among all edges entering those nodes. For simplicity, we assume that $b > 0$ and $s > 0$. However, it is still possible to generalize the result shown in the following to $b = 0$ or $s = 0$.

The situation is represented in Fig. 9.4. In the following, we neglect all edges in the rest of the network, and we will also neglect multiple edges with identical rates. As a matter of fact, we only consider three links: from MR number 1 to the gateway, from the rest of the network to the gateway and from the rest of the network to MR number 1, having rates a, b and s, respectively. With a slight abuse of notation, we will call them with their rate value, for brevity. The lower bound $T^L_{\min}$ is derived considering all the traffic in the rest of the network (equal to $Q - q$ packets) to be always available

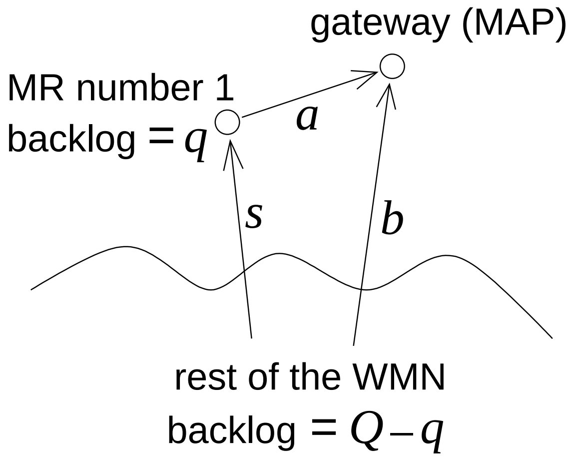

Fig. 9.4. Notations used to derive the lower bound $T^L_{\min}$.

on "border" nodes which can use these links. Actually, this is an optimistic assumption as these packets can be instead queued at other nodes which are not directly connected to the gateway, or to MR number 1. The derivation of $T^L_{\min}$ is obtained through the following theorem.

Theorem 1. *A lower bound on* $T_{\min}$ *is given by*

$$T^L_{\min} = \frac{q}{a} + \frac{Q-q}{s+b}\left(1 + \frac{s}{a}\right). \tag{9.13}$$

Proof. First, observe a general property. When edge a is active, i.e., MR number 1 sends packets to the gateway, no other transmission to these nodes can be activated due to the half-duplex constraint. At most, it is possible to activate in parallel some transmissions within the "rest of the network," but this has no effect whatsoever, since we are under the optimistic assumption that all the traffic which is not queued at MR number 1 is always available for transmission on links b and s. We can therefore neglect these transmissions as they can not improve the lower bound $T^L_{\min}$. Thus, it is not restrictive to assume that all the traffic available at MR number 1 is transmitted first, which takes a time equal to q/a. Then, $T^L_{\min} = q/a + T_1$, where T_1 is a lower bound on the delivery time in the same network, where however q has been delivered to the gateway. Since MR number 1 has now no packets in the queue, links b and s need to be activated. The best possibility (i.e., the one minimizing the delivery time) is that they can operate perfectly in parallel. During such a parallel transmission, assume that x and y are the amounts of traffic sent over link s and b, respectively. After this transmission, no packets are left in the rest of the network, so $x + y = Q - q$. Moreover, the minimum transmission time is obtained when x and y take exactly the same time to be transmitted.

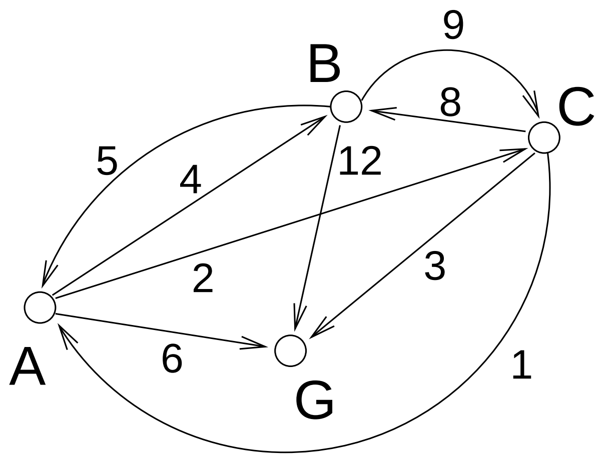

Fig. 9.5. An example of network topology.

This means that

$$\begin{cases} x + y = Q - q \\ x : s = y : b \end{cases} . \quad (9.14)$$

This system of equations can be solved so as to obtain

$$x = s\frac{Q - q}{s + b} , \quad y = b\frac{Q - q}{s + b} . \quad (9.15)$$

The parallel transmission over b and s is also found to have a duration of $(Q - q)/(s + b)$. After its termination, an amount of traffic equal to y has been delivered to the gateway, whereas x is still in queue at MR number 1. The best possibility to transmit x is to use link a, which takes a time equal to $(s/a)(Q - q)/(s + b)$. Thus, collecting all these results,

$$T_1 = \frac{Q - q}{s + b}\left(1 + \frac{s}{a}\right) \quad (9.16)$$

and the theorem is proved. ■

This result has many practical consequences. For example, not only does a impact on $T_{\min}$, but so do b and s. In particular, if $s \ll a$, the gateway bottleneck is worsened, since packets arrive at MR number 1 with very low rate, thus alternate paths (hence with rate b lower than a) have to be used. On the same line, link a can not always be used for transmitting packets, since it can not be activated when MR number 1 is receiving. If b is considerably lower than a, there may be a decrease in the network throughput. These considerations give some practical guidelines for network deployment. First of all, it is important to have several "good" links to the gateway, i.e., b should be close to a, and there should be multiple non-interfering paths to the gateway, so as to allow parallel allocation of links to the gateway and to some of the neighbors of the gateway. Instead, if all routes to the gateway traverse the same node, the single gateway bottleneck is worsened. Moreover, the rate of connections to the gateway should be high, but it is also important to have a good relaying speed to the gateway neighbors (i.e., high s).

For what concerns numerical evaluations, we found that this lower bound works well in practical cases. In particular, it is much stricter than the trivial lower bound given by Q/a, and it also has the advantage of limiting the analysis to three numerical values, i.e., the best rate and the second best rate of edges entering the gateway and the best rate of edges entering MR number 1. To give an idea of this, consider[3] the sample WMN represented in Fig. 9.5. Numbers reported on the edges denote their rates (again, with the same abuse of notation, we speak, e.g., of edge 1 to indicate the one from node C to node A). Assume that node G is the gateway (this is also implicitly taken into account in the figure, where no edges from G are depicted), and that $q_i(0) = 24$ for every node. In such a case, $T^U_{\min} = 9$ (shortest paths are through direct links for all MRs but for node C, whose best path is through edges 8 and 12, with an overall rate of 24/5), which is indeed the actual value of $T_{\min}$ if the 01protocol model holds. As $a = 12$, $b = 6$, $s = 8$, the "trivial" lower bound Q/a is 6, whereas $T^L_{\min} = 7.71$. This latter value is much more accurate than the former, as the minimal length of the schedule is 8 time slots (the optimal schedule corresponds in fact to activating edges 8 and 6 simultaneously for three slots, then edge 8 alone for one slot and finally edge 12 for four slots). The slightly lower value of $T^L_{\min}$ with respect to the real value is a consequence of the fact that the parallelism of edges 8 and 6 is not perfect (even though the round-up still eliminates this issue). Moreover, the activation of this parallel transmission is possible only if it does not violate any additional interference constraint. In fact, if for example the 11protocol interference model is assumed, as here the network topology is a clique, we obtain $\mathcal{I}(e) = \mathcal{E}$ for all e, thus we fall again in the case described by the upper bound.

This suggests that, according to the interference model, the network performance in terms of $T_{\min}$ moves from the upper to the lower bound (or close to it). In the next subsection, we will show a more extensive analysis of this behavior.

9.6.2 Numerical results

In this section we analyze, through an example, the resulting network capacity when different interference models are adopted. To this end, we focus on the

[3] We remark that this topology, and also the one shown in the next subsection, have only the value of examples and are not proposed in this paper as realistic or efficient network deployments.

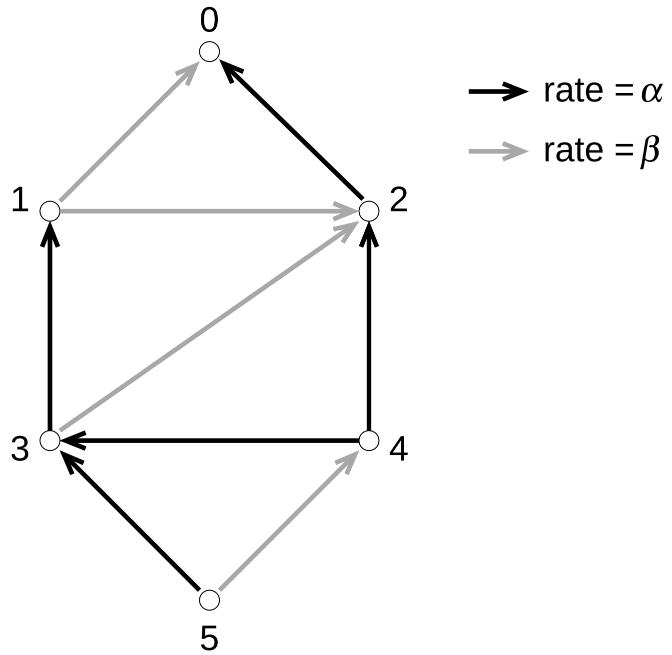

Fig. 9.6. Example topology: the gateway is node 0, edge rates are either α (black links) or β (grey links).

network topology reported in Fig. 9.6 which consists of six nodes. Node 0 is assumed to be the gateway. The links between nodes are represented by directional edges whose rate is either α or β. With respect to protocol models, the disturbance range is assumed to be equal to the transmission range, i.e., $\vartheta = 1$. Thus, a node can disturb only the nodes it can transmit to, and vice versa. The nodes that do not have a direct link toward the gateway exploit their neighboring nodes to relay their packets. We assume that each node is provided with an equal amount of traffic to forward to the gateway. The network capacity is evaluated by means of $\lceil T_{\min}/q \rceil$, i.e., the number of slots needed to deliver the overall network workload to the gateway, normalized over the initial amount of traffic of the nodes. To some extent, we can draw in this way general conclusions on the network capacity irrespective of the initial traffic load of the nodes. Depending on the interference model, our analysis is carried out either with the theoretical results described in subsection 9.6.1 or through numerical evaluations, performed with an exhaustive search over all possible link activation patterns.

In Fig. 9.7 we plot this metric versus the ratio α/β, considering a case where $\alpha\beta = 1$, for different interference models. As can be seen, the 01protocol model curve (also corresponding to the analytical upper bound for the performance of any MAC) always lies above the other ones due to the restric-

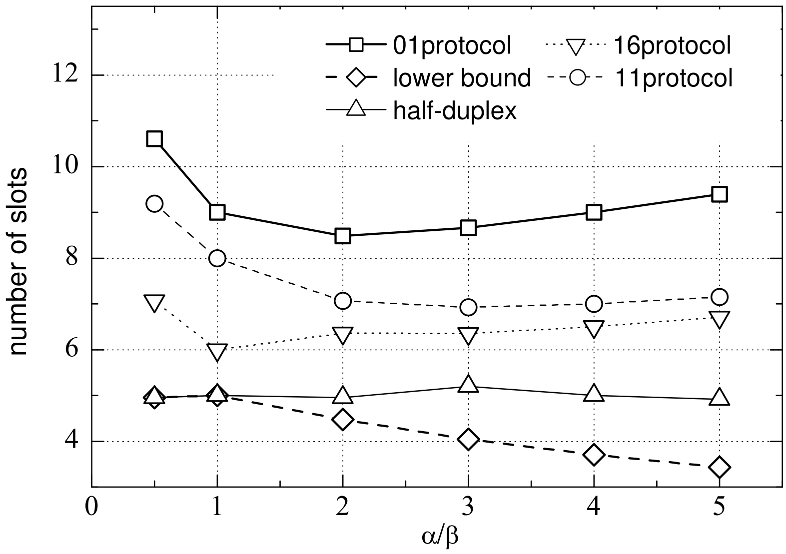

Fig. 9.7. Numerical results. For all protocol models $\vartheta = 1$ is considered.

tive constraint that at most one link can be active at a time. Even though the curves exhibit slightly variable behavior when α/β is changed, there is in any case a significant gap (a factor of 1.8 or more) between the 01protocol model and the theoretical lower bound curves. Instead, the half duplex performance, derived through exhaustive search in the least restrictive condition of simultaneous link activation, is well approximated by the analytical lower bound. Especially, the lower bound is fairly tight when $\alpha \leq \beta$.

Any value between the 01protocol model and the half-duplex curves may potentially be achievable depending on the interference model. In particular, if the physical model is utilized, the performance will span between the two extreme curves almost with continuity. If a protocol model is used instead, the behavior is more difficult to change. The performance of both 11protocol and 16protocol can be observed in the figure to be according to the reasonings of Subsection 9.5.3. For example, the 11protocol is closer to the 01protocol than the 16protocol due to its more restrictive assumptions. Recall that, in the 11protocol, all the nodes falling into the disturbance range of both the transmitter and the receiver need to be silenced. This assumption is relaxed in the 16protocol which thus permits more parallelism of the transmissions. However, both the 11protocol and 16protocol models obtain a significantly higher $T_{\min}$ than when only the half-duplex constraint is assumed, but no further constraint is imposed.

The condition $\alpha = \beta$ corresponds to the case where the link rates in the network become homogeneous, i.e., all nodes transmit at the same rate. As envisioned in the previous discussion, this is a good condition for achieving high throughput, since the lack of potential bottlenecks caused by slower links permits to efficiently exploit the overall network capacity. As soon as the link rates in the network become heterogeneous, we observe that the performance degrades in this sense. Note that, because $\alpha\beta = 1$, when α is increased, it also happens that β is inversely proportionally decreased. For $\alpha > \beta$ link rates are higher on aggregate: e.g., when $\alpha/\beta = 4$, the average link rate is equal to 4/3 instead of 1. However, this does not correspond to an improvement in the scheduling efficiency, because there are high rate links to the gateway, but also strong variability of the link rates is present. Due to the aforementioned bottlenecks, the case only constrained by half-duplex limitations keeps its performance as almost constant. The performance of 01protocol, 11protocol and 16protocol are instead slightly worse than the homogeneous network case.

Conclusions

In this paper, we have investigated some research issues arising in the context of link activation for WMNs. Specifically, we revisited the classic problems of routing and link scheduling over multi-hop wireless networks to provide the reader with a clear overview of the hottest topics in WMNs. After a brief introduction discussing preliminary concepts of wireless networks, we have critically reviewed the recent literature in this field, highlighting pros and cons of possible approaches to the problem. We have proposed an approach which jointly considers the routing and link scheduling problems. To this aim, we have introduced and described theoretical models to characterize wireless networks, which include the nodes' transmission/reception constraints and the interference of wireless links. Within this theoretical framework, we have discussed the characteristics of the most common MAC protocols. Finally, we have derived theoretical performance bounds for network capacity and have compared these bounds to the results obtained by the presented models in a sample topology.

We believe that these results can be useful in many ways. For sure, one possibility is to use them as guidelines for WMN deployment so as to avoid bottlenecks in the network and allow instead high data rates to the end users, which corresponds to the major objective of such systems. At the same time, we remark how our findings highlight the need for a proper interference characterization. This point seems often neglected, as testified by the widespread usage of the protocol interference models, which do not correctly describe the underlying physical aspects and also lead to pessimistic performance results. We believe that further investigation on wireless interference models and their impact on MAC and network layer aspects is an interesting scientific challenge. Finally, efficient link activation strategies exhibit a performance with a high degree of variability between the evaluated analytical upper and lower bounds. In this respect, novel proposals which are able to fill this gap are clearly emphasized as very promising directions of future research.

References

1. I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks (Elsevier)*, vol. 47, pp. 445–487, Mar. 2005.
2. R. Bruno, M. Conti, and E. Gregori, "Mesh networks: commodity multihop ad hoc networks," *IEEE Commun. Mag.*, vol. 43, pp. 123–131, Mar. 2005.
3. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.
4. *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, 2004.
5. K. N. Ramachandran, M. M. Buddhikot, G. Chandranmenon, S. Miller, E. M. Belding-Royer, and K. C. Almeroth, "On the design and implementation of infrastructure mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, Sept. 2005.
6. M. Alicherry, R. Bhatia, and L. B. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1960–1971, Nov. 2006.
7. J. Tang, G. Xue, and W. Zhang, "Interference-aware topology control and QoS routing in multi-channel wireless mesh networks," in *Proc. ACM MobiHoc*, Urbana-Champaign, IL, USA, May 2005, pp. 68–77.
8. R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks," in *Proc. IEEE INFOCOM*, vol. 1, San Francisco, CA, USA, 2003, pp. 702–711.
9. H. Balakrishnan, C. L. Barrett, V. S. A. Kumar, M. V. Marathe, and S. Thite, "The distance-2 matching problem and its relationship to the MAC-Layer capacity of ad hoc wireless networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1069–1079, Aug. 2004.
10. G. Brar, D. Blough, and P. Santi, "Computationally efficient scheduling with the physical interference model for throughput improvement in wireless mesh networks," in *Proc. ACM MobiCom*, Los Angeles, CA, USA, Sept. 2006, pp. 2–13.
11. L. Tassiulas and A. Ephremides, "Jointly optimal routing and scheduling in packet ratio networks," vol. 38, no. 1, pp. 165–168, Jan. 1992.
12. M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 868–880, Aug. 2005.
13. ——, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom*, Cologne, Germany, 2005, pp. 73–87.
14. K. Xu, M. Gerla, and S. Bae, "How effective is the IEEE 802.11 RTS/CTS handshake in ad hoc networks," in *Proc. IEEE Globecom*, Taipei, Taiwan, R.O.C., 2002, pp. 72–76.
15. J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Commun. Mag.*, vol. 10, no. 5, pp. 8–14, Oct. 2003.
16. W. Wang and X. Liu, "A framework for maximum capacity in multi-channel multi-radio wireless networks," in *Proc. IEEE CCNC*, vol. 2, 2006, pp. 720–724.
17. N. Ben Salem and J.-P. Hubaux, "A fair scheduling for wireless mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, 2005, invited paper.
18. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.

19. C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Measurement-based models of delivery and interference in static wireless networks," in *Proc. ACM SIGCOMM*, Pisa, Italy, 2006, pp. 51–62.
20. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. ACM MobiCom*, Philadelphia, PA, USA, July 2004, pp. 114–128.
21. Y. Yang, J. Wang, and R. Kravets, "Designing routing metrics for mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, Sept. 2005.
22. H. Wu, F. Yang, K. Tan, J. Chen, Q. Zhang, and Z. Zhang, "Distributed channel assignment and routing in multiradio multichannel multihop wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1972–1983, Nov. 2006.
23. T. Salonidis and L. Tassiulas, "Distributed on-line schedule adaptation for balanced slot allocation in wireless ad hoc networks," in *Proc. IEEE IWQoS*, Montreal, QC, Canada, June 2004, pp. 20–29.
24. P. Djukic and S. Valaee, "Distributed link scheduling for TDMA mesh networks," in *Proc. IEEE ICC 2007*, Glasgow, Scotland, to appear.
25. K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, no. 4, pp. 471–487, July 2005.
26. M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Proc. ACM MobiHoc*, Urbana-Champaign, IL, USA, May 2005, pp. 78–89.
27. H. Wei, S. Ganguly, R. Izmailov, and Z. J. Haas, "Interference-aware IEEE 802.16 WiMax mesh networks," in *Proc. IEEE VTC 2005 Spring*, Stockholm, Sweden.
28. A. P. Subramanian, M. M. Buddhikot, and S. Miller, "Interference aware routing in multi-radio wireless mesh networks," in *Proc. WiMesh*, Reston, VA, USA, 2006, pp. 55–63.
29. S. O. Krumke, M. V. Marathe, and S. Ravi, "Models and approximation algorithms for channel assignment in radio networks," *Wireless Networks*, vol. 7, no. 6, pp. 575–584, Nov. 2001.
30. P. Kyasanur and N. H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proc. ACM MobiCom*, Cologne, Germany, 2005, pp. 43–57.
31. A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for IEEE 802.11 wireless networks," in *Proc. IEEE BroadNets*, San José, CA, USA, 2004, pp. 344–354.
32. A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful," in *Proc. ACM SIGMetrics/Performance*, Saint Malo, France, 2006, pp. 63–74.
33. F. M. E. Amaldi, A. Capone, "Planning UMTS base station location: optimization models with power control and algorithms," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 939–952, Sept. 2003.
34. B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with QoS constraints," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2127–2136, Nov. 2006.
35. A. J. Viterbi, *CDMA: principles of spread spectrum communication*. Addison Wesley Longman Publishing Co., Inc.
36. J. W. Wallace and M. A. Jensen, "Modeling the indoor MIMO wireless channel," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 5, pp. 591–599, May 2002.

37. A. Scaglione and Y.-W. Hong, "Opportunistic large arrays: cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2082–2092, Aug. 2003.
38. Y. Sagduyu and A. Ephremides, "Crosslayer design for distributed MAC and network coding in wireless ad hoc networks," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, 2005, pp. 1863–1867.
39. L. Lai, K. Liu, and H. E. Gamal, "The three-node wireless network: achievable rates and cooperation strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 805–828, 2006.
40. S. M. Das, H. Pucha, D. Koutsonikolas, Y. C. Hu, and D. Peroulis, "DMesh: Incorporating practical directional antennas in multichannel wireless mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2028–2039, Nov. 2006.
41. J. A. Stine, "Exploiting smart antennas in wireless mesh networks using contention access," *IEEE Wireless Commun. Mag.*, vol. 13, no. 2, pp. 38–49, Apr. 2006.
42. D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," in *Proc. ACM MSWiM*, Venice, Italy, 2004, pp. 78–82.
43. S. Basagni, M. Conti, S. Giordano, and I. Stojmenović, Eds., *Mobile Ad Hoc Networking*. New York: IEEE press and John Wiley & Sons, 2004.
44. H. Zhai and Y. Fang, "Medium access control protocols in mobile ad hoc networks: problems and solutions," in *Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless and Peer-to-Peer Networks*, J. Wu, Ed. Boca Raton, FL, USA: Auerbach Publications, Taylor & Francis Group, 2006, ch. 15, pp. 231–250.
45. A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" in *Proc. Qshine*, no. 2, Waterloo, ON, Canada, 2006, invited paper.
46. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.