On max-min fair flow optimization in wireless mesh networks

Michal Pioro, Mateusz Zotkiewicz, Barbara Staehle, Dirk Staehle and Di Yuan

Michal Pioro, Mateusz Zotkiewicz, Barbara Staehle, Dirk Staehle and Di Yuan, On max-min fair flow optimization in wireless mesh networks, 2014, Ad hoc networks, (13), 134-152.
http://dx.doi.org/10.1016/j.adhoc.2011.05.003

On max-min fair flow optimization in wireless mesh networks

Michał Pióro[a,b], Mateusz Żotkiewicz[b], Barbara Staehle[c], Dirk Staehle[c], Di Yuan[d]

[a] *Lund University, Sweden*
[b] *Warsaw University of Technology, Poland*
[c] *Wuerzburg University, Germany*
[d] *Linköping University, Sweden*

Abstract

The paper is devoted to modeling wireless mesh networks (WMN) through mixed-integer programming (MIP) formulations that allow to precisely characterize the link data rate capacity and transmission scheduling using the notion of time slots. Such MIP models are formulated for several cases of the modulation and coding schemes (MCS) assignment. We present a general way of solving the max-min fairness (MMF) traffic objective for WMN using the formulated capacity models. Thus the paper combines WMN radio link modeling with a non-standard way of dealing with uncertain traffic, a combination that has not, to our knowledge, been treated so far by exact optimization models. We discuss several ways, including a method based on the so called compatible or independent sets, of solving the arising MIP problems. We also present an extensive numerical study that illustrates the running time efficiency of different solution approaches, and the influence of the MCS selection options and the number of time slots on traffic performance of a WMN. Exact joint optimization modeling of the WMN capacity and the MMF traffic objectives forms the main contribution of the paper.

Keywords: wireless mesh network, max-min fairness, mixed-integer programming

Email addresses: `Michal.Pioro@eit.lth.se` (Michał Pióro), `mzotkiew@tele.pw.edu.pl` (Mateusz Żotkiewicz), `bstaehle@informatik.uni-wuerzburg.de` (Barbara Staehle), `dstaehle@informatik.uni-wuerzburg.de` (Dirk Staehle), `diyua@itn.liu.se` (Di Yuan)

Contents

1. Introduction

Wireless mesh networks (WMN) offer common and affordable access to the Internet in metropolitan and residential areas. The core of a WMN consists of a set of fixed mesh nodes – routers and Internet gateways – interconnected by radio links that typically follow the Wi-Fi IEEE 802.11-family standards. Other standards, such as Bluetooth IEEE 802.15.5, WiMAX IEEE 802.16a, and IEEE 802.20 can also support WMNs. Mesh clients, being either fixed or mobile, connect to mesh routers to obtain access to Internet gateways either over direct links or via multi-hop mesh routes. WMN is a cost-efficient approach for Internet access with a bandwidth in the range of 50-200 Mbps. The WMN solution is competitive to the wired Internet access offered by cable network providers or by mobile operators. WMNs are decentralized, non-hierarchical networks, typically deployed by communities of users (see [1, 4]), and based on commonly available off-the-shelf wireless communication equipment (see [2, 3, 5]). The idea of WMN stems from the ad-hoc networking paradigm and, as such, fits very well the decentralized philosophy of the Internet. For comprehensive surveys of WMN, we refer to [7, 19].

Although WMNs are relatively cheap and easy to deploy, achieving efficient and fair resource allocation is not straightforward. One particular issue is how to effectively allocate the offered network capacity among the routes between the gateways and the routers. Without network optimization, which ideally should be simple, fast, and distributed, a WMN can behave poorly, delivering significantly lower throughput than it can potentially achieve. For WMN, network optimization tasks range from transmission scheduling, through channel assignment, transmission power adjustment and rate adaptation, to routing. In this context, traffic engineering, which is a key aspect in operating communication networks, is of high significance to WMN. From the mathematical optimization standpoint, traffic engineering in WMN poses challenges that are not present in classical network flow optimization, necessitating novel modeling and optimization concepts to account for:

- transmission scheduling on radio links realized by the MAC (medium access control) layer using multiple access schemes such as CSMA (contention-aware carrier sense multiple access, IEEE 802.11) or TDMA (contention-free reservation-based time division multiple access, IEEE 802.16)
- need of (dynamic) channel assignment in multi-channel WMNs
- possibility of node power control for interference mitigation
- possibility of link rate control for adapting the transmission to the channel propagation and interference conditions
- uncertain nature of traffic to be routed between WMN routers and gateways.

Because of interference, it is not likely that transmissions can simultaneously take place on all links of a WMN. Therefore, scheduling algorithms are used to organize the transmissions with the objective of end-to-end throughput optimization. Design and implementation of MAC scheduling is a challenging task, and its impact on traffic performance is a key element in mathematical optimization models for WMN traffic engineering. Moreover, in some types of WMNs, more than one radio channel can be used. This adds a new dimension (and difficulty) to WMN optimization. Finally, additional aspects that increase the complexity of WMN optimization are power control and rate adaptation.

In the context of traffic engineering, the uncertainty in the requirements on the amount of end-to-end information flows poses a major challenge. In WMN, the traffic matrix representing the demands (information flows) to be routed between WMN routers and gateways is virtually unknown. In fact, the classical notion of a traffic matrix is not well suited for WMN design, as the sets of mesh clients connected to the mesh routers are highly dynamic and their traffic profiles are nearly impossible to assess. In this paper, the way of dealing with traffic uncertainty is to incorporate fairness in assigning bandwidth to WMN routes. i.e., to end-to-end flows. As traffic demands in WMN are mostly elastic and QoS is considered only implicitly through lower bounds imposed on the bandwidth assigned to the routes, bandwidth allocation within given bounds using a fairness criterion is appropriate for dealing with the main traffic characteristic of WMN – the aggregates of mesh clients attached to mesh routers are usually downloading as much contents as possible, certainly within some reasonable bounds. To this end, we apply the max-min fairness (MMF) concept for WMN traffic engineering.

In the paper we present a framework for studying multi-commodity flow formulations, and propose mixed-integer programming (MIP) methods and algorithms for optimizing WMN traffic engineering with objectives involving MMF, taking into account peculiarities of radio link modeling, including scheduling, channel assignment, and rate control. The developed methods are investigated by numerical studies for typical WMN scenarios. The results provide insights for the development of link control and routing mechanisms for WMN.

The MMF criterion used in the paper applies to traffic throughput. The throughput is a bandwidth vector assigned to traffic demands in their downstream directions, assuming that a (single) routing path for each demand is fixed and given. The demands for traffic are elastic and therefore they will consume any assigned bandwidth. In effect, the demands compete for link capacities, and the objective of optimization becomes the MMF assignment of bandwidth to the demands. Thus, the paper combines WMN radio link modeling with a non-conventional way of dealing with unknown traffic, a combination that, to our knowledge, has not been considered so far in terms of exact optimization models. This combination forms the key contribution of the paper. More specifically, in the paper we present results along the following lines of research.

- We develop a generic MIP framework to characterize the space of network link capacity vectors being feasible in terms of the signal to interference plus noise ratio (SINR).
- Based on the capacity characterization, we develop a set of MIP formulations for some representative WMN link control mechanisms, maximizing the network traffic throughput with the MMF criterion.
- For the selected link control mechanisms, we provide details of mathematical modeling, using the case with unconstrained scheduling, shortest path routing and MMF traffic flow assignment as the reference solution.
- The developed algorithms are applied in a numerical study of a class of IEEE 802.11 WMNs in order to compare the effectiveness of various link control mechanisms under the fairness objective.

The paper is organized as follows. Section 2 is devoted to a survey of the field. In Section 3, we introduce basic notions and notation. Section 4 presents a set of MIP models representing the link capacity region for a variety of assumptions on radio links in WMNs. Section 5 contains three parts. First, we introduce the notion of MMF and a general algorithm for solving non-convex MMF problems. Second, based on the models from Section 4, we detail the problem settings to be used in our numerical studies. Third, we present the solution approach using the notion of a compatible set. Section 6 presents a heuristic method for solving a WMN traffic engineering problem with MMF. Then, in Section 7 we report and discuss numerical results to illustrate the effectiveness of the optimization approaches, and provide insights into the relation between traffic throughput and system modeling parameters. Conclusions are given in Section 8. Finally, in the two appendices we give some details of MMF modeling.

2. Survey of the field

There is a considerable literature on WMN networking technology and its applications, see, for example, surveys in [7, 8, 19], and introductory discussions in [26, 32]. Also, optimization techniques have been extensively studied in the literature to, generally speaking, maximize the network capacity. Mixed-integer programming models related to WMN network optimization are treated in quite a large number of papers such as [9, 10, 11, 14, 21, 23, 41, 50].

A general (rough) conclusion that can be drawn from these studies is that already for medium networks (with 30 nodes, say) the exact MIP formulations, which are appropriate for the kind of WMN optimization we consider, are hardly solvable to optimality (or even to near-optimality) through direct application of even the top commercial MIP solvers (such as CPLEX). It seems that there is a lot of room for strengthening the existing formulations (i.e., improving the upper bound, see the remark after formula (7) in Subsection 4.1), simplifying them (to decrease the number of binary variables, see Subsection 4.5), and for the use of branch-and-price-and-cut approach in its more sophisticated forms, in order to achieve more time efficient solution algorithms. Such improvements constitute an important direction of further research.

While the aforementioned papers use MIP modeling, their results are not directly usable for our purpose as they assume that demand traffic matrices are known and fixed, i.e., the simplest assumption on traffic data. The assumption is often improper, as uncertainty of traffic demand has to be taken into account in modeling networks carrying unpredictable Internet traffic generated by diverse data, voice and video applications. Today's networks ought to be made robust to traffic uncertainty – this has motivated network planners to extend classical optimization models with demand descriptions that can effectively capture the uncertainty in question. Research in this area (see a recent survey in [22]) falls under two categories: robust optimization and stochastic optimization. Robust optimization assumes a given set of demand matrices, either a large discrete set (multi-scenario description), or a continuous set, typically a polyhedron (polyhedral description) [15], and optimizes the network making sure that each demand matrix from the assumed set will be supported. Stochastic optimization in turn works with a single demand matrix treated as a stochastic matrix, modeling the point-to-point (point-to-multipoint) demands as correlated random variables, and optimizes the network under the requirement that the actually realized demand matrix will be supported with a sufficiently high probability [42]. However, these views of the characteristics of traffic matrix may not be well suited for WMN design for the reasons discussed in the introduction (highly variable sets of mesh clients connected to mesh routers with unpredictable traffic profiles). This, as already mentioned, justifies less classical approaches to traffic modeling as the one used in this paper, that is, max-min fairness (for the notion of MMF see for example [35, 38, 39]).

In [13], it is observed that maximizing the overall throughput of a WMN tends to favor users located close to mesh gateways, while users that have to be reached with more hops are discriminated in capacity allocation. Thus fairness is an issue in experienced service quality, in addition to dealing with traffic uncertainty. The MMF approach tackles fairness by not only maximizing the flows equally assigned to the routes, but also increasing the flows for which this is possible in subsequent iterative steps. In other words, a MMF solution means that no route can gain higher flow without having to decrease the flow on a route where the current flow is lower. Application of the MMF-type traffic engineering to WMN has been considered in [13, 27, 33, 43, 48]. All these papers present heuristic methods (some of them very promising and useful, as in [13, 33, 43]) to optimize the WMN link data rate assignment. By the nature of heuristic methods, the solutions are sub-optimal, and hence do not give a precise solution of globally optimal MMF resource allocation. Our objective is to incorporate the MMF performance metric into MIP formulations for WMN optimization in a rigorous yet efficient way based on our previous contributions in [37, 38, 39]. We note here that an MMF solution is a limiting case of the maximization of the sum of certain concave utility functions (see for example [34]) – this fact can be used for finding approximate MMF solutions as proposed for WMN in [33].

An important, from the perspective of our work, recent paper is [21] where the authors introduce optimization models of WMN for link control (scheduling, channel assignment, power control, and link rate control) and traffic routing. Their basic approach is to solve linear relaxations of the MIP problems formulated in terms of "compatible sets of links" (i.e., sets of links that can transmit simultaneously), and to round-off the fractional solutions obtained through column generation to reach sub-optimal integer solutions. In fact, the use of compatible sets (also called independent sets) was considered as early as in the 90s, see [46]. As a matter of fact, assuming fixed power transmission, the capacity maximization problem can be solved by finding maximum weighted independent sets. Using this approach, for infinite time horizon researchers were maximizing the transmission rates [36, 46] while for finite time horizon, they were minimizing the number of time slots to accommodate the given traffic load, see [49] and references therein.

One capacity aspect modeled in our paper and in the previous work considering WMN optimization is adaptive modulation and coding, also called link rate adaption. Several works have addressed how to take advantage of the multi-channel and multi-rate capabilities of wireless network interfaces used in WMNs, and exploit the channel and path diversity [10, 17, 20]. All these studies point out that the transmission rate used by the wireless interfaces is one of the most important factors that influence the WMN performance. A wireless network interface with multi-rate capabilities should select, at run-time, the transmission rate based on the medium conditions achieving the best link performance. To this end, several rate adaptation algorithms have been proposed in the literature [13, 24, 43]. The approaches either differ in the metrics used to estimate the link quality, or in the rate-selection decision process. Some of the proposed rate adaptation schemes, being compliant with the current 802.11 technology, have been implemented in commodity hardware [13, 24] and used in off-the-shelf products. Recently, experimental studies (e.g., [12]) have investigated how effectively these adaptive rate adaptation algorithms perform in practical settings.

3. Notation

The topology of a WMN network is modeled by a directed graph $\mathcal{N} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V}$ is the set of nodes, $v \in \mathcal{V}$, and $\mathcal{V} = \mathcal{G} \cup \mathcal{R}$, where $\mathcal{G}$ and $\mathcal{R}$ denote the sets of gateways and mesh routers, respectively. The set $\mathcal{E}$ represents the (radio) links, modeled as directed arcs e, $e \in \mathcal{E}$. The head and tail node of a link $e \in \mathcal{E}$ is denoted by $a(e)$ and $b(e)$, respectively, that is, when $e = vw, v, w \in \mathcal{V}$ then $a(e) = v$ and $b(e) = w$. We assume that if $vw \in \mathcal{E}$ then also $wv \in \mathcal{E}$, that is, the arc $e' = wv$

being opposite to an existing arc $e = vw$ also exists (although may not be used). Further, $\delta^+(v)$ and $\delta^-(v)$ denote, respectively, the sets of outgoing and incoming arcs from/to node $v \in \mathcal{V}$, and $\delta(v)$ is the set of all arcs incident to node v, $\delta(v) = \delta^+(v) \cup \delta^-(v)$. A (radio) link $e = vw$ is provided, i.e., $vw \in \mathcal{E}$ only when nodes v and w are able to communicate. As explained below, this is the case when the signal to noise ratio for the v to w communication is greater than a certain threshold.

Below, we are using the standard (dual) description of the quantities related to radio transmissions, using either the linear scale or the logarithmic scale (related to mW or dBm units, respectively). Each quantity expressed in the logarithmic scale and corresponding to a quantity Q expressed in the linear scale will be denoted by $\hat{Q}$. The well known relation between the two quantities is as follows (log denotes the decimal logarithm):

$$\hat{Q} = 10 \log Q, \quad Q = 10^{\frac{\hat{Q}}{10}}.$$

Let P_{vw} denote the power, expressed in mW, received at node $w \in \mathcal{V}$ when node $v \in \mathcal{V}$ is transmitting (as defined above, the same power expressed in dBm is denoted by $\hat{P}_{vw}$). In the numerical studies of this paper we assume the case of 802.11 WMNs operating with an OFDM PHY in the 5 GHz band and use the generic path gain model from [25]. In the decibel scale for a reference distance of 10 m and a path loss exponent of 4, the power received by w when v is transmitting is given by

$$\hat{P}_{vw} = \hat{P} + \hat{G}_{vw} = \hat{P} - 140.046 - 40 \cdot \log d_{vw}, \tag{1}$$

where d_{vw} is the distance between nodes v and w in km and the quantity $\hat{G}_{vw} = -140.046 - 40 \cdot \log d_{vw}$ is called the path gain. Note that we assume that all the nodes use the same transmission power P, where $P = 100$ mW ($\hat{P} = 20$ dBm).

The quantity N will denote the ambient noise power which is set to the product of the thermal noise spectral density $S = -174$ dBm/Hz and the system bandwidth $W = 20$ MHz, i.e., $\hat{N} = -101$ dBm, and in consequence $N = 10^{-10.1}$ mW.

At a given time instance a node can transmit using one of the available *modulation and coding schemes* (MCS). The set of such MCSs is denoted by $\mathcal{M}$, and a particular MCS by $m \in \mathcal{M}$. Table 1, reproduced from [43], shows the set of MCSs available for IEEE 802.11a. Each MCS $m \in \mathcal{M}$ has its signal to interference plus noise ratio (SINR) requirement, i.e., a certain (decoding) threshold γ^m which must be obeyed in order for a transmission to be successful. The SINR thresholds $\hat{\gamma}^m$ and the maximal feasible transmission distances which allow to meet a frame error rate (FER) of 1% when an IP packet with 1500 B payload is transmitted over an additive white gaussian noise channel with bandwidth $W = 20$ MHz are obtained by link level simulations and are given in Table 1. Note that in Table 1, $\hat{\gamma}^m$ is given in the logarithmic scale (in dB) while its value in the

Table 1: IEEE 802.11a MCS, FER $\leq 1\%$, 1500 Byte payload, channel model (1)

MCS m	raw rate B^m	SINR threshold $\hat{\gamma}^m$	max. link length d^m
BPSK 1/2	6 Mbps	3.5 dB	273.5 m
BPSK 3/4	9 Mbps	6.5 dB	230.0 m
QPSK 1/2	12 Mbps	6.6 dB	228.0 m
QPSK 3/4	18 Mbps	9.5 dB	193.7 m
16-QAM 1/2	24 Mbps	12.8 dB	160.2 m
16-QAM 3/4	36 Mbps	16.2 dB	131.7 m
64-QAM 2/3	48 Mbps	20.3 dB	103.8 m
64-QAM 3/4	54 Mbps	22.1 dB	93.5 m

linear scale ($\gamma^m = 10^{\frac{\hat{\gamma}^m}{10}}$) is used in the balance of this paper. The resulting (raw) data rate of a transmission using MCS m is specified by the quantity B^m in Table 1.

The signal to noise ratio (SNR) of a (potential) link $e = vw$ is a constant value equal to

$$\Gamma'_e = \frac{P_{vw}}{N}. \tag{2}$$

We assume that in a network with the given set $\mathcal{M}$ of available MCSs , a link e is provided ($e \in \mathcal{E}$) only if $\Gamma'_e \geq \gamma^m$ for at least one $m \in \mathcal{M}$. In fact, assuming channel model (1), this condition is equivalent to that $d_{vw} \leq d^m$ where d_{vw} is the distance from node v to node w and the maximum distance values d^m are given in Table 1. The set of MCSs for which $d_{vw} \leq d^m$ could be denoted by $\mathcal{M}(e)$ but in the paper we assume (merely to simplify the considerations) that $\mathcal{M}(e) = \mathcal{M}, e \in \mathcal{E}$. The question which MCS to use for which link is addressed by the formulations presented in Sections 4 and 6.1. In the sequel the superscript m will be skipped for the case of a WMN with only one available MCS.

Note that $\Gamma'_e \geq \gamma^m$ means that link e can successfully transmit when there are no other (simultaneous) transmissions in the network. If at certain instance of time t other transmission are on, this may not be possible as other nodes can interfere with the transmission on e. To take this fundamental fact into account define the set $\mathcal{A}(t)$ of nodes active at time instance t ($\mathcal{A}(t) \subseteq \mathcal{V}$). Then the corresponding SINR of link $e = vw$ is defined as

$$\Gamma_e(t) = \frac{P_{vw}}{N + \sum_{a \in \mathcal{A}(t) \setminus \{v\}} P_{aw}} = \frac{P_{vw}}{N + I_{vw}(t)} \tag{3}$$

where the time varying interference, $I_{vw}(t)$, is computed as the sum of the powers received from the nodes which are transmitting at the same time as v. For of convenience, we will drop the time indicator t unless it is necessary. If the interference is zero (no node is transmitting at the same time as v) then $\Gamma_e = \Gamma'_e$, otherwise $\Gamma_e < \Gamma'_e$. We observe that link e can successfully transmit using MCS m at the time instant t only if $\Gamma_e(t) \geq \gamma^m$, that is if

$$\frac{P_{vw}}{N + \sum_{a \in \mathcal{A}(t) \setminus \{v\}} P_{aw}} \geq \gamma^m. \tag{4}$$

Inequality (4) is referred to as the SINR constraint.

Finally, $\mathcal{T}$ denotes the set of (abstract) time slots modeling transmission scheduling. Each slot $t \in \mathcal{T}$ represents a time interval of length τ, with a specified subset of active radio links that can successfully transmit at the same time. The subsets of links active in consecutive time slots $t \in \mathcal{T}$ are specified by a particular scheduling solution. The quantity $T = \tau|\mathcal{T}|$ expresses the length of the *basic cycle* of network operation. A network continuously repeats the cycle, thus potentially operates in an infinite time horizon using time slots of very short duration.

The following list summarizes the introduced notation.

$\mathcal{E}$ – set of directed links (arcs), $e \in \mathcal{E}$

$a(e), b(e)$ – tail and head nodes of link $e \in \mathcal{E}$; we assume that if $vw \in \mathcal{E}$ then also $wv \in \mathcal{E}$

$\mathcal{V}$, $\mathcal{G}$, $\mathcal{R}$ – sets of nodes, gateways, and mesh routers, respectively

$\delta^+(v)$, $\delta^-(v)$ – outgoing and incoming sets of arcs to node $v \in \mathcal{V}$, respectively

$\delta(v)$ – set of all arcs incident to node $v \in \mathcal{V}$; $\delta(v) = \delta^+(v) \cup \delta^-(v)$

P_{vw} – power received at node $w \in \mathcal{V}$ when node $v \in \mathcal{V}$ is transmitting

N – thermal noise power (assumed to be the same for all nodes $v \in \mathcal{V}$)

Γ'_e – signal to noise ratio (SNR) of link e

Γ_e – signal to interference plus noise ratio (SINR) of link e

$\mathcal{M}$ – set of available MCSs, $m \in \mathcal{M}$

γ^m – SINR requirement or decoding threshold for MCS $m \in \mathcal{M}$ (γ if only one MCS is used)

B^m – data rate of MCS $m \in \mathcal{M}$ (B if only one MCS is used)

$\mathcal{T}$ – set of time slots modeling transmission scheduling

τ – the length/duration of a time slot

T – length of a cycle; $T = \tau|\mathcal{T}|$.

We end this section with specifying the basic variables used in the optimization models considered in the next sections.

Y_{et} – binary variable indicating whether link e is scheduled to be active in time slot $t \in \mathcal{T}$. If a link is active, then (4) must be satisfied for the SINR threshold of the appropriate MCS.

y_{et}^m – binary variable indicating whether link e is active and uses MCS m in time slot $t \in \mathcal{T}$.

X_{vt} – binary variable indicating whether node v transmits (is active) in time slot $t \in \mathcal{T}$.

x_{vt}^m – binary variable indicating whether node v transmits using MCS m in time slot $t \in \mathcal{T}$.

Z_v^m – binary variable indicating which MCS is used at node v in all time slots (only when the static MCS assignment is used).

r_{et} – nonnegative continuous capacity of link e achieved in time slot $t \in \mathcal{T}$.

c_e – nonnegative continuous capacity of link e, equal to the total volume of data that can be send on link e during time T.

4. Modeling link capacity space

This section is devoted to modeling feasible sets of radio link capacity reservations corresponding to possible variants of WMN systems. For this purpose we use mixed-integer programming formulations, that is, systems of linear equalities/inequalities in continuous/integer variables. Starting with a formulation for a simple WMN system, we subsequently present its extensions capturing more advanced features of WMN system configurations. The models of this section are independent of any particular optimization objective formulations and are in this sense general. They will be used in the subsequent sections for defining and solving network optimization problems involving MMF. In all the cases we assume that at any time slot each mesh node is either active (and in this case it transmits) or is inactive, and so are the links.

Certainly, many other variants of capacity reservation models relevant to WMN systems could be considered besides those studied in this section. One important class of such extensions is obtained when the power control is admitted in the nodes (i.e., when P_{vw} become variables). As discussed in [21], this allows to limit the interferences and in consequence to increase the network capacity.

Another extension is to model multiple radio channels with separate interference domains. A channel is assigned to a link either statically (the same channel for all time slots) or dynamically (channels assigned to a link can change from a time slot to a time slot). Clearly, interference constraints (4) would be then considered separately for transmissions on different channels. Such a multi-transceiver, multiple-MCS, multi-channel model consistent with the notation used in this paper can be found in [40]. Also, the OFDMA technology can be modeled in a similar way with the OFDMA subcarriers (roughly) corresponding to transmission channels. Still another extension could be obtained when multiple directed antennas can be used in the nodes. Then the interference constraint (4) would be transformed to a much less demanding one.

In the paper we limited the set of the considered models to keep the paper volume on an acceptable level.

4.1. Single MCS

The following formulation describes a very simple WMN system where each node has only one transceiver, and only one fixed MCS is used by all the nodes.

$$\sum_{e\in\delta(v)} Y_{et} \leq 1 \qquad v\in\mathcal{V}, t\in\mathcal{T} \tag{5a}$$

$$\sum_{e\in\delta^{+}(v)} Y_{et} = X_{vt} \qquad v\in\mathcal{V}, t\in\mathcal{T} \tag{5b}$$

$$Y_{et}\frac{P_{a(e)b(e)}}{N+\sum_{v\in\mathcal{V}\setminus\{a(e)\}} P_{vb(e)}X_{vt}} \geq Y_{et}\gamma \qquad e\in\mathcal{E}, t\in\mathcal{T} \tag{5c}$$

$$c_e = \tau B\sum_{t\in\mathcal{T}} Y_{et} \qquad e\in\mathcal{E}. \tag{5d}$$

The constraints assure that:

(5a): at most one link e incident to node v can be active in time slot t.

(5b): exactly one link e outgoing from node v is active in time slot t if the node is active in this time slot.

(5c): SINR constraint (see (4)); if link e is active in time slot t, i.e., if $Y_{et} = 1$, then its SINR must be greater or equal to the decoding threshold.

(5d): total effective capacity of link e is the sum of its capacities over all time slots; note that because B is expressed in Mbps, the capacity is the total number of bits transmitted on the link in the time frame T containing all time slots (i.e., the entire duration of the cycle, $T = \tau|\mathcal{T}|$).

Notice that in formulation (5), X_{vt} are auxiliary variables and thus they can either be eliminated or their binarity can be skipped. Moreover, observe that (5c) is not linear but bi-linear. It can be made linear by introducing additional (continuous) variables z_{evt} to express the product $Y_{et}\cdot X_{vt}$, i.e., $z_{evt} = 1$ if both Y_{et} and X_{vt} are equal to 1, and 0, otherwise. This is achieved by adding the constraints

$$z_{evt} \geq Y_{et} + X_{vt} - 1 \qquad v\in\mathcal{V}, e\in\mathcal{E}, t\in\mathcal{T} \tag{6a}$$

$$z_{evt} \leq Y_{et},\ z_{evt} \leq X_{vt} \qquad v\in\mathcal{V}, e\in\mathcal{E}, t\in\mathcal{T} \tag{6b}$$

$$z_{evt} \geq 0 \qquad v\in\mathcal{V}, e\in\mathcal{E}, t\in\mathcal{T} \tag{6c}$$

and changing (5c) to

$$NY_{et} + \sum_{v \in \mathcal{V} \setminus \{a(e)\}} P_{vb(e)} z_{evt} \leq \frac{1}{\gamma} P_{a(e)b(e)} Y_{et} \qquad e \in \mathcal{E}, t \in \mathcal{T}. \tag{7}$$

It is important to note that although variable Y_{et} is not necessary on the left hand side of (5c), its presence strengthens formulation (5) because it makes it possible to avoid using the so called "big M" in constraints (6) and (7), contrary to the analogous linearization in [21]. The lack of "big M" in our formulations considerably decreases the upper bounds of the corresponding linear relaxations, and therefore improves their tractability (see Section 7).

In effect, we arrive at a mixed-integer programming (MIP) formulation (mixed-binary, in fact) that expresses constraints for feasible allocations of link capacities, and, in fact, for transmission scheduling within individual time slots. The so specified optimization space $\mathcal{C}$ for variables $c = (c_e,\ e \in \mathcal{E})$ can then be used to formulate optimization problems involving optimization of traffic objectives for WMN. In Section 5 this is done for the MMF allocation of flows to paths.

4.2. Single MCS, node capacity split

In the system modeled by (5), the entire transmission capacity τB of active node $v \in \mathcal{V}$ in a time slot is allocated only to one selected link outgoing from node v. This assumption can in fact be alleviated by allowing the system to arbitrarily split the transmission capacity within a time slot among all links outgoing from a node. Such a split could be realized through a suitable protocol (there is a similarity to the routing protocol realizing the ECMP feature in OSPF). As shown in Section 7, this new degree of freedom (referred to as node capacity split) in allocating capacity to links can help to improve the efficiency of network control for example by using fewer time slots for achieving the same capacity allocation. Below we present a formulation assuming the node capacity split. Note that, in the considered problem, if multiple outgoing links of a node become active during a time slot, they are restricted to use a common MCS.

$$\sum_{e \in \delta^-(v)} Y_{et} \leq 1 \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{8a}$$

$$Y_{et} + Y_{e't} \leq 1 \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{8b}$$

$$Y_{et} \leq X_{vt} \qquad v \in \mathcal{V}, e \in \delta^+(v), t \in \mathcal{T} \tag{8c}$$

$$z_{evt} \geq Y_{et} + X_{vt} - 1 \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T} \tag{8d}$$

$$z_{evt} \leq Y_{et},\ z_{evt} \leq X_{vt} \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T} \tag{8e}$$

$$NY_{et} + \sum_{v \in \mathcal{V} \setminus \{a(e)\}} P_{vb(e)} z_{evt} \leq \frac{1}{\gamma} P_{a(e)b(e)} Y_{et} \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{8f}$$

$$r_{et} \leq \tau B Y_{et} \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{8g}$$

$$\sum_{e \in \delta^+(v)} r_{et} \leq \tau B \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{8h}$$

$$c_e = \sum_{t \in \mathcal{T}} r_{et} \qquad e \in \mathcal{E}. \tag{8i}$$

The constraints assure that:

(8a): at most one link e incoming to node v can be active in time slot t.

(8b): only one arc in any pair of opposite arcs can be active in time slot t.

(8c): potentially all links outgoing from any node v can be active in time slot t.

(8d)-(8f): admissible SINR for link e in time slot t; note that the linearization discussed in Section 4.1 is used.

(8g)-(8h): volume of data transmitted from node v in a time slot t can be split among the active links outgoing from the node.

(8i): total effective capacity of link e is the sum of its capacities over all time slots.

Note that, because of (8c), variables X_{vt} cannot be eliminated from the formulation like in (5). However, still their binarity can be relaxed.

Contrary to (5a), in (8a) only links incoming to v are considered. It may create an impression that node v can receive and transmit simultaneously. This is not the case, because of (8f), i.e., if v transmits then it jams all the transmissions to v. Also note that we have assumed that a node cannot use more than one incoming link in a time slot. This assumption is not only justified, but also inevitable. Assume that node v can use two incoming links, say wv and $w'v$. Then the nodes w and w' cannot transmit simultaneously when one of them is transmitting to v, as they would jam each other. But they do, as they both transmit during the whole time slot. Thus, we reach a contradiction.

After straightforward adaptations, the node capacity split extension applies to all subsequent model formulations in this section. Still, we will not include this option in the rest of the models presented in this section. Consequently, we will not indicate this fact in the titles of the subsections.

4.3. Static allocation of MCS

The next problem we consider assumes that a set of possible MCSs is available. Still, each node uses only one MCS which is fixed in all time slots. This is referred to as static allocation of MCS to nodes, a technique which is used by all state-of-the-art WiFi access points.

$$\sum_{m\in\mathcal{M}} Z_v^m = 1 \qquad v \in \mathcal{V} \tag{9a}$$

$$x_{vt}^m \le Z_v^m \qquad v \in \mathcal{V}, m \in \mathcal{M}, t \in \mathcal{T} \tag{9b}$$

$$X_{vt} = \sum_{m\in\mathcal{M}} x_{vt}^m \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{9c}$$

$$\sum_{e\in\delta^+(v)} y_{et}^m = x_{vt}^m \qquad v \in \mathcal{V}, m \in \mathcal{M}, t \in \mathcal{T} \tag{9d}$$

$$Y_{et} = \sum_{m\in\mathcal{M}} y_{et}^m \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{9e}$$

$$\sum_{e\in\delta(v)} Y_{et} \le 1 \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{9f}$$

$$z_{evt}^m \ge y_{et}^m + X_{vt} - 1 \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{9g}$$

$$z_{evt}^m \le y_{et}^m,\ z_{evt}^m \le X_{vt} \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{9h}$$

$$N y_{et}^m + \sum_{v\in\mathcal{V}\setminus\{a(e)\}} P_{vb(e)} z_{evt}^m \le \frac{1}{\gamma^m} P_{a(e)b(e)} y_{et}^m \qquad e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{9i}$$

$$c_e = \tau \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{M}} B^m y_{et}^m \qquad e \in \mathcal{E}. \tag{9j}$$

The constraints assure that:

(9a): every node selects one MCS.

(9b): MCS used by a node in any time slot in which the node is active must be the one selected for this node.

(9c): defines X_{vt} indicating whether or not node v is active in time slot t.

(9d): if node v is active in time slot t then exactly one link outgoing from node v is active in time slot t and uses the MCS of the node.

(9e): $Y_{et} = 1$ if, and only if, link e is active and uses one of the available MCSs.

(9f): at most one link incident to a node can be active in any time slot.

(9g)-(9i): admissible SINR for link e with MCS m in time slot t.

(9j): total effective capacity of link e is the sum of its capacities over all time slots and MCSs.

A modified version of the above formulation was considered in [43, 44]. In the modification, MCSs are fixed to links rather than to nodes so that each link uses the same MCS in all time slots (when active) but different links outgoing from a node can use different MCSs. This particular version of (9) allows to address the trade-off between spatial reuse and throughput and is used in the corresponding formulations of Subsection 5.2 and assumed for the heuristic presented in Section 6.

4.4. Dynamic allocation of MCS

The following formulation corresponds to a system where multiple MCSs are available, and the nodes can use different MCSs in different time slots. The latter feature is referred to as dynamic allocation of MCS to nodes.

$$X_{vt} = \sum_{m \in \mathcal{M}} x_{vt}^m \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{10a}$$

$$Y_{et} = \sum_{m \in \mathcal{M}} y_{et}^m \qquad e \in \mathcal{E}, t \in \mathcal{T} \tag{10b}$$

$$\sum_{e \in \delta(v)} Y_{et} \le 1 \qquad v \in \mathcal{V}, t \in \mathcal{T} \tag{10c}$$

$$\sum_{e \in \delta^+(v)} y_{et}^m = x_{vt}^m \qquad v \in \mathcal{V}, m \in \mathcal{M}, t \in \mathcal{T} \tag{10d}$$

$$z_{evt}^m \ge y_{et}^m + X_{vt} - 1 \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{10e}$$

$$z_{evt}^m \le y_{et}^m,\ z_{evt}^m \le X_{vt} \qquad v \in \mathcal{V}, e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{10f}$$

$$N y_{et}^m + \sum_{v \in \mathcal{V} \setminus \{a(e)\}} P_{vb(e)} z_{evt}^m \le \frac{1}{\gamma^m} P_{a(e)b(e)} y_{et}^m \qquad e \in \mathcal{E}, t \in \mathcal{T}, m \in \mathcal{M} \tag{10g}$$

$$c_e = \tau \sum_{t \in \mathcal{T}} \sum_{m \in \mathcal{M}} B^m y_{et}^m \qquad e \in \mathcal{E}. \tag{10h}$$

The constraints assure that:

(10a): node v can use at most one MCS in time slot t.

(10b): link e can use at most one MCS in time slot t.

(10c): at most one link e incident to node v can be active in time slot t.

(10d): exactly one link e outgoing from node v is active and uses MCS m in time slot t if the node is active in this time slot and uses this MCS.

(10e)-(10g): admissible SINR for link e using MCS m in time slot t (this constraint is active if link e is active and uses MCS m in time slot t).

(10f): total effective capacity of link e is the sum of its capacities over all MCS in all time slots.

4.5. A simplified interference model

An important modification of the considered feasible link capacity space characterization that leads to a significant decrease of the time needed to solve the related optimization problems (see the formulations in Section 5 and the numerical results in Section 7) is a simplification of the interference model. In the cases considered earlier in this section, a full interference model was assumed. However, it is possible to reasonably simplify it like in [13, 43] where only the first order interference was considered using the notion of the so called collision domains. The idea of the simplification is to assume that if any subset of nodes $\mathcal{W} \subseteq \mathcal{V} \setminus \{a(e)\}$, when simultaneously active, interferes with transmission on link e then transmission on link e is already interfered when any single node $v \in \mathcal{W}$ is active. Certainly, in general this is not the case and therefore may lead to network capacity overestimation.

In order to formally incorporate the simplified interference model into for example (5), we replace (5c) with the following constraint:

$$Y_{et} \frac{P_{a(e)b(e)}}{N + P_{vb(e)} X_{vt}} \geq Y_{et}\gamma \quad e \in \mathcal{E},\ t \in \mathcal{T},\ v \in \mathcal{V} \setminus \{a(e)\}. \tag{11}$$

Now it is possible to rewrite (11) in a linear form (with no extra variables, contrary to (6)-(7)) in the following way:

$$Y_{et} + X_{vt} \leq 1, \quad e \in \mathcal{E},\ t \in \mathcal{T},\ v \in \mathcal{W}_e \tag{12}$$

where

$$\mathcal{W}_e = \{v \in \mathcal{V} \setminus \{a(e)\} :\ \frac{P_{a(e)b(e)}}{N + P_{vb(e)}} < \gamma\}. \tag{13}$$

As illustrated in Section 7, it turns out that constraint (12), being what we may call the first order SINR constraint, leads to much more efficient optimization problem formulations than (6) and (7) and therefore is computationally advantageous. Moreover, taking into account only first order SINR allows for devising efficient heuristics – see Section 6.

In the case of multiple MCSs, the corresponding modification of the SINR constraint is analogous. Hence, the modification can be applied to all problems considered in this paper by replacing the original SINR constraints (involving additional variables z) with an appropriate constraint similar to (12).

5. MMF optimization of demand flows

5.1. Application of the MMF approach

Below we shall formulate a general, from the point of view of the models of Section 4, max-min fair (MMF) flow optimization problem. We assume that link capacity reservation variables

$c = (c_e : e \in \mathcal{E})$ belong to a feasible set $\mathcal{C} \subseteq \mathbf{R}^{|\mathcal{E}|}$. The set $\mathcal{C}$ can be defined as in any of the formulations described in Section 4. Note that all of those sets are not convex.

Let $\mathcal{P} = \{p_1, p_2, \ldots, p_D\}$ be the given set of paths between nodes and gateways. Each such path p_d is supposed to carry the entire downstream flow f_d from a selected gateway (gateways form a subset $\mathcal{G}$ of $\mathcal{V}$) to a destination node in $\mathcal{D} = \mathcal{V} \setminus \mathcal{G}$, where $D = |\mathcal{D}|$. (Upstream flow can be accounted in a similar way, although it is not considered here, see also Section 1.) Each path p_d is treated as a subset of links, i.e., $p_d \subseteq \mathcal{E}, d = 1, 2, \ldots, D$. For each link $e \in \mathcal{E}$, the set of all indices of paths in $\mathcal{P}$ that contain this link will be denoted by $\mathcal{Q}_e = \{d : \ e \in p_d, \ 1 \leq d \leq D\}$. The MMF flow optimization problem is defined as follows:

$$\text{lexmax} \quad [(f_1, f_2, \ldots, f_D)] \tag{14a}$$

$$c \in \mathcal{C} \tag{14b}$$

$$\sum_{d \in \mathcal{Q}_e} f_d \leq c_e \qquad\qquad e \in \mathcal{E}. \tag{14c}$$

In (14a), $[f] = ([f]_1, [f]_2, \ldots, [f]_D)$ denotes the vector $f = (f_1, f_2, \ldots, f_D)$ sorted in the non-decreasing order, so that objective (14a) consists in lexicographical maximization of the sorted vector of path-flows $([f]_1 \leq [f]_2 \leq \ldots \leq [f]_D)$.

Problem (14) can be solved sequentially through the so called *conditional means approach* described in Section 4.2 of [38]. The idea of the approach is to express the consecutive entries of vector $[f]$ (these entries are referred to as *optimization criteria*) and maximize them one by one, keeping the previously optimized entries at their optimal values. For the details of the derivations given below, the reader is referred to [38].

Let the *cumulated optimization criteria* be defined as $F_d = [f]_1 + [f]_2 + \ldots + [f]_d, \ d = 1, 2, \ldots, D$. Each such cumulated criterion F_d (for a fixed $d, \ 1 \leq d \leq D$) can be expressed as the solution of a linear programming problem in continuous variables b_d and $g_{dj}, j = 1, 2, \ldots, D$:

$$F_d = \max \, \{d b_d - \sum_{j=1}^{D} g_{dj}\} \tag{15a}$$

$$b_d - f_j \leq g_{dj} \qquad\qquad j = 1, 2, \ldots, D \tag{15b}$$

$$g_{dj} \geq 0 \qquad\qquad j = 1, 2, \ldots, D. \tag{15c}$$

The justification and derivation of (15) is given in Appendix A. Note that the entities f_d in (15b) are the entries of the flow vector f – they are variables in formulation (14) but are given constants in formulation (15). Because of (15a), the following algorithm solves the MMF flow allocation problem.

Algorithm 1 (Algorithm for problem (14))

Input: Optimization space $\mathcal{C}$.

Output: MMF solution $f^0 = (f_1^0, f_2^0, \ldots, f_D^0)$ and corresponding optimal capacities c^0.

Step 0: Set $d = 1$.

Step 1: Solve the mathematical program:

$$\text{maximize} \quad db_d - \textstyle\sum_{j=1}^{D} g_{dj} \tag{16a}$$

$$\text{s.t.} \quad c \in \mathcal{C} \tag{16b}$$

$$\textstyle\sum_{d \in \mathcal{Q}_e} f_d \leq c_e, \quad e \in \mathcal{E} \tag{16c}$$

$$F_d^0 \leq ib_j - \textstyle\sum_{j=1}^{D} g_{ij}, \quad i = 1, 2, \ldots, d-1 \tag{16d}$$

$$b_i - f_j \leq g_{ij}, \quad j = 1, 2, \ldots, D,\ i = 1, 2, \ldots, d \tag{16e}$$

$$g_{ij} \geq 0, \quad j = 1, 2, \ldots, D,\ i = 1, 2, \ldots, d \tag{16f}$$

and denote the resulting optimal objective by F_d^0.

Step 2: If $d < D$ put $d := d+1$ and go to Step 1. Otherwise, stop: $f_1^0 = F_1^0, f_d^0 = F_d^0 - F_{d-1}^0, d = 2, 3, \ldots, D$ and the last optimized capacities c^0 compose the optimal solution).

It is worth mentioning that the difficulty in solving problems (16) for consecutive $d, d = 1, 2, \ldots, D$ lies in non-convexity of the set $\mathcal{C}$ and not in additional (continuous) variables b_d $(d = 1, 2, \ldots, D)$ and g_{dj} $(d, j = 1, 2, \ldots, D)$ nor in (linear) constraints (16c)-(16f).

Remark: The number of steps of the algorithm is equal to the number of demands, i.e., to D. In each step we fix one (consecutive) value of the MMF vector of bandwidth assignments. In fact there exists another approach (called the *distribution approach*, see Section 4.3 of [38]) to non-convex MMF problems. The complexity of this alternative approach is similar to that of the above approach. The distribution approach requires that the set of values assumed by the criteria $f_1, f_2, \ldots, f_D$ is finite, and then the number of steps is equal to the number of distinct values assumed by the criteria in the optimal vector f^0.

5.2. The considered MMF problems

For the numerical studies presented in Section 7 we have selected the following link capacity models corresponding to different WMN systems described in Section 4:

$\mathcal{C}_{static}^{FI}$ – model (9) with static MCS allocation with the modification described at the end of Section 4.3 (*static allocation with full interference*)

$\mathcal{C}_{static}^{SI}$ – model (9) with static MCS allocation with the modification described at the end of Section 4.3 and with constraints (9g)-(9i) substituted with (12) (*static allocation with simplified interference*)

$\mathcal{C}^{FI}_{dynamic}$ – model (10) with dynamic MCS allocation of Section 4.4 (*dynamic allocation with full interference*)

$\mathcal{C}^{SI}_{dynamic}$ – model (10) with dynamic MCS allocation of Section 4.4 and with constraints (10e)-(10g) substituted with an appropriate version of (12) (*dynamic allocation with simplified interference*).

Additionally we will consider modifications of the above models assuming node capacity split described in Section 4.2. When this option is assumed, we will denote this by adding $/NS$ to the superscript, for example $\mathcal{C}^{FI/NS}_{static}$. The resulting MMF optimization problems are then obtained by substituting $\mathcal{C}$ in constraint (14b) of the generic MMF problem (14) by:

Problem SA/FI (*Static Allocation with Full Interference*): $\mathcal{C} := \mathcal{C}^{FI}_{static}$

Problem DA/FI (*Dynamic Allocation with Full Interference*): $\mathcal{C} := \mathcal{C}^{FI}_{dynamic}$

Problem SA/FI/NS (*Static Allocation with Full Interf. and Node Capacity Split*): $\mathcal{C} := \mathcal{C}^{FI/NS}_{static}$

Other MMF problems that correspond to other feasible capacity sets defined previously are obtained analogously. Hence, problems abbreviated with DA/FI/NS, DA/SI, DA/SI/NS, SA/SI and SA/SI/NS are defined by substituting $\mathcal{C}$ in constraint (14b) with $\mathcal{C}^{FI/NS}_{dynamic}$, $\mathcal{C}^{SI}_{dynamic}$, $\mathcal{C}^{SI/NS}_{dynamic}$, $\mathcal{C}^{SI}_{static}$ and $\mathcal{C}^{SI/NS}_{static}$, respectively.

5.3. Non-compact formulation and compatible sets generation

Problems of the form (16) solved in Step 1 of the MMF algorithm (Algorithm 1) for the formulations described at the end of the previous subsection (SA/FI, DA/FI, etc.) are hard (see Section 7). Fortunately, it is possible to considerably simplify them (especially problems of the DA type, see below) by using the so called *compatible sets* [21]. When a single MCS is assumed, a compatible set (CS) is defined as a subset $\mathcal{E}_i$ of links ($\mathcal{E}_i \subseteq \mathcal{E}$) that can transmit simultaneously without interfering with each other. In other words, a compatible set is defined by $\mathcal{E}_i = \{e \in \mathcal{E} : \ Y_e = 1\}$ for any set of (feasible) link variables $Y_e, \ e \in \mathcal{E}$ fulfilling (19b)-(19h). For the case with multiple MCS, the definition of $\mathcal{E}_i$ includes not only a subset of links but also a list of particular MCSs that each link is using. Hence, $\mathcal{E}_i = \{(e, m) \in \mathcal{E} \times \mathcal{M} : \ y^m_e = 1\}$ where variables y^m_e form a feasible solution for a corresponding problem from Section 4.3 or Section 4.4 with only one time slot, i.e., when $|\mathcal{T}| = 1$ and the subscript t is skipped in formulations (9) and (10).

Let $\mathcal{I}$ be a family of compatible sets. Using $\mathcal{I}$, we can introduce alternative formulations of the previously considered MIP problems, i.e., formulations involving time slots $t \in \mathcal{T}$. In fact, the compatible-set formulations are continuous approximations of the MIP problems, with continuous variables z_i that represent the portion of time T utilizing the transmission scheme defined by a compatible set $i \in \mathcal{I}$. The compatible-set formulations are non-compact, since $|\mathcal{I}|$ grows exponentially in the network size. One of advantages of using compatible sets is that there is no need of treating time slots explicitly, as the optimization task boils down to determining how many time slots a compatible set should be allocated within the whole cycle $T = \tau|\mathcal{T}|$ – this number is simply proportional to $\frac{z_i}{T}$. This type of approximation has been shown to be very accurate in [18]. Note that when the length T of the cycle grows, the approximation will approach an integer optimum.

As an example of a compatible-set formulation, consider the problem solved in the first step of the MMF algorithm (maximization of a minimal flow) for the simplest WMN system considered in Section 4.1 defined by the set of constraints (5a)-(5b), (6), (7), and (5d), i.e., for all DA/FI, and SA/FI with $|\mathcal{M}| = 1$.

$$\max \quad f \tag{17a}$$

$$[\alpha] \quad \sum_{i \in \mathcal{I}} z_i = T \tag{17b}$$

$$[\lambda_d \geq 0] \quad f \leq f_d \qquad d \in \mathcal{D} \tag{17c}$$

$$c_e = \sum_{i \in \mathcal{I}} B_{ei} z_i \qquad e \in \mathcal{E} \tag{17d}$$

$$[\pi_e \geq 0] \quad \sum_{d \in \mathcal{Q}_e} f_d \leq c_e \qquad e \in \mathcal{E} \tag{17e}$$

$$z \geq 0. \tag{17f}$$

In the above formulation, $d \in \mathcal{D}$ are the demands (abbreviated $d = 1, 2, \ldots, D$ earlier in this section), and B_{ei} is the rate allocated to link $e \in \mathcal{E}$ in compatible set $i \in \mathcal{I}$, i.e., either B or 0, depending on whether link e is active or not in compatible set i. The entities shown within the brackets denote the dual variables. Note that the non-compactness of formulation (17) is implied by the exponential number of potential compatible sets.

The dual problem (for the notion of duality see for example [31]) corresponding to the primal problem (17) with the current list of compatible sets $\mathcal{I}$ reads:

$$\begin{aligned}
\min \quad & \alpha T && && \text{(18a)}\\
& \sum_{d\in\mathcal{D}} \lambda_d = 1 && && \text{(18b)}\\
& \sum_{e\in\mathcal{E}} \pi_e B_{ei} \leq \alpha && i \in \mathcal{I} && \text{(18c)}\\
& \lambda_d \leq \sum_{e\in p_d} \pi_e && d \in \mathcal{D} && \text{(18d)}\\
& \lambda, \pi \geq 0. && && \text{(18e)}
\end{aligned}$$

Let α^*, π^* be an optimal solution of the dual problem. A compatible set is generated by solving the following problem.

$$\begin{aligned}
\max \quad & \sum_{e\in\mathcal{E}} \pi_e^* B_e && && \text{(19a)}\\
& \sum_{e\in\delta(v)} Y_e \leq 1 && v \in \mathcal{V} && \text{(19b)}\\
& \sum_{e\in\delta^+(v)} Y_e = X_v && v \in \mathcal{V} && \text{(19c)}\\
& Z_{ev} \geq Y_e + X_v - 1 && v \in \mathcal{V}, e \in \mathcal{E} && \text{(19d)}\\
& Z_{ev} \leq Y_e,\ Z_{ev} \leq X_v && v \in \mathcal{V}, e \in \mathcal{E} && \text{(19e)}\\
& NY_e + \sum_{v\in\mathcal{V}\setminus\{a(e)\}} P_{vb(e)} Z_{ev} \leq \frac{1}{\gamma} P_{a(e)b(e)} Y_e && e \in \mathcal{E} && \text{(19f)}\\
& B_e = BY_e && e \in \mathcal{E} && \text{(19g)}\\
& Y_e \in \{0,1\} && e \in \mathcal{E}. && \text{(19h)}
\end{aligned}$$

Above, B_e is the rate allocated to link e in the optimized compatible set, i.e., either B or 0. The variables and constraints have the same interpretation as in Section 4.1 assuming only one time slot t (and hence skipping the subscript t) as we are looking for a compatible set for an arbitrary portion of time. If the result of the optimization is greater than α^* then the generated compatible set defined by $\{e \in \mathcal{E} : Y_e = 1\}$ is added to $\mathcal{I}$ as it may improve the optimal primal solution.

Note that generation of compatible sets for problem DA/FI (based on the link capacity model (10) from Section 4.4) looks essentially the same. The only difference is that in the compatible set generation subproblem (19), constraints (19b)-(19g) should be substituted by their counterparts based on constraints from (10) involving index m.

For problem SA/FI (based on (9) from Section 4.3), the use of compatible sets becomes more complicated as it requires additional constants an variables to formulate an appropriate (non-compact) primal problem. Denote by u_{vi}^m a binary constant equal to 1 if, and only if, a compatible set i consists of node v transmitting using MCS m. By x_v^m denote a binary variable equal to 1 if node v can use MCS m, and 0 otherwise. Then the problem can be formulated as follows.

$$
\begin{aligned}
& & \max \quad f & & \quad (20\text{a})\\
&[\alpha] & \sum_{i \in \mathcal{I}} z_i = T & & \quad (20\text{b})\\
&[\beta_{vi}^m \geq 0] & u_{vi}^m z_i \leq x_v^m T & \qquad v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I} & \quad (20\text{c})\\
&[\zeta_v \geq 0] & \sum_{m \in \mathcal{M}} x_v^m \leq 1 & \qquad v \in \mathcal{V} & \quad (20\text{d})\\
&[\lambda_d \geq 0] & f \leq f_d & \qquad d \in \mathcal{D} & \quad (20\text{e})\\
& & c_e = \sum_{i \in \mathcal{I}} B_{ei} z_i & \qquad e \in \mathcal{E} & \quad (20\text{f})\\
&[\pi_e \geq 0] & \sum_{d \in \mathcal{Q}_e} f_d \leq c_e & \qquad e \in \mathcal{E} & \quad (20\text{g})\\
& & z \geq 0. & & \quad (20\text{h})
\end{aligned}
$$

Formulation (20) extends (17) with two additional constraints (20c) and (20d) that let each node use only one MCS. The problem dual to (20) reads:

$$
\begin{aligned}
\min \quad & \alpha T + \sum_{v \in \mathcal{V}} \zeta_v & & \quad (21\text{a})\\
& \sum_{d \in \mathcal{D}} \lambda_d = 1 & & \quad (21\text{b})\\
& \sum_{e \in \mathcal{E}} \pi_e B_{ei} - \sum_{v \in \mathcal{V}} \sum_{m \in \mathcal{M}} \beta_{vi}^m u_{vi}^m \leq \alpha & \qquad i \in \mathcal{I} & \quad (21\text{c})\\
& \lambda_d \leq \sum_{e \in p_d} \pi_e & \qquad d \in \mathcal{D} & \quad (21\text{d})\\
& \sum_{i \in \mathcal{I}} \beta_{vi}^m T \leq \zeta_v & \qquad v \in \mathcal{V}, m \in \mathcal{M} & \quad (21\text{e})\\
& \lambda, \pi \geq 0. & & \quad (21\text{f})
\end{aligned}
$$

Although formulation (21) looks more complicated than (18), it turns out that compatible sets are generated in the same way as before, i.e., by solving (19). The reason is that in an optimal dual solution all variables β_{vi}^m can be made equal to 0.

Solving (20) is much more complicated than solving (17), because it consists of binary variables and in consequence requires the branch-and-price approach [6, 39]. When at a certain node of the branch-and-bound tree a variable x_v^m becomes fixed (to 0 or 1) then all the compatible sets used in the corresponding branch-and-bound subtree must obey this. This requires invoking a compatible set generation procedure in each node of the branch-and-bound tree.

It is worth to notice that the compatible set generation problem becomes a lot simpler when the simplified interference model (SI) is used. In such a case, the generation problem is just a well-known Maximum Weight Matching extensively studied in the literature. Moreover, it was proved that the scheduling scheme that solves Maximum Weight Matching can achieve optimal throughput performance, see [46] for the back-pressure scheme, and [36] for the max weight scheduling.

Certainly, for further steps of the MMF algorithm from Subsection 5.1, the compatible set generation problem remains the same. However, since the CS formulations (as (17)) are linear, a more efficient MMF algorithm specialized for convex MMF problems is applicable. Such an algorithm is presented in Appendix B.

5.4. Practicability of the models' solutions

Note that when solving (17) or (20) we obtain a vector z^* of optimal periods of time each compatible set should be used. Then we may split the time z_i^* allocated to each compatible set i among the time slots. If the number of time slots $|\mathcal{T}|$ is large (recall that $|\mathcal{T}|\tau = T$), then the resulting solution will be an excellent approximation of an exact solution of a corresponding problem involving time slots as in formulations of Section 4.

First note that the number of non-zero columns in an optimal solution provided by Simplex cannot be greater than the number of rows, which in case of (17) is $|\mathcal{D}| + |\mathcal{E}| + 1$. Assume that all paths have positive flows $f_d^* > 0, d \in \mathcal{D}$, so all the columns corresponding to f_d and a column corresponding to f are non-zero. Thus only $|\mathcal{E}|$ of other variables can be greater than zero. It implies that no more than $|\mathcal{E}|$ compatible sets can be used in an optimal solution.

Assume now that the optimal vector z^* and the number of time slots $|\mathcal{T}|$ are given. It is easy to see that in order to obtain a vector $\hat{z}$, such that $\hat{z}_i \geq z_i^*$ and $\hat{z}_i = \tau L$, (L non-negative integer), for all $i \in \mathcal{I}$, we need at most $|\mathcal{T}| + |\mathcal{E}| - 1$ time slots. Analogously, to obtain a vector $\hat{z}$ for a vector z, such that $z_i = z_i^* \frac{|\mathcal{T}|-|\mathcal{E}|+1}{|\mathcal{T}|}$, for all $i \in \mathcal{I}$, we need no more than $|\mathcal{T}|$ time slots. Thus the result provided by the compatible-set approach translated into a given number of time slots cannot be farther than $\frac{|\mathcal{E}|-1}{|\mathcal{T}|} \cdot 100\%$ from the optimum.

6. A heuristic approach to SA/SI

The link capacity optimization space of WMNs, as introduced in Section 4, is described by a large number of integer variables and is therefore hard to efficiently handle in optimizations. Therefore, heuristic methods for WMN optimization are in place. In this section we describe a heuristic sub-optimal method, referred to as LBA (load based algorithm), for optimizing the MMF throughput in WMN, namely for solving problem SA/SI. LBA is a two phase algorithm. In the first phase (Subsection 6.1), LBA determines the assignment of MCSs to the links in a simple way called adaptive modulation and coding (AMC). The resulting MCSs $m(e), e \in \mathcal{E}$ and the corresponding link data rates $B^{m(e)}$ are then used in the second phase (Subsection 6.2) for calculating the achievable MMF solution for given routing paths (recall that in this paper the routing paths are fixed, see Subsection 5.1). The LBA approach assumes that the obtained MMF solution can be achieved in a real WMN by some suitable transmission scheduling protocol.

6.1. Phase 1: adaptive modulation and coding

Certainly, in problem SA/SI (and as matter of fact in SA/FI) assigning fixed MCSs $m(e), e \in \mathcal{E}$ to links should be optimized as it not only influences the link data rates $B_e = B^{m(e)}, e \in \mathcal{E}$ but also, through $\gamma^{m(e)}$, the admissible SINR ratios (3) (see Table 1). When active, a link which uses an MCS with a high data rate admits less concurrent transmissions than a link which uses an MCS with a lower data rate [25]. In effect, providing links that communicate at the highest possible data rate (although commonly assumed in the literature, see for example [30, 29]) may not be advantageous since an MCS with a smaller data rate used for link e would decrease the link rate B_e, but at the same time would increase the spatial reuse. This tradeoff should definitely be considered in optimizations. If the use of some link e simultaneously with a link in its neighborhood becomes possible, this clearly increases the throughput of the end-to-end data flows. Thus, such a more conservative link rate assignment strategy (i.e., to possibly use MCS smaller than implied by $d_{vw} \leq d^m$, see Table 1) can be advantageous for increasing the MMF network throughput [43], and also for the performance of a random access WMN [44].

Contrary to exact optimization of the MCS assignment which is difficult (as it requires exact solving of SA/SI), fixing reasonable MCSs in advance leads to a significant simplification of subsequent optimization of the MMF allocation vector. Hence, if this assignment is appropriate we may expect that it will allow for near-optimal MMF allocation. Pre-selection of appropriate MCSs for the links as a subproblem of the optimization problem SA/SI was considered in a previous work [43, 44]. Below we briefly summarize this approach (called AMC) and assume the computed MCS for the second phase of LDA.

Consider a fixed link $e \in \mathcal{E}$ and suppose that the transmissions on all other links in the network are specified at each time instant t. Let $\Gamma_e(t)$ (see (3)) be the resulting SINR of link e. Then, an optimal MCS assignment mechanism would clearly select the transmission $m(e)$ at the highest data rate that still allows to decode the transmitted signals at any time instant t when link e is active (i.e., $\Gamma_e(t) \leq \gamma^{m(e)}$). Following [47], define the link rate function B (i.e., a function that maps the link SINR to the maximum link data rate available through one of $m \in \mathcal{M}$) as follows:

$$B(\Gamma_e) = \max_{m \in \mathcal{M}} \{B^m : \gamma^m \leq \Gamma_e\}. \tag{22}$$

If the link rates are to be statically assigned before a WMN becomes operational, such an AMC mechanism cannot be used since SINR is time varying ($\Gamma_e = \Gamma_e(t)$, see (3)) and not predictable in advance. In contrast, the SNR coefficient ($\Gamma'_e = \frac{P_{a(e)b(e)}}{N}$, see (2)) is a constant (depends on the constant path gain only). Therefore, we alternatively could use the link SNR for assigning data rates (MCS) to links through the equation.

$$B_e = B(\Gamma'_e). \tag{23}$$

Clearly, (23) results in link data rates which are larger or equal to those computed by (22).

For a more conservative link rate assignment, i.e., a more robust MCS choice which enables a receiver to successfully decode a transmission despite a certain amount of interference, we use the link SNR together with the so called *interference buffer* $\Delta\Gamma_e$. This quantity can be interpreted as a safety margin to account for a certain amount of interference which realistically decreases the link rate. An interference buffer for link $e \in \mathcal{E}$ guaranteeing the successful transmission on the link in the presence of interference would formally be given by

$$\Delta\Gamma_e = \max_{0 \leq t \leq T} \frac{\Gamma_e(t)}{\Gamma'_e}. \tag{24}$$

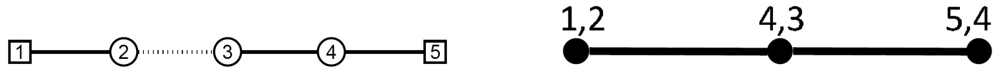
A conservative link rate assignment strategy using the interference buffer consists in replacing (23) by

$$B_e = \max_{m \in \mathcal{M}} \{B^m : \gamma^m \leq \frac{\Gamma'_e}{\Delta\Gamma_e}\}. \tag{25}$$

Using this function for link rate assignment would hence allow to transmit successfully in the presence of a number of interfering nodes (using any $\Delta\Gamma_e > 1$ would increase the likelihood of successful transmission with respect to (23)). Still, as $\Delta\Gamma_e$ depends on the time-varying interference, it is not known in advance and the values of $\Delta\Gamma_e, e \in \mathcal{E}$ (equivalently, assignment of MCSs to links) should in fact be optimized (with the MIP techniques applied to models of Section 5). However, a simplified assignment of MCS to links achieved through specifying a suitable fixed $\Delta\Gamma_e > 1, e \in \mathcal{E}$ would greatly simplify the subsequent optimization process, in effect transforming a multi-MCS problem SA/SI to its single MCS counterpart (the only difference would be that the fixed link rate values B_e would in general be different for different links). Our previous results [43, 44] indicate the usefulness of using a small interference buffer for the link rate assignment for all links, and hence this is the solution we use in LBA.

6.2. Phase 2: max-min fair throughput computation

In this subsection we present two algorithms of the LBA method for computing a MMF flow solution in the case when a selected single MCS $m(e)$ is already fixed to each link $e \in \mathcal{E}$ (which means that also the data rate of each link link is fixed to $B_e = B^{m(e)}$). This assignment is done heuristically by means of the simple AMC method described in the previous subsection. The two algorithms are referred to as nominal load based algorithm (NLBA), and effective load based



(a) Network topology

(b) Contention graph

Figure 1: Exemplary network topology and corresponding contention graph

algorithm (ELBA). Both NLBA and ELBA assume the simplified interference model described in Section 4.5.

For the load computation, NLBA uses the notion of a collision domain [13, 43]. A *collision domain* $\mathcal{F}_e$ of link $e \in \mathcal{E}$ consists of the link e itself and all links $g \in \mathcal{E}$ such that transmission on g disturbs transmission on e or vice versa. Formally:

$$\mathcal{F}_e = \{e\} \cup \{g \in \mathcal{E} \setminus \{e\} : \ \frac{P_{a(e)b(e)}}{N + P_{a(g)b(e)}} < \gamma^{m(e)} \ \vee \ \frac{P_{a(g)b(g)}}{N + P_{a(e)b(g)}} < \gamma^{m(g)}\}. \tag{26}$$

For the same purpose, ELBA uses the notion of a *maximal clique* in the contention graph $\mathcal{G}^c = (\mathcal{E}, \mathcal{K})$. The vertices of $\mathcal{G}^c$ are the links $e \in \mathcal{E}$ between the mesh nodes (eventually after discarding the links not used in the routing paths). An edge $k \in \mathcal{K}$ between two links $e, g \in \mathcal{E}$ exists if, and only if, the two links are *contending*, i.e., cannot be used in parallel [27]. We define Ω as the set of all maximal cliques in $\mathcal{G}^c$. In each clique $\mathcal{C} \in \Omega$ one link may be used at any time instant.

As an example consider the collision domain $\mathcal{F}_{43}$ of link $e = 43$ in the example shown in Figure 1a. For the sake of simplicity, we consider only the links which are used for routing purposes and discard the rest. Consequently, assuming that we are to establish the flow from gateway 1 to node 2, and the flows from gateway 5 to nodes 3 and 4, we consider only links 12, 54, and 43 (depicted by solid lines), and discard links 23, 32, 21, 34 and 45 as they are not used for routing purposes. Assuming a realistic interference scenario, the collision domain of link 43 is $\mathcal{F}_{43} = \{43, 54, 12\}$.

The contention graph for the topology of Figure 1a is shown in Figure 1b. The corresponding set of maximal cliques is given by $\Omega = \{\{12, 43\}, \{43, 54\}\}$.

The main (and virtually only) difference between the two algorithms is that NLBA uses the notion of the nominal load related to collision domains: $L^{nom}_{\mathcal{F}(e)}, e \in \mathcal{E}$, while ELBA – the notion of effective load related to maximal cliques: $L^{eff}_{\mathcal{C}}, \mathcal{C} \in \Omega$. These entities are defined as follows:

$$L^{nom}_{\mathcal{F}(e)} = \frac{T}{\sum_{e \in \mathcal{F}(e)} \frac{n_e}{B_e}} \qquad e \in \mathcal{E} \tag{27a}$$

$$L^{eff}_{\mathcal{C}} = \frac{T}{\sum_{e \in \mathcal{C}} \frac{n_e}{B_e}} \qquad \mathcal{C} \in \Omega. \tag{27b}$$

In (27), n_e denotes the number of routes traversing link e, i.e., $|\mathcal{Q}_(e)|$. Note that when B_e is the same (and equal to B) for all links, then $L^{nom}_{\mathcal{F}_e} = \frac{TB}{\sum_{e \in \mathcal{F}_e} n_e}$ and $L^{eff}_{\mathcal{C}} = \frac{TB}{\sum_{e \in \mathcal{C}} n_e}$.

NLBA assumes that only one link in a collision domain can be active at a given time instant while ELBA assumes this for a maximal clique. In either case the maximal flow that can equally be assigned to the network routes (the smallest value in the MMF allocation vector) is given by

$$f^{nom} = \min_{e \in \mathcal{E}} L^{nom}_{\mathcal{F}_e}, \qquad f^{eff} = \min_{e \in \mathcal{C}} L^{eff}_{\mathcal{C}}. \tag{28}$$

We illustrate the difference again with the help of Figure 1a assuming that all links have the data rate equal to B. The minimum in the left hand side of (28) is achieved for $\mathcal{F}_{43}$ with $\sum_{e \in \mathcal{F}_{43}} n_e = 4$ (flow 3 induces transmissions on two links and flows 4 and 2 induce transmission on one link each). This results in the nominal load of $L^{nom}_{\mathcal{F}_{43}} = \frac{TB}{4}$ and hence in the first step of MMF we get $f^{nom}_2 = f^{nom}_3 = f^{nom}_4 = \frac{1}{4}TB$. Analogously the minimum of the right hand side of (28) is achieved for $\mathcal{C} = \{43, 54\}$ with $L^{eff}_{\mathcal{C}} = \frac{TB}{3}$ ($\sum_{e \in \mathcal{C}} n_e = 3$) and hence $f^{eff}_2 = f^{eff}_3 = f^{eff}_4 = \frac{1}{3}TB$.

Having computed the values in (28) we assign the resulting flows to all routes in the network. We fix the corresponding values in the MMF vector for all the routes traversing the selected bottleneck collision domain (NLBA) or clique (ELBA). Then we delete these routes from the network and reduce the link data rates accordingly ($B_e := B_e - f^{nom} n_e$ or $B_e := B_e - f^{eff} n_e$). After that we repeat the procedure to increase the flows on the remaining routes to get the next value in the MMF vector, and so on. This procedure is a variant of the so called water-filling algorithm (see [16, 35, 39]).

In general, the nominal load (computed for a collision domain of link e) is smaller than or equal to the effective load (of a maximal clique containing the link), as the latter accounts for the possibility that two links in a collision domain can be used in parallel. In the example topology of Figure 1a, transmissions on links 12 and 54 cannot take place simultaneously with transmission on link 43, but they can take place in parallel to each other. The transmission from 4 to 3 could thus be scheduled for $\frac{1}{3}$ of the time. For $\frac{2}{3}$ of the time, transmissions on 12 and on 54 can take place in parallel. Transmission on link 54 is split equally to the flow from 5 to 4 and to the flow from 5 to 3. In effect, the final MMF allocation computed with ELBA is $f^{eff}_3 = f^{eff}_4 = \frac{1}{3}B$ and $f^{eff}_2 = \frac{2}{3}B$ while with NLBA it does not change after the first step.

An algorithm based on the effective LBA concept was proposed in [13]. However, it is correct (i.e., gives an optimal MMF vector for a fixed MCS allocation) only for the case of a tree-type mesh network with just one Internet gateway, as [13] does not show how to correctly compute the effective load for a general mesh network. Such a correct computation is given in [45] (based on ideas from [27]) resulting in the method ELBA described above.

To conclude, for fixed $m(e), e \in \mathcal{M}$, NLBA is a sub-optimal heuristic method for computing optimal MMF flow allocations while ELBA, as discussed in [45] is exact. This in particular means that ELBA is an exact method for SA/SI when $|\mathcal{M}| = 1$. Still, in general computing maximal cliques is an $\mathcal{NP}$-complete problem what may become an issue for large networks. As already mentioned, both NLBA and ELBA assume an idealistic channel access scheme guaranteeing a specific scheduling of channel access to each node that would realize the calculated MMF flow.

7. Numerical study

In this section we present a numerical study illustrating computational efficiency of the optimization models discussed in the previous sections and traffic efficiency (in terms of the resulting MMF vector of flows allocated to demands) of various variants of the WMN systems.

The reported results were obtained from the LP and MIP models implemented using Visual C++ under Windows XP, and executed on a single core Intel 2.4 GHz CPU with 3.92 GB RAM using the CPLEX 12.1 LP/MIP solver [28]. The heuristics as well as the algorithms NLBA and ELBA were implemented in MATLAB under Windows XP and executed on a dual core Intel 3.16 GHz CPU with 3.72 GB RAM.

In the numerical results we concentrate on four main problems, namely DA/FI, DA/SI, SA/FI, and SA/SI. We solve them using two approaches:

TS – time slot approach that considers time as a discrete resource and takes into account a notion of time slots; the approach is based on the MIP models of Section 4.

CS – compatible-set approach that considers time as a continuous resource and does not take into account a notion of time slots; the approach, based on column generation, is described in Subsection 5.3.

We compare running times of the algorithms and the resulting MMF vectors. Moreover, we investigate the impact of the number of possible MCSs on the results, and (in case of TS) the impact of the assumed number of time slots. We also compare the optimal MMF vector with the results provided by the LBA heuristics. Finally, we show how introducing NS (node capacity split) influences the optimal MMF vector and the running time of TS.

7.1. Example networks

Different variants of the MMF optimization problems (see Subsection 5.2) were solved for a number of randomly generated example topologies shown in Figure 2. Each of those topologies was generated using a grid with length d_x of $n_x \times n_y$ points where each of the grid points is chosen to be a mesh router or a mesh gateway with probability p_r and p_g respectively. This methodology, based on a binomial point process on a grid, guarantees the topologies that are likely to occur in reality (it is common that there is a certain minimal distance between the mesh nodes). As the topological characteristics may strongly influence the performance of WMN and of our algorithms, we generate sparse and dense topologies in a square with length 750 and 1050 m, respectively. For the sparse topologies shown in Figures 2a-2d, the parameters $d_x = 25$ m, $n_x = n_y = 30$ m and the probabilities $p_r = 0.02$ and $p_g = 0.003$ are used. The dense topologies depicted in Figures 2e-2f, are created with $d_x = 30$ m, $n_x = n_y = 30$ m and the probabilities $p_r = 0.04$ and $p_g = 0.006$.

In Figure 2, mesh routers are represented by circles and mesh gateways by squares. The routing topology is created by a routing protocol abstracts a sophisticated mesh routing protocol that was used in [43]. Paths rooted in the gateways are established by iteratively connecting the neighbor which is reachable with the highest link rate. Routing links are depicted by solid lines, all other links $e \in \mathcal{E}$ are depicted by dotted lines. For LBA, MCS assignment (shown in the figure) is done by means of AMC discussed in Subsection 6.1 using $\Delta\Gamma = 1$, i.e., each link e uses the highest feasible link rate given the SNR between its end points. The resulting link rates are represented by colors which are summarized by the legend shown in Figure 2a.

7.2. Computational efficiency

Below we discuss the running times obtained for different variants of the MMF problems. In Table 2, running times of algorithms maximizing the minimum flow in the presented models are shown. The first three columns of the table describe tested networks and contain: network name, number of nodes, and number of links. Remaining eight columns contain running times of the appropriate algorithms solving the four considered problems (DA/FI, DA/SI, SA/FI, and SA/SI), all approached with both TS and CS. Note that some results are marked with an asterisk. Those are the cases that were not solved within a time limit of three hours.

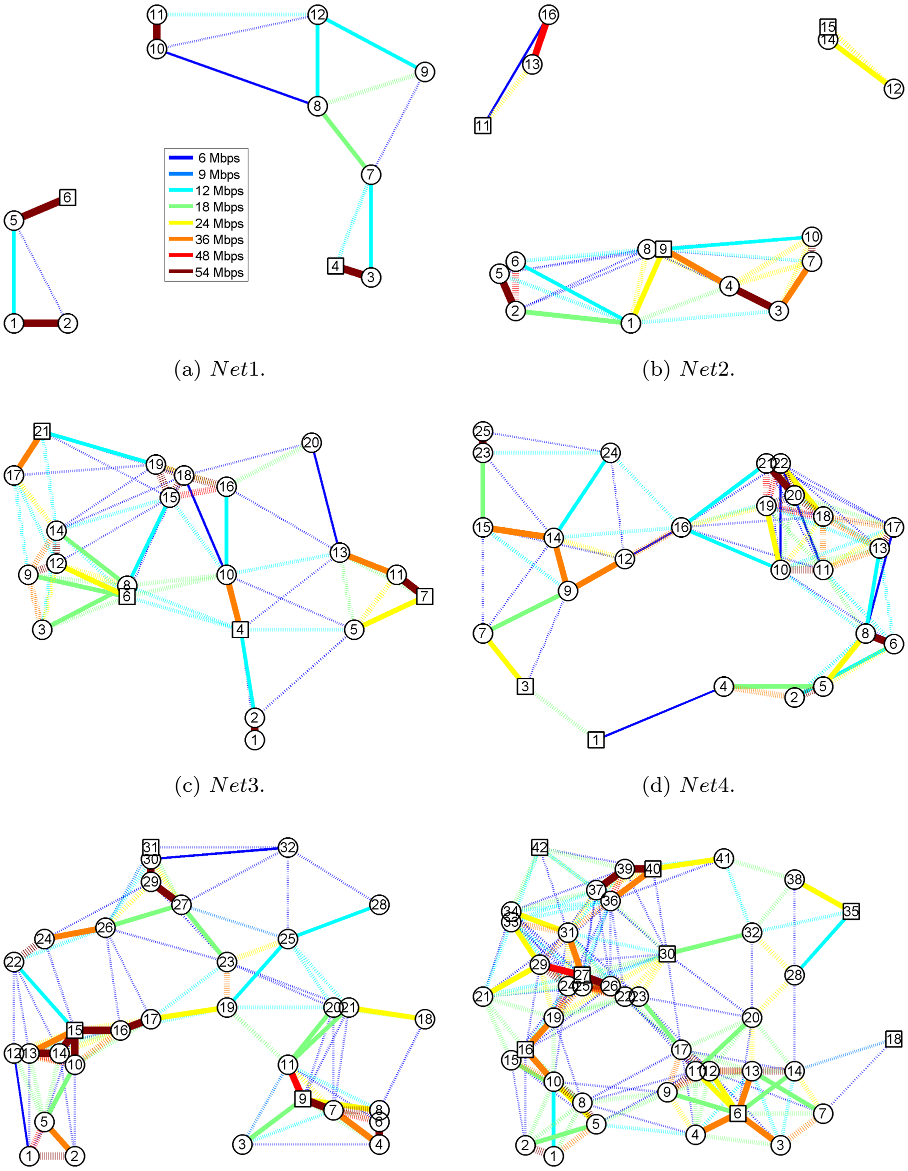
In the considered cases, all eight MCSs were available, and the number of time slots $|\mathcal{T}|$ for TS was set to ten.

The obtained results confirms the difficulty of the FI cases, thus justifying introducing the simplified interference model (SI). They also show that CS is in general superior to TS in terms of the running time. An interesting observation regarding this fact is that the SA cases seem to be much more difficult than the DA cases for CS. The reason is that they require a branch-and-price algorithm, while for TS a branch-and-cut algorithm is sufficient (note that we were able to use a commercial B&C solver).

The time required for computing the MMF solution for the LBA heuristics for all the considered topologies was in the worst case equal to 90 seconds, and in most cases much smaller. Thus the running time for the heuristics is not explicitly given in the following.

7.3. Traffic efficiency

In Table 3, the minimum flow for each case considered in Table 2 is presented. Note that for SA/SI solved using CS the results are presented in the following way: a (b), where a is a result returned by the exact method, while b is the best result provided by the heuristic methods of Section 6. As before, the cases that were not solved within a 3 hour time limit are marked with an asterisk.



(a) *Net*1.

(b) *Net*2.

(c) *Net*3.

(d) *Net*4.

(e) *Dense*1.

(f) *Dense*2.

Figure 2: Tested networks.

Table 2: Running times

			TS				CS			
network			DA		SA		DA		SA	
name	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	FI	SI	FI	SI	FI	SI	FI	SI
$Net1$	12	10	3m50s	<1s	3s	7s	<1s	<1s	7s	<1s
$Net2$	16	13	3h*	58s	1h13m	17s	<1s	<1s	2m48s	1m12s
$Net3$	21	17	3h*	5s	3h*	10s	7s	6s	19m59s	7m20s
$Net4$	25	23	3h*	3s	3h*	30s	13s	11s	3h*	3h*
$Dense1$	32	29	3h*	11m28s	3h*	39m33s	4m41s	35s	3h*	3h*
$Dense2$	42	34	3h*	3h*	3h*	48m53s	23m1s	56s	3h*	3h*

Table 3: Maximal minimum flow

			TS				CS			
network			DA		SA		DA		SA	
name	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	FI	SI	FI	SI	FI	SI	FI	SI
$Net1$	12	10	0.6	0.6	0.6	0.6	0.75	0.76	0.73	0.73 (0.66)
$Net2$	16	13	1.2*	1.2	1.2	1.2	1.62	1.65	1.5	1.5 (1.2)
$Net3$	21	17	1.2*	1.2	0.9*	1.2	1.32	1.44	1.26	1.41 (1.19)
$Net4$	25	23	0.0*	0.15	0.15*	0.15	0.22	0.23	0.21*	0.22* (0.22)
$Dense1$	32	29	0.0*	0.6	0.0*	0.6	0.92	0.99	0.72*	0.74* (0.80)
$Dense2$	42	34	0.0*	1.2*	0.0*	1.2	1.57	1.70	1.25*	1.26* (1.44)

Table 4: Impact of the number of available MCSs

# MCSs	1	2	3	4	5	6	7	8
time	1s	18s	14s	26s	28s	1m3s	2m4s	3m50s
result	0.3	0.45	0.6	0.6	0.6	0.6	0.6	0.6

As expected, the results obtained by CS are better than the results obtained by TS. This fact is self-explanatory. Another fact that might have been considered obvious is that results for SI should have been better than those for FI. It is usually the case. However, when TS was taken into account, simplifying the interference model was sometimes the only way to obtain a non-zero feasible result. Thus, due to the inability to solve big problem instances for FI models by MIP solvers, results provided for SI models sound like a good alternative. Note that results found by the heuristic method are better than the result found by the exact methods for $Dense1$ and $Dense2$. The reason is that the exact algorithm timed out, thus the solution it returned is not optimal.

In Table 4, the impact of the number of available MCSs on the minimum flow and the running time is shown. The impact has been measured for $Net1$ network, DA/FI problem, solved using TS with ten time slots. The MCSs were taken from Table 1 in the order presented there, i.e., if one MCS is available then it is the first one from the table, if two are available then they are the first two ones from the table, and so on.

As expected both running time of the algorithm and the minimum flow increase with the number of available MCSs. An interesting observation is that while the increase in time grows with the number of MCSs, the minimum flow rises significantly only in the beginning and seems to converge when bigger numbers of MCSs are available. The reason is that MCSs with higher required SINR can be used rarely in a network, thus their impact on the final result is limited. Note that in many cases an MCS with a high SINR cannot be used at all not because of the interferences but solely because of the noise. This fact highlights the importance of the choice of the number of available MCSs.

In Table 5, the impact of the number of time slots $|\mathcal{T}|$ on the results and the running time is presented. The conditions are like in the previous case, i.e., network $Net1$, DA/FI problem, solved using TS. Note that in this case the number of available time slots is changing from 3 to 12. The number of available MCSs is constant and equal to 8. In the first row of the table the number of available time slots is presented. The next row contains running times of the algorithm solving the considered instances of the problem, while in the last row the minimum flows obtained while solving the problems are shown.

Note that increasing the number of time slots does not always result in the increase of the

Table 5: Impact of the number of available time slots

# time slots	3	4	5	6	7	8	9	10	11	12
time	1s	1s	6s	8s	43s	40s	4s	3m50s	4m23s	5m22s
result	0.0	0.25	0.4	0.5	0.51	0.6	0.67	0.6	0.55	0.6

Table 6: Results for algorithms providing the full MMF vector

MMF step	1	2	3	4	5	6	7	8	9	10	11	12	13
time	72s	55s	55s	59s	59s	65s	71s	35s	10s	3s	3s	1s	1s
optimal	1.5	1.5	1.5	1.5	1.5	1.5	1.5	2.4	2.4	2.4	2.7	7.94	7.94
ELBA	1.2	1.2	1.2	1.2	1.2	1.2	1.2	2.58	2.58	2.58	2.58	8.49	8.49
NLBA	0.9	0.9	0.9	0.9	0.9	0.9	2.33	3.27	3.27	3.27	3.27	9.47	9.47

minimum flow. The reason is that the increased number of time slots decrease τ (assuming constant T) and this may limit the amount of data transmitted in crucial time slots. Obviously, adding the n^{th} time slot cannot deteriorate the result by more than $1/n$.

In Table 6, the time needed to solve each MMF step to optimality is presented together with the obtained optimal result and results returned by heuristic methods presented in Section 6. In this case $Net2$ network was used, and SA/SI case was solved using CS. In order to solve the full MMF problem, the algorithm of Section 5 was employed.

In the first row of the table the number of the considered step of the algorithm, i.e., d in Algorithm 1, is shown. The next row presents the time needed to solve the step of the first row to optimality. The following three rows contain results returned by the three considered algorithms, i.e., exact algorithm of Section 5.2, and two heuristic approaches (ELBA, NLBA) of Section 6. The results should be read as follows. The first column contains the minimum flow, the second column contains the second minimum flow, etc. The results show that the heuristics do not provide optimal solutions. Thus, as far as it is possible with respect to the running time, the exact approach should be used in order to solve the problem. Note that the time needed to solved the consecutive steps of Algorithm 1 decreases. The reason is that the feasible region is continuously diminished by adding constraints (16d), (16e), and a constraint resulting from (16a) after each step of the algorithm.

In Table 7, the impact of NS on the results and the time is presented. In this case all the networks from Table 2 were used, and DA/SI/NS case was solved using TS with ten available time slots. Note that the NS case cannot be solved with the form of CS described in Subsection 5.3.

It is clearly visible in the table that NS adds difficulty to the problem. The algorithm solving the problem reached a 3-hour time limit in three test cases out of six for the case with NS, while

Table 7: Impact of the node split

	without NS		with NS	
name	time	result	time	result
$Net1$	<1s	0.6	38s	0.7
$Net2$	58s	1.2	1h6m	1.35
$Net3$	5s	1.2	3h*	1.2
$Net4$	3s	0.15	12m50s	0.15
$Dense1$	11m28s	0.6	3h*	0.6
$Dense2$	3h*	1.2	3h*	1.2

it reached the time limit only once for the case without NS. On the other hand, in two test cases the result provided for NS were significantly better ($> 10\%$) that the corresponding results for the case without NS.

7.4. Practical hints on problem solving

The difficulty of the considered problems lays in the presence of a significant number of binary variables in the formulations. Usually it is possible to reduce their number, as the integrity of some of them is implied by the integrity of others. Fortunately, in the considered cases the integrity of all binary variables is imposed by the integrity of link variables, i.e., Y_{et}, y_{et}^m, Y_e in case of the column generation, and y_e^m in SA case solved using CS. Thus, the integrity of all other variables, i.e., X_{vt}, x_{vt}^m, Z_v^m, and z_{evt}, can be relaxed.

The branch-and-price approach for solving (20) was implemented in C++. The branch-and-price tree was traversed using the depth-first search (DFS) algorithm. The order of links in DFS was with respect to the priority, which for link e was computed as:

$$\sum_{v \in \mathcal{V}} P_{va(e)} + P_{vb(e)}.$$

Each node of the tree, together with a procedure generating interesting compatible sets, was solved using CPLEX.

7.5. Summary of the numerical study

In this section a numerical study illustrating computational efficiency of the optimization models discussed in the previous sections and traffic efficiency (in terms of the resulting MMF vector of flows allocated to demands) of various variants of the WMN systems has been presented.

We have concentrated on four problems: DA/FI, DA/SI, SA/FI, and SA/SI. The problems have been solved using two approaches: TS and CS. Running times of the algorithms and the quality of the obtained MMF vectors have been compared, showing that CS outperforms TS in both running time and the quality of the solution. Also, the impact of the number of possible MCSs on the results, and (in case of TS) the impact of the number of available time slots has been discussed.

We also compared the optimal MMF vector to results provided by our heuristic approaches. The results clearly show that there is a space for improvements in this field, and the gap should be filled by more efficient heuristic methods that are able to optimize the link rate assignment and the time of usage of each link in some kind of an interlacing way.

Finally, we showed how introducing NS influences the optimal MMF vector and the running time of TS. Our results confirm that NS can be desirable from the point of view of the objective function, at least for a small number of time slots (we can expect that the gain from NS will gradually disappear with increasing $|\mathcal{T}|$). However, introducing NS significantly increases the running times.

8. Conclusion

The paper is devoted to WMN modeling using mixed-integer programming (MIP) formulations that allow to precisely characterize the link data rate capacity space and transmission scheduling within time slots. Such MIP models have been formulated for several cases of the modulation and coding schemes (MCS) assignment. On top of that we have presented a general way of solving the MMF traffic objective for WMN and use it for the formulated capacity models. This is a novel feature of the paper. We have discussed several ways of solving the considered MMF problems (compact MIP formulations based on time slots, non-compact LP/MIP formulations based on compatible sets, heuristic methods). We have made an extensive numerical study that illustrates the running time efficiency of the different solution approaches, and the influence of the MCS selection options (static, dynamic, node capacity split) and the number of time slots on the traffic performance of a WMN.

As for the running times, a general conclusion is that the exact approaches (especially those based on generation of compatible sets) are quite effective and can be applied even to network examples of realistic sizes (especially when combined with the simplified interference model). When the exact approach fails, the heuristics (for static MCS allocation) can be used since in practice they indicate short running times. Still, the heuristics tend to underestimate the MMF vector and need some improvement of the first phase (allocation of MCSs to links).

As for the traffic performance, dynamic MCS allocation (DA) seems to be only slightly better than static allocation (SA). Adding the node capacity split feature increases the traffic performance (for small number of time slots) but on the expense of increased running times. We have also investigated the influence of the number of MCS and the number of time slots on the MMF vector. It seems that limiting the number of MCS and time slots is acceptable. We obtained the same results for the case with three MCSs and the case with eight MCSs. What is more, the result for nine time slots was only 10% away from the theoretical maximum obtained for an infinite number of time slots (the case equivalent to CS).

Basically, the presented approach assumes (abstract) time slots and provides the transmission plan for each slot. In practice, the calculated schedule should somehow be realized in a network. One such solution would be to implement a centrally preplanned transmission scheduling by means of TDMA. This aspect, however, has not been studied in the paper – it needs a considerable effort in terms of further research, This in particular concerns our heuristics, as they do not even use the notion of a time slot. Another direction of further research is improvement of the MIP models formulated in Section 4 to make them applicable to larger network instances. This can be achieved by introducing additional cuts and other means of integer programming. Finally, improving the first phase of the heuristics is an important research issue.

Acknowledgement While preparing the paper, the cooperation of Di Yuan and Michał Pióro was supported by the eLLIIT project funded by the Swedish Foundation for Strategic Research. The cooperation of Barbara Staehle, Dirk Staehle, M. Pióro and Mateusz Żotkiewicz was supported by the FP7 Network of Excellence Euro NF. M. Żotkiewicz (and partly M. Pióro) was supported by the Polish Ministry of Science and Higher Education (grant no. N517 397334 "Optimization Models for NGI Core Networks"). M. Pióro was also partly supported by the Swedish Research Council (grant no. 621-2006-5509 "Modeling and Design of Core Internet Network"). The work of Barbara Staehle and Dirk Staehle was funded by the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 01 BK 0800, GLab).

References

[1] *Bay area wireless users group*, http://www.bawug.com.

[2] *Belair networks*, http://www.belairnetworks.com.

[3] *Nortel*, http://www.nortel.com.

[4] *Seattle wireless*, http://www.seattlewireless.net.

[5] *Tropos networks*, http://www.tropos.com.

[6] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, *Network flows: Theory, algorithms, and applications*, Prentice Hall, 1993.

[7] I.F. Akyildiz and X. Wang, *A survey on wireless mesh networks*, IEEE Communications Magazine **43** (2005), no. 9, 23–30.

[8] I.F. Akyildiz, X. Wang, and W. Wang, *Wireless mesh networks: a survey*, Computer Networks **47** (2005), no. 4, 445–487.

[9] H.M.K. Alazemil, A. Das, R. Vijaykumar, and S. Roy, *Fixed channel assignment algorithm for multi-radio multi-channel mesh networks*, Wireless Communications and Mobile Computing **8** (2008), 811–818.

[10] M. Alicherry, R. Bhatia, and L.E. Li, *Multi-radio, multi-channel communication: Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks*, Proceedings ACM MobiCom (Cologne, Germany), August 2005.

[11] E. Amaldi, A. Capone, M. Cesana, and F. Malucelli, *Optimization models for the radio planning of wireless mesh network*, Lecture Notes in Computer Science (2007), no. 4479, 287–298.

[12] E. Ancillotti, R. Bruno, and M. Conti, *Design and performance evaluation of throughput-aware rate adaptation protocols for IEEE 802.11 wireless networks*, Elsevier Performance Evaluation **66** (2009), no. 10, 811–825.

[13] B. Aoun and R. Boutaba, *Max-min fair capacity of wireless mesh networks*, Proceedings of IEEE MASS (Vancouver, Canada), October 2006.

[14] B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, *Gateway placement optimization in wireless mesh networks with QoS constraints*, IEEE Journal in Selected Areas in Communications **24** (2006), no. 11, 2127–2136.

[15] W. Ben-Ameur and H. Kerivin, *New economical virtual private networks*, Communications of the ACM **46** (2003), no. 6, 69–73.

[16] D. Bertsekas and R. Gallager, *Data networks*, Prentice Hall, 1992.

[17] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, *Architecture and evaluation of an unplanned 802.11b mesh network*, Proceedings of ACM MobiCom (Cologne, Germany), August 2005.

[18] P. Björklund, P. Värbr, and D. Yuan, *A column generation method for spatial TDMA scheduling in ad hoc networks*, Ad Hoc Networks **2** (2004), no. 4, 405–418.

[19] R. Bruno, M. Conti, and E. Gregori, *Mesh networks: commodity multihop ad hoc networks*, IEEE Communications Magazine **43** (2005), no. 3, 123–131.

[20] J. Camp, J. Robinson, C. Steger, and E. Knightly, *Measurement driven deployment of a two-tier urban mesh access network*, Proceedings of ACM MobiSys (Uppsala, Sweden), June 2006.

[21] A. Capone, G. Carello, I. Filippini, S. Gualandi, and F. Malucelli, *Routing, scheduling and channel assignment in wmn networks: optimization, models and algorithms*, Ad Hoc Networks (2010), no. 8, 545–563.

[22] C. Chekuri, *Routing and network design with robustness to changing or uncertain traffic demands*, SIGACT News **38** (2007), no. 3, 106–129.

[23] A.K. Das, H.M.K. Alazemi, R. Vijayakumar, and S. Roy, *Optimization models fixed channel assignment in wireless mesh networks with multiple radios*, Proceedings of IEEE SECON (Santa Clara, CA, USA), September 2005.

[24] M. Genetzakis and V. Siris, *A contention-aware routing metric for multi-rate multi-radio mesh networks*, Proceedings of IEEE SECON (Boston, MA, USA), June 2008.

[25] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[26] G.R. Hiertz, S. Max, E. Weiss, L. Berelmann, D. Deteneer, and S. Mangold, *Mesh technology enabling ubiqitous wireless network*, Proceedings of WICON (Boston, MA, USA), August 2006.

[27] X.L. Huang and B. Bensaou, *On max-min fairness and scheduling in wireless ad-hoc networks: Analytical framework and implementation*, Proceedings of MobiHoc (Long Beach, CA, USA), October 2001.

[28] ILOG, *Cplex 12.1 user's manual*, ILOG, 2009, `ftp://ftp.software.ibm.com/software/websphere/ilog/docs/optimization/cplex/ps_usrmancplex.pdf`.

[29] T.S. Kim, H. Lim, and J.C. Hou, *Improving spatial reuse through tuning transmit power carrier sense threshold, and data rate in multihop wireless networks*, Proceedings of MobiCom (Los Angeles, CA, USA), September 2006.

[30] M. Lacage, M.H. Manshaei, and T. Turletti, *IEEE 802.11 rate adaptation: a practical approach*, Proceedings of MSWiM (Venice, Italy), October 2004.

[31] L. Lasdon, *Optimization theory for large systems*, MacMillan, 1970.

[32] M.J. Lee, J. Zheng, Y-B. Ko, and D.M. Shrestha, *Emerging standards for wireless mesh technology*, IEEE Wireless Communications **13** (2006), no. 2, 56–63.

[33] X. Lin and N.B. Shroff, *Joint rate control and scheduling in multihop wireless networks*, Proceedings of IEEE Conference on Decision and Control (Atlantis, Paradise Island, Bahamas), December 2004.

[34] J. Mo and J. Walrand, *Fair end-to-end window-based congestion control*, IEEE/ACM Transactions on Networking **8** (2000), 556–567.

[35] D. Nace and M. Pióro, *Max-min fairness and its applications to routing and load-balancing in communication networks – a tutorial*, IEEE Communications Surveys and Tutorials **10** (2008), no. 4, 5–17.

[36] M.J. Neely, E. Modiano, and C. Li, *Fairness and optimal stochastic control for heterogeneous networks*, Proceedings of IEEE INFOCOM (Miami, FL, USA), March 2005.

[37] P. Nilsson and M. Pióro, *Solving dimensioning tasks for proportionally fair networks carrying elastic traffic*, Performance Evaluation **49** (2002), 371–383.

[38] W. Ogryczak, M. Pióro, and A. Tomaszewski, *Telecommunications network design and max-min optimization problem*, Journal of Telecommunications and Information Technology (2005), no. 3, 1–9.

[39] M. Pióro and D. Medhi, *Routing, flow, and capacity design in communication and computer networks*, Morgan Kaufman, 2004.

[40] M. Pióro, B. Staehle, D. Staehle, and Di Yuan, *Optimization models for mesh networks involving max-min fairness*, Tech. report, Department of Electrical and Information Technology, Lund University, Sweden, `www.eit.lth.se/staff/michal.pioro/reports/meshed.pdf`, April 2010.

[41] A. Raniwala, K. Gopalan, and T.C. Chiueh, *Centralized channel assignment and routing algorithms for multichannel wireless mesh networks*, ACM Mobile Computing and Communications Review **8** (2004), no. 2, 50–65.

[42] S. Sen, R.D. Doverspike, and S. Cosares, *Network planning with random demand*, Telecommunication Systems **3** (1994), 11–30.

[43] B. Staehle, D. Staehle, and R. Pries, *Effects of link rate assignement on the max-min fair throughput of wireless mesh networks*, Proceedings of ITC (Paris, France), September 2009.

[44] ———, *Effects of link rate assignment in IEEE 802.11 mesh networks*, Proceedings of European Wireless Conference (Lucca, Italy), April 2010.

[45] D. Staehle, B. Staehle, and R. Pries, *Max-min fair throughput in multi-gateway multi-rate mesh networks*, Proceedings of IEEE VTC (Taipei, Taiwan), May 2010.

[46] L. Tassiulas and A. Ephremides, *Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks*, IEEE Transactions on Automatic Control **37** (1992), no. 12, 1936–1948.

[47] S. Toumpis and A Goldsmith, *Capacity regions for wireless ad hoc networks*, IEEE Transactions on Wireless Communications **2** (2003), no. 4, 736–748.

[48] R. Vannier and I. Lassous, *Towards a practical and fair rate allocation for multihop wireless networks based on simple node model*, Proceedings of MSWiM (Vancouver, British Columbia, Canada), October 2008.

[49] P-J. Wan, *Multiflows in multihop wireless networks*, Proceedings of MobiHoc (New Orleans, LA, USA), May 2009.

[50] X. Zhou and L. Hang (eds.), *Resource management and routing in wireless mesh network*, Computer Communications (2008), no. 31, 1255–1258.

Appendix A. Derivation of formulation (15)

Consider a fixed, given (criteria) vector $f = (f_1, f_2, \ldots, f_D) \in \mathbb{R}^m$. Let $[f] = ([f]_1, [f]_2, \ldots, [f]_D)$ denote the version of f sorted in the non-decreasing order and let the cumulated optimization criteria vector $F = (f_1, f_2, \ldots, f_D)$ be defined by $F_d = [f]_1 + [f]_2 + \ldots + [f]_d,\ d = 1, 2, \ldots, D$. We will show that each such cumulated criterion F_d (for a fixed d, $1 \leq d \leq D$) can be expressed as the solution of a linear programming problem in continuous variables b and $g_j, j = 1, 2, \ldots, D$ (see Section 4.2 of [38] and the references therein):

$$F_d = \max \{db - \sum_{j=1}^{D} g_j\} \tag{A.1a}$$

$$b \leq f_j + g_j \qquad j = 1, 2, \ldots, D \tag{A.1b}$$

$$g_j \geq 0 \qquad j = 1, 2, \ldots, D. \tag{A.1c}$$

The fact that the optimal objective (A.1a) does indeed express the sum of the d smallest elements of f can be verified intuitively. The formal proof is as follows.

Consider the following binary minimization problem:

$$z = \min \sum_{j=1}^{D} f_j u_j \tag{A.2a}$$

$$\sum_{j=1}^{D} u_j = d \tag{A.2b}$$

$$u_j \in \{0, 1\} \qquad j = 1, 2, \ldots, D. \tag{A.2c}$$

The minimal objective z^0 defined in (A.2a) is equal to F_d because according to (A.2b) each feasible vector of variables u contains exactly d entries u_j equal to 1, and in the optimal solution u^*, the entries $u_j^0 = 1$ will correspond to the first d smallest entries in vector f.

Moreover, we can relax the binarity of the variables and solve the corresponding linear program in continuous u with constraint $0 \leq u_j \leq 1,\ j = 1, 2, \ldots, D$ instead of (A.2c):

$$\min \sum_{j=1}^{D} f_j u_j \tag{A.3a}$$

$$[\,b\,] \qquad \sum_{j=1}^{D} u_j = d \tag{A.3b}$$

$$[\,g_j \geq 0\,] \qquad u_j \leq 1 \qquad j = 1, 2, \ldots, D \tag{A.3c}$$

$$u_j \geq 0 \qquad j = 1, 2, \ldots, D. \tag{A.3d}$$

The quantities in brackets are the dual variables corresponding to the constraints. It is now a straightforward exercise to see that the problem dual to (A.3) is exactly the problem formulated in (A.1).

Appendix B. MMF algorithm for the CS formulations

Algorithm 2 (Algorithm for problem (17))

Input: A procedure for generating compatible sets.
Output: MMF solution $f^0 = (f_1^0, f_2^0, \ldots, f_D^0)$ and corresponding optimal capacities c^0.

Step 0: Set $\mathcal{B} = \emptyset$ (the set of blocking demands).

Step 1: Solve (through CS generation) the linear program:

$$\max \quad F \tag{B.1a}$$

$$F \leq f_d \qquad d \in \mathcal{D} \setminus \mathcal{B} \tag{B.1b}$$

$$f_d = f_d^0 \qquad d \in \mathcal{B} \tag{B.1c}$$

$$\sum_{i \in \mathcal{I}} z_i = T \tag{B.1d}$$

$$c_e = \sum_{i \in \mathcal{I}} B_{ei} z_i \qquad e \in \mathcal{E} \tag{B.1e}$$

$$\sum_{d \in \mathcal{Q}_e} f_d \leq c_e \qquad e \in \mathcal{E} \tag{B.1f}$$

$$z \geq 0 \tag{B.1g}$$

(where $\mathcal{D} = \{1, 2, \ldots, D\}$).

Step 2: Denote the resulting optimal objective by F^0 and the optimal values of dual variables corresponding to (B.1b) by $\lambda_d^0,\ d \in \mathcal{D} \setminus \mathcal{B}$.

Put $f_d^0 = F^0,\ d \in \{d \in \mathcal{D} \setminus \mathcal{B} :\ \lambda_d^0 > 0\}$ and $\mathcal{B} := \mathcal{B} \cup \{d \in \mathcal{D} \setminus \mathcal{B} :\ \lambda_d^0 > 0\}$.

Step 3: If $\mathcal{B} \subsetneq \mathcal{D}$ come back to Step 1. Otherwise, stop: $f^0 = (f_1^0, f_d^0, \ldots, f_D^0)$ and the last optimized capacities c^0 compose the optimal solution.

The correctness of Algorithm 2 follows from the convexity of problem (B.1), non-negativity of λ_d ($d \in \mathcal{D} \setminus \mathcal{B}$), the dual constraint $\sum_{d \in \mathcal{D} \setminus \mathcal{B}} \lambda_d = 1$, and the complementary slackness property. For details the reader is referred to Chapter 8 in [39] (see also [35, 38]).