Opportunistic Routing with Adaptive CSMA/CA in Wireless Mesh Networks

System Architecture Group
Technical Report No. SAR-PR-2010-04

Mathias Kurth
Humboldt University Berlin, Germany
kurth@informatik.hu-berlin.de

May 3, 2010

In this paper, we address the question of how opportunistic routing (OR) should be used in wireless mesh networks (WMNs) based on carrier sense multiple access (CSMA) in order to efficiently utilize multi-user and spatial diversity. OR mitigates the channel impairments using multiple candidate receivers, which comes at the expense of reduced spatial reuse. Based on a cross-layer optimization framework, we design a cross-layer protocol for congestion control, opportunistic routing and CSMA scheduling for WMNs that is able to handle the described tradeoff. It is completely distributed and relies on local and neighboring information only.

We present a Markov model of CSMA/CA with node-oriented carrier sensing (CS). We show that CSMA with node-based CS can be operated in a simple and distributed way if the receiver blocking problem is handled properly. In addition, we illustrate how a CSMA/CA protocol for WMNs can be designed according to the proposed model based on hierarchical busy tones.

In the derivation of the protocol, several practical questions have to be addressed. In order to prevent the breakdown of CSMA due to collisions, the contention aggressiveness has to be adapted. We show how an efficient working point in terms of throughput and collisions can be approached dynamically via feedback loops. Furthermore, we illustrate that virtual packets can be used to control the end-to-end delay and reduce the convergence time while their costs in terms of throughput are generally lower compared to state-of-the-art approaches. In addition, we show how the back-pressure routing can be heuristically guided via a WMN

routing metric in order to trade off the incurring costs of additional relays in CSMA scheduling.

Analytic results suggest that neither spatial nor multi-user diversity dominates the other. Thus, both forms of diversity should be exploited systematically and dynamically. We illustrate how the cross-layer protocol handles the tradeoff between spatial and multi-user diversity. In particular, it is crucial that a transmitter does not decide for a single anycast link in advance. Instead, all anycast links should contend for medium access, and the topology "decides" on the links to activate according to the CS relationship. Considering the simulation results, we conclude that the proposed protocol is able to increase the throughput and fairness performance of WMNs in relation to state-of-the-art protocols. However, a fundamental tradeoff arises between throughput and delay on the one hand and the degree of multi-path and opportunism on the other. The optimality gap to the theoretical potentials increases with the utilized degree of multi-path and opportunism. The more relays are used, the higher are the expected benefits of spatial and multi-user diversity, but they are generally diminishing and will eventually be eaten up by the costs of every additional relay in terms of CSMA efficiency.

Contents

1. Introduction **5**

2. Related Work **8**
2.1. Network Utility Maximization 8
2.2. Utility-Optimal CSMA 10
2.3. Opportunistic Routing 15

3. A Node-Oriented Model of CSMA in WMNs **17**
3.1. CSMA/CA as a Markov Random Field 18
3.2. Operating CSMA/CA in WMNs 21

4. CSMA/CA with Hierarchical Busy Tones **22**
4.1. Link Level Evaluation 25
4.1.1. Synchronization Rate and Probing Overhead 25
4.1.2. Collisions in CSMA/HBT 27
4.1.3. Mutual Exclusion under Physical Interference 29

5. Optimization Framework for Opportunistic Routing with Adaptive CSMA/CA **34**
5.1. Optimization Problem 36
5.2. Congestion Control 39
5.3. Opportunistic Routing 39
5.4. CSMA Scheduling 42

6. Design of a Cross-Layer Protocol **42**
6.1. Packet Transition and Forwarder Selection 43
6.2. Efficiency-Collisions Tradeoff in Multi-Hop Routing 44
6.2.1. Achieving High Efficiency with Limited TAs 45
6.2.2. Enforcing Proportional Fairness 47
6.3. Convergence Period and Queueing Delays 50
6.3.1. Providing Service during Convergence 52
6.3.2. Back-Pressure with Reduced End-To-End Delays 54
6.4. Compensation for Candidate Set Dynamics 60
6.5. Route Pre-Selection via Routing Metrics 63

7. Evaluation in Illustrative Scenarios **64**
7.1. Efficiency Adaptation with Limited TAs 66
7.2. Fairness with Limited TAs 67
7.3. Step Size Adaption during and after Convergence 69
7.4. End-To-End Delay and Short Time Fairness 71
7.5. Candidate Set Dynamics 76
7.6. Multi-User Diversity 78
7.7. Spatial Diversity 81

8. System Level Evaluation **84**

8.1. Potential of Multi-User and Spatial Diversity on Random Topologies . 85
8.2. A Single Flow on Random Topologies 89
8.3. Multiple Flows on Random Topologies 97

9. Conclusion and Outlook **101**
9.1. Theory-Practice Gaps, Limitations and Future Directions 103

A. Partial Balance in the Extended CSMA Markov Chain **108**

B. Statistical Entropy of the Extended CSMA Markov Chain **108**

C. Anycast Goodput Region With Polynomial Number of Constraints **110**

D. Simultaneous Active Constraints of the Anycast Goodput Region **111**

E. Opportunistic Back-Pressure Routing with Linear Queueing Complexity **113**

References **116**

Acronyms **127**

1. Introduction

Within the last few years, research and engineering in the field of the physical layer (PHY) for wireless communications has made tremendous progress. Techniques like multiple-input multiple-output (MIMO), spatial multiplexing, beam-forming, space-time coding (STC) and orthogonal frequency-division multiple access (OFDMA) have been developed, to mention only some of them. In most cases, the innovation has been driven by the demands of the cellular world. Furthermore, the design principles behind cellular systems enable the rapid assimilation of innovative PHY technologies and concepts, so that we already have all the above mentioned techniques readily available in today's cellular systems.

On the other hand, wireless mesh and ad-hoc networks cannot keep up with the speed of innovation at the PHY. The distinguishing feature of WMNs is the coupling of wireless link capacities at the PHY, so that local decisions of nodes are sub-optimal. In particular, the scheduling problem NP-hard [127]. If we assimilate advanced PHY techniques into WMNs via the transparent exchange of the PHY only, then it is questionable whether they are efficiently utilized in order to increase the overall network performance. It is rather necessary to understand and design WMNs from a holistic cross-layer perspective.

The Internet protocol stack as well as the ISO/OSI reference model has a layered architecture. Layering is a principle of software engineering that tries to reduce the complexity of a system through abstraction and modularization [13], and it has been applied to the Internet stack and the OSI model for that reason with great success [14, 137]. Cross-layer design based on network utility maximization (NUM) does *not* intend to dismiss the layered architecture of network stacks. In contrast, it provides a formal justification for layering that is essential to systematically understand and design WMNs. Using the words of Yi *et al.* [127], it "becomes a foundation to understand, in a conceptually simple way, the complexities of network architectures: 'who does what' and 'how to connect them' ."

We will illustrate the need for a cross-layer design using two examples. The TCP resides at the transport layer. It provides a byte-stream oriented, reliable and ordered end-to-end service, and it is responsible for congestion control amongst others. The rationale of the IEEE 802 family of standards is to define a common view on the medium access control (MAC) and PHY, so that a basic inter-connection between different types of local and metropolitan area networks is possible [44]. In particular, the network layer is provided with a common abstraction of the lower layers, which allows the operation of TCP in Ethernet local area networks (LANs) as well as in multi-hop WMNs. However, TCP has been designed for wire-line networks. The exchange of the lower layers from a black-box perspective results in an underutilization of the WMN resources since the error characteristic of the wireless channel is different from the design assumptions [27]. Furthermore, optimization approaches that operate on the lower layers only may have the adverse effects on the throughput of TCP,

as it was shown for the adaption of the modulation and coding scheme (i.e. bit-rate) at intermediate links [111].

As second example, we will consider opportunistic routing (OR) in WMNs, which is also the main topic of this paper. OR is one of the above mentioned advanced concepts. Its intention is to benefit from multi-user diversity (MUD) in WMNs. In contrast to wire-line networks, bit and frame errors are common in WMNs [1]. OR turns this drawback into an opportunity. Due to the broadcast nature of the wireless medium, the risk of transmission errors can be significantly reduced through the usage of additional candidate receivers. The transmitter does not have to specify the relay for the next hop a priori. Instead, the routing decision is made a posteriori relying on the information which candidate actually received the frame. However, every additional candidate causes costs in terms of spatial resources. Thus, the *tradeoff between multi-user gain and spatial reuse* is central to OR. A small number of candidates offers higher spatial reuse, whereas transmission errors become more likely, and vice versa. Thus, for the system the question arises: What type of diversity should be used, and to which extent?

In WMNs, the design of efficient MAC protocols is challenging due to the complexity of the underlying scheduling problem. For distributed solutions, the complexity translates to excessive message passing that is prohibitively resource consuming. Thus, we might ask how much throughput efficiency we have to sacrifice for distributed schedulers that do not rely on message passing like carrier sense multiple access (CSMA). And in fact, the question has been answered recently and the answer is: There is no efficiency loss, i.e. CSMA is throughput-optimal [48]. Furthermore, the result is constructive in a way that it can be embedded into an optimization framework in order to understand and design cross-layer CSMA protocols.

In this paper, we address the question of how OR should be used in WMNs based on CSMA in order to efficiently utilize multi-user and spatial diversity. Our objective is to design a cross-layer protocol and to evaluate its properties using a prototype for a detailed network level simulator. The primary performance metric is utility, i.e. throughput subject to fairness constraints. We are targeting at applications that support elastic traffic and do not impose (tight) end-to-end delay or jitter constraints like FTP. As a first step in understanding OR, we have chosen a WMN with stationary memoryless unreliable links as target environment, i.e. the frame error process is assumed to be independent and identically distributed (i.i.d.) in time. This assumption covers several well-known channel models like the additive white Gaussian noise (AWGN) channel, the frequency-selective wideband fading channel, which arises from delayed multi-path components at the receiver, and the fast fading channel, where the fading process evolves faster than the symbol duration [103]. Note that the slow fading channel and the Gilbert-Elliot model are not covered [37, 103]. In particular, the temporal uncorrelated channel is the best case for OR, since the transmitter cannot use any historical information to deduce the instantaneous channel conditions. The more reliable the channel can be predicted beforehand, the lower the benefits of OR become in relation to opportunistic scheduling [115]. We do not constrain our research focus to a particular technology like IEEE 802.11 or to the ISO/OSI layering in general. However, our solution will fit within the OSI model, as we will

see. The contributions of this paper are as follows.

- We present a Markov model for CSMA/CA, in which CS is processed *per node* (cf. section 3). In contrast to state-of-the-art approaches that assume a link-based CS [47], our model captures the capabilities of available CS hardware more closely [62] and allows for a better spatial reuse. In addition, we show that CSMA with node-based CS can be operated in a simple and distributed way if the receiver blocking problem is handled properly.

- We design a CSMA/CA protocol for WMNs according to the model above based on the idea of dual busy tone multiple access (DBTMA) [42] (cf. section 4). The throughput-optimality of CSMA relies on the prerequisite that the contention among WMN nodes is fair in a sense that every transmitter can increase its service rate at each time through a more aggressive contention. We show how this prerequisite can be ensured within an IEEE 802.11-like protocol for WMNs without excessively sacrificing spatial resources.

- We present a cross-layer algorithm for congestion control, opportunistic routing and CSMA scheduling for WMNs (cf. section 5). The algorithm is completely distributed and relies on local and neighboring information only. It furthermore handles transmission errors and selects bit-rates on anycast links.

- Based on the algorithm, we propose a cross-layer protocol that can readily be implemented in a wireless network simulator (cf. section 6). In particular, this involves the following three issues.
  We define a working point for CSMA in terms of contention aggressiveness that ensures a high efficiency while avoiding the breakdown of CSMA due to collisions. The working point is approached via an intra-flow feedback loop, whereas an inter-flow feedback loop considers the fairness among competing flows.
  Furthermore, we present an approach to control the throughput-delay tradeoff in CSMA via virtual packets. This way, we are able to reduce the convergence time and control the end-to-end delay at a level of practical relevance, whereas its costs in terms of throughput are lower compared to state-of-the-art delay reduction approaches [47].
  In addition, we propose to pre-select the routing paths according to a WMN metric. For CSMA, every additional relay incurs costs in terms of throughput. Using the pre-selection, it is possible to pre-exclude relays that are supposed to have only a marginal contribution to the overall performance.

- We evaluate the proposed protocol through analysis and simulation (cf. section 7 and section 8). The analytic characterization of random topologies suggests that both multi-user and spatial diversity should be exploited systematically and dynamically, since the potential for either of them depends on the topology and the protocol parameters and no form of diversity dominates the other.

We illustrate how the tradeoff between spatial reuse and multi-user gain is handled within the simulator prototype. In particular, it is crucial that a transmitter does not decide for a single anycast link in advance. Instead, all anycast links should contend for medium access, and the topology "decides" on the links to activate according to the CS relationship.
Considering the simulation results, we conclude that the prototype is able to increase the throughput and fairness performance of WMNs in relation to state-of-the-art single-path and OR protocols. On the other hand, the optimality gap to the theoretical potentials increases with the utilized degree of multi-path and opportunism. A more robust MAC leads to fewer opportunities on the higher layers and thus smaller OR gains. Furthermore, a fundamental tradeoff arises between utility and delay on the one hand and the degree of multi-path and opportunism on the other. The more relays are used, the higher are the expected benefits of spatial and multi-user diversity, but they are generally diminishing and will eventually be eaten up by the costs of every additional relay in terms of CSMA efficiency.

2. Related Work

Our objective is to understand and control opportunistic routing (OR) in WMNs based on carrier sense multiple access (CSMA). The paper builds upon a broad foundation of related work. The most important categories of related work cover the concepts OR and CSMA, of course. However, we need a further means which allows us to understand the cross-layer interaction between both concepts and provides us hints how to design a practical cross-layer protocol. We have found such a means within network utility maximization (NUM). In the following, we will introduce NUM, CSMA and OR in short. For the sake of presentation, we postpone the presentation of further related work to the sections it is referenced.

2.1. Network Utility Maximization

The characteristic property of WMNs is the coupling of link capacities at the PHY. In order to achieve objectives like high throughput, it is necessary to understand and control the cooperation between different layers of the network stack. Network utility maximization (NUM) provides us the necessary tools to understand and design the cross-layer interaction of layered protocol stacks. The concept of NUM is derived from a new perspective on the network that goes back to Kelly *et al.*: They understand the network itself and its operation as an optimizer for a particular optimization problem [59]. Following the theory of optimization decomposition, NUM gives us a formal justification for the layering of protocol stacks. In particular, the concept of layering as optimization decomposition (LAD) understands the modularization process of the network stack as the decomposition of the optimization problem into smaller sub-problems. The inter-connection between them determines the cross-layers interaction. This way, the layers of the network stack are integrated into

a single and coherent theory. They carry out local decisions in order to cooperatively solve a global optimization problem. The framework of NUM can be used in two different ways. On the one hand, we may ask which optimization problem an existing system solves. Through reverse-engineering, this question has been answered for several protocols like TCP. On the other hand, we can specify the objectives and constraints of a system to design and use LAD to derive a network architecture. Note that the same NUM problem can lead to several alternative designs depending on how the problem is decomposed. Further information on NUM and LAD can be found in [19, 32, 109, 127]. The concept of NUM is possibly best described by Yi *et al.* [127] as follows.

> The theory of decomposition of NUM thus becomes a foundation to understand, in a conceptually simple way, the complexities of network architectures: "who does what" and "how to connect them".

In this paper, we are particularly interested in cross-layer designs for WMNs that target at throughput maximization for elastic traffic. In this case, the optimization objective is cumulated utility subject to congestion control, flow conservation and scheduling (and possibly further) constraints [18, 25, 74]. The utility function handles the tradeoff between throughput and fairness between competing flows. A logarithmic utility function results in proportional fairness, which is regarded as suitable form of fairness for WMNs [96, 97]. The optimization problem naturally decomposes vertically. Congestion control is situated at the transport layer and controls the packet ingress according to the queue length at the source. The differential queue back-pressure between neighboring nodes determines the routing decision. Furthermore, the back-pressure determines the MAC layer scheduling. The throughput-optimal algorithm is maximal-weight scheduling (MWS) [113]. In particular, each link is weighted by the differential queue back-pressure. A schedule is a set of links that can be activated concurrently within the same timeslot. The max-weight scheduler is a collision-free, TDMA-like scheduler that selects (one of) the schedule(s) with the highest cumulated weight within each time slot. However, MWS is a NP-hard problem [127]. Thus, the scheduling becomes the bottleneck for a distributed realization of a cross-layer protocol.

Considering the complexity of MWS, the research took two different directions from here on [127]. On the one hand, there are approaches that use heuristics to approximate the max-weight scheduler with reduced complexity. However, even the approximation problem is complex, so that these approaches generally achieve only a fraction of the MWS performance [127]. In addition, they have in common that they rely on message passing and require synchronization to ensure collision-free schedules. On the other hand, a direction in research started from existing random access protocols and considered the question whether and to which extent these protocols achieve optimality. We consider this direction of research in the next section.

2.2. Utility-Optimal CSMA

In the context of cross-layer optimization, random access has also been considered for scheduling. The scheduling problem for ALOHA type networks consists of the adaptation of the persistency probabilities. In particular, slotted ALOHA has been reverse-engineered using the LAD approach [19] and the problem has a desirable structure that allows for a distributed and greedy solution. In particular, the ALOHA capacity region is convex in the logarithmic domain of access probabilities, which is an important prerequisite for the applicability of the NUM approach. Based on this insight, several cross-layer protocols for single-hop and multi-hop communication have been proposed [41, 71, 75]. However, ALOHA scheduling is generally sub-optimal, i.e. it is not possible to achieve the MWS capacity region [127]. We do not go into details with this type of scheduling, since we focus on CSMA throughout the paper. Nonetheless, further information can be found in [127].

Recently, it has been shown that the so-called *utility-optimal CSMA (UO-CSMA)* [1] is able to approximate the MWS arbitrarily close even without message passing. The research directions on utility-optimal CSMA (UO-CSMA) can be categorized as follows.

- Distributed CSMA as Markov random field (MRF), properties of non-adaptive CSMA [24, 58, 72, 73, 92, 102, 118, 123].
- Utility-optimality in continuous time [47, 48, 51, 81, 102].
- Utility-optimality in slotted time with collisions and delayed CS [26, 47, 49, 50, 52, 53, 60, 77, 78, 90, 91, 93, 108].
- Characterization of the delay performance and the short-time fairness; delay-aware scheduling policies [24, 51, 54, 68, 78, 79, 102, 127].
- Different weight functions [33, 102].
- Alternative contention procedures [17, 57].
- Extension to multiple radio frequency (RF) channels [93], multiple modulation and coding schemes (MCSs) [57], MIMO [95], flow level dynamics [10] and optical networks [101, 107].
- Design of protocols and evaluation through simulation and prototypes [2, 3, 70, 80, 110, 124, 130].

Except for the last point, the work is generally theoretical in nature. In the following, we will go through the most important contributions.

The breakthrough that has led to the development of UO-CSMA was the observation that *distributed CSMA* in WMNs can be described as Markov random field (MRF) [24, 58, 72, 73, 102, 123]. Using this approach, each link of the WMN is represented

[1] Also called *adaptive CSMA*. We use both terms interchangeably.

as a node in the link conflict graph (LCG). The link is either in the state transmitting or idle. The state of the network is determined by the joint state of all links, and it forms a Markov chain, the so-called CSMA Markov chain. However, not all possible network states are valid, since the CS prevents the activation of conflicting links according to the LCG. The idle and active durations of a link are exponentially distributed with rate parameters λ and μ, which are either fixed with non-adaptive CSMA or traffic-adaptive with $CSMA(\lambda, \mu)$. The CSMA Markov chain $CSMA(\lambda, \mu)$ is a time-reversible spatial process and its stationary distribution is a MRF [47]. Spatial processes extend the Markov property from time to space, i.e. the state of every link is conditionally independent from all other links it is not in conflict with [47, 58]. It is an *idealized CSMA* model since the CSMA backoff periods, i.e. the idle durations, are continuous in time and instantaneous propagation is assumed, so that no CS collisions occur. Non-adaptive CSMA protocols like IEEE 802.11 try to push the system into states that are the maximal independent sets (MISs) of the LCG, so that the sum-throughput of the network is maximized. However, the network and user objectives are generally different from sum-throughput, which explains the unfairness that is reported wit non-adaptive CSMA [28, 30].

Adaptive CSMA tries to bring the network operation into agreement with the network and user objectives. The canonical form of UO-CSMA for single-hop flows, which accounts for CSMA scheduling and congestion control, is as follows [127]:

UO-CSMA: The transmitter of link l runs $CSMA(\lambda[t], \mu)$ using

$$q_l[t+1] = \left[q_l[t] + \frac{b[t]}{W'(q_l[t])} \left(U'^{-1}(\frac{W(q_l[t])}{V}) - S_l[t] \right) \right]_{q^{\min}}^{q^{\max}}, \tag{1}$$

$$\lambda_l[t+1] = \mu^{-1} \exp\left(W(q_l[t+1])\right), \tag{2}$$

where W is an increasing *weight function*, V is a positive parameter that controls the throughput-delay tradeoff, U is an increasing *utility function* that controls the throughput-fairness tradeoff and $S_l[t]$ is the throughput on link l during slot t. The variable q_l corresponds to the queue length when using a constant *step size* $b[t]$.

The rationale of the algorithm is as follows. We can think of the U'^{-1} term in (1) as the congestion-controlled arrival rate on link l. Thus, q increases if the arrival rate exceeds the service rate S and declines in the opposite case, i.e. it captures the queueing process on link l. The medium access rate λ on link l is adjusted in relation to q in (2), i.e. the scheduler contends for the medium more aggressively if it cannot realize its anticipated flow rate, and it becomes less aggressive if it consumes more resources as requested by the congestion controller. The algorithm does not assume slot-synchronization, it does not need to know the arrival rates in advance and it is furthermore completely distributed, i.e. each transmitter uses local information only.

The above algorithm has been found to be asymptotically utility-optimal in continuous time where collisions due to equal backoff counters do not occur [47, 48, 51, 81, 102]. A further assumption is a time-scale separation, i.e. the underlying CSMA

Markov chain is assumed to converge quickly in-between the parameter updates in (2). The optimization problem that belongs to the solution in (1) and (2) is the following [78].

$$\begin{aligned} \max_{\gamma,\pi} \; & V \sum_{l \in \mathcal{L}} U(\gamma_l) - \sum_{m \in \mathcal{N}} \pi_m \log(\pi_m) \\ \text{s.t.} \; & \gamma_l \leq \sum_{m \in \mathcal{N}: m_l = 1} \pi_m, \quad \sum_{m \in \mathcal{N}} \pi_m = 1 \end{aligned} \tag{3}$$

In the formulation above, γ_l denotes the long-term throughput on link l, $\mathcal{L}$ is the set of links, $\mathcal{N}$ is the set of schedules and π is the distribution of the CSMA Markov chain over all schedules $m \in \mathcal{N}$. The notation m_l refers to the state of link l within schedule m, which could either be transmitting (1) or inactive (0). Surprisingly, the objective (3) contains the entropy of the CSMA Markov chain, in addition, and we will consider the consequences within one of the next paragraphs.

The utility-optimality of UO-CSMA has been extended to slotted time without the time-scale separation [26, 47, 49, 50, 52, 53, 60, 77, 78, 90, 91, 108]. In particular, Jiang *et al.* and Liu *et al.* use a very intuitive idea to establish the result [47, 78]. The average backoff λ^{-1} in (2) can be traded off against the average channel holding time μ^{-1} within $CSMA(\lambda, \mu)$. Thus, instead of decreasing the backoffs, which would result in severe collisions in slotted time, the idea is to hold the channel longer and thus transmit more frames per successful contention attempt. In addition, the use of request to send (RTS)/clear to send (CTS) keeps potential collision durations short. This way, the impact of collisions remains arbitrarily small while the utility-optimality is preserved. However, the short-term fairness seriously suffers in the proposed approach. In contrast, Ni *et al.* use a slightly different approach [90, 91]. They assume that the time is slotted into super-frames consisting of a signaling and a data subframe. Collisions occur only within the signaling frame using a collision avoidance (CA) technique based on RTS/CTS. Interestingly, the modeling approach leads to the same product form distribution and thus to same optimality results. Furthermore, Proutiere *et al.* have shown that a time-scale separation between the discrete update process of the CSMA parameters and the continuous evolution of the network dynamics is not necessary [93]. In particular, if the network dynamics can be modeled as a continuous-time Markov process, then UO-CSMA achieves asymptotical optimality with properly diminishing step sizes even in the case that the control parameters are updated at the same time-scale. Furthermore, constant but small step sizes achieve a weak convergence in the sense that the system converges into a small neighborhood of the optimum. A remarkable point in the work of Proutiere, Liu *et al.* is the assumption that the queue length at the optimal working point must be within the specified limits in (1) for the parameter set used [78, 93]. Later in section 6.2, we will consider this issue in detail and present an algorithm that dynamically adapts the system parameters accordingly.

The results above are promising. However, now we should take a look at the other side of the coin and ask about the price we have to pay. In particular, the MWS problem is NP-hard and does not allow for polynomial time approximations [127, 129].

How is it possible that UO-CSMA solves the problem asymptotically throughput optimal at a significant low level of complexity? The reason lies within a three-dimensional tradeoff between throughput, complexity and delay. We can think of it as follows. If we increase the parameter V in (1), the queue length as well as the medium access rates increase and we approach a more efficient working point in terms of throughput or utility. According to [47, 48, 78], the optimality gap of the problem (3) is bounded as follows.

$$\left| \sum_{l \in \mathcal{L}} \left(U(\gamma_l[\infty]) - U(\gamma_l^\star) \right) \right| \leq \log |\mathcal{M}| / V \leq |\mathcal{L}| \log(2) / V \tag{4}$$

In the expression above, $\gamma_l^\star$ is the optimal throughput under MWS and $\gamma_l[\infty]$ denotes the solution to (3) with diminishing step sizes. The utility gap, i.e. the right-hand side (RHS) of (4), is determined by the number of active links $|\mathcal{L}|$ and the parameter V. The derivation of the RHS relies on the fact that the entropy of the distribution of the CSMA Markov chain in (3) is upper bounded. Thus, the algorithm converges into a small neighborhood of the MWS solution that depends on both $|\mathcal{L}|$ and V. An alternate interpretation of (3) and (4) is as follows. Without external energy ($V = 0$), the system (3) follows the principle of maximal entropy. However, in order to achieve our objectives we have to push the system in the desired direction using $V > 0$. The more energy we dedicate to the system, the more the system behaves according to our intentions.

However, we have to pay for a small utility gap in terms of short-term MAC fairness and end-to-end delay [24, 51, 54, 68, 78, 79, 102, 127]. The intuition behind this result is as follows. Let us consider an even number of nodes that are placed on a circle with equal distances. Single-hop traffic flows exist in counterclockwise direction and the primary interference model is used. In that case, the MISs consists of two schedules that contain the even and the odd links, respectively. Increasing the parameter V leads to longer queues and shorter backoffs. Thus, it takes longer to switch from one MIS to another [24], i.e. the system "locks into" a particular schedule [79]. The consequences are interesting: The throughput benefits from it since the system spends more time in the MIS. Remember that throughput is an asymptotic concept that averages over a long duration. On the other hand, the short-term fairness declines since the inactive links have to wait longer to regain medium access. The end-to-end delay suffers as well from larger queues according to Little's law [7].

The derivation of exact expressions for the delay performance is challenging due to the coupling of queues. The interested reader may refer to [90] for an overview of the current delay approximation and bounding approaches. However, a very general result from collision-free MWS is the three-dimensional tradeoff between throughput, delay and complexity [127, 129]. As shown in Fig. 1, greedy schedulers have a lower complexity, but they achieve only a fraction of the throughput region of MWS. Interestingly, we can trade complexity off for delay with a scheduling policy called randomized pick-and-compare (RPC) [127]. Instead of calculating the MWS for each time slot anew, the idea of RPC is to persist within "good" schedules, so that the

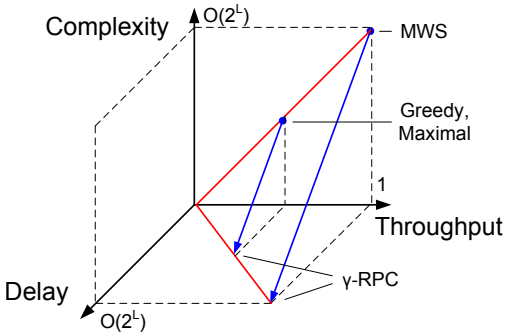


Figure 1: Three-dimensional tradeoff between throughput, delay and complexity (According to [127, 129]).

costs of finding them amortize over time. Hence, RPC sacrifices delay, while the throughput remains unaffected since it is an asymptotic concept. It is assumed that the tradeoff structure of UO-CSMA is comparable [76, 78, 79, 127].

Thus, delay cannot be eliminated, but unnecessary latency can be removed and the throughput-delay tradeoff can be controlled. For the latter goal, two approaches have been presented in recent literature. According to Jiang *et al.*, the congestion controller should inject slightly more (virtual) qs than (physical) packets [47]. This way, the actual arrival rate of physical packets in (1) becomes less than the (physical) service rate of the link, so that the (physical) queues tend to zero. The price to pay is throughput, since we have to emit *dummy packets* if the physical queue runs empty. Lotfinezhad *et al.* presents another interesting approach to control the tradeoff that does not rely on dummy packets [79]. In particular, they propose to regularly "reset" the state of the CSMA Markov chain to zero, which prevents an excessive locking of the system within a schedule. Interesting approaches exist also for the former goal of removing unnecessary delays. One observation with back-pressure routing is that a lot of packets get stuck within intermediate queues only to provide differential backlog. However, these physical packets can readily be substituted by virtual packets that are nothing more than counters [12, 83, 99]. Ni *et al.* propose a dual scheduler with two different policies [90, 91]. If the back-pressure of the considered link exceeds a threshold, the throughput-optimal policy is used. In the remaining case, the authors propose to use a heuristic policy that has a better delay performance. Furthermore, Moeller *et al.* report delay improvements from changing the queuing discipline to last in, first out (LIFO) [83].

The influence of the weight function in (1) is less well understood [70]. Rajagopalan *et al.* propose to weight the queues with the very slow increasing function $W(q) = \log\log q$. It is argued that the shallow slope is necessary for unsaturated sessions [102], whereas the requirements on the weight function can be relaxed with saturated sessions, because the decreasing step size makes the network dynamics slow enough so that noise is averaged out [127]. However, the shallow slope leads to longer queues

and delays. Ghaderi *et al.* present another weight function that is able to reduce the end-to-end delays and the convergence time of the system [33].

There are also some initial results targeting the feasibility of UO-CSMA on state-of-the-art hardware [2, 3, 70, 80, 110, 124, 130]. In most papers, the shortcomings of commodity IEEE 802.11 hardware are discussed, since only marginal parts of the MAC operation are accessible. For example, it turns out that the realization of adaptive backoffs is difficult. Interestingly, the MAC of wireless sensor network (WSN) hardware allows for a better access [110]. The perspective of the early papers is different from UO-CSMA [2, 3, 110, 124]. In particular, the authors intend to find heuristic and empirical solutions that approximate the collision-free MWS with IEEE 802.11 as good as possible. On the other hand, Lee *et al.* present a prototype and some initial results for UO-CSMA using simulation and a real world test-bed [70, 130].

2.3. Opportunistic Routing

Opportunistic routing (OR) is a concept to benefit from multi-user diversity (MUD) in WMNs. The wireless communication is prone to transmission errors in a different and more serious way than communication via wires. For example, empirical results suggest that the majority of links in a WMN are of intermediate quality [1]. Thus, instead of transparently hiding the characteristic of wireless links from the upper layers, the advent of cross-layer routing metrics like ETX [20] and ETT [22], for example, indicates that there are strong incentives for WMN-aware network and routing layers. The rationale of OR is to mitigate transmission errors via the simultaneous usage of multiple candidate receivers. Due to the broadcast nature of the wireless medium, the transmitter does not have to specify the relay on the next hop a priori. Instead, the routing decision is made a posteriori relying on the information which candidate actually received the frame. The additional process of determining the next-hop relay within the MAC transaction is called *forwarder selection* or alternatively *relay (self)-selection*.

Most of the recent literature reviews OR from a MAC centric perspective and focuses on the forwarder selection. A detailed review can be found in [139]. In the following, we highlight the most important contributions only. Selection diversity forwarding (SDF) as proposed by Larsson *et al.* uses a four-way handshake for forwarder selection [65]. The approach is refined to a two-way handshake by Valenti *et al.* [116] and by Biswas *et al.* [8]. The latter authors propose a technique called *slotted acknowledgement (ACK)* in order to orthogonalize the ACKs of all candidates in time. Alternative designs that estimate the instantaneous channel conditions and select the bit-rates accordingly have been proposed by Larsson *et al.* [66], Zubow *et al.* [142], Wang *et al.* [121, 122], Ji *et al.* [46] and Bletsas *et al.* [9]. Assuming that the wireless channel evolves on a larger time scale compared to the frame duration, an additional probing stage is able to determine the instantaneous channel quality to all candidates, so that the best bit-rate for the following data transmission can be estimated. The overhead of the slotted ACK can be further reduced by compression [140], which relies on the CS capabilities of the transceiver to detect missing ACK frames. Bletsas *et al.* propose an alternative to the slotting of the probe replies: The

candidates have to contend among each other according to their instantaneous channel conditions [9]. Thus, the actual data transmission stage is similar to unicast. Yang *et al.* propose a similar contention approach for the ACK stage of the MAC transaction [126]. Furthermore, the overhead of relay self-selection can be significantly reduced via logarithmic splitting algorithms [4, 122].

In recent literature, two categories of forwarder selection policies can be found. With relay *self*-selection, we refer to a distributed agreement among all candidates. The objective is that the best candidate according to a metric like expected transmission count (ETX) [8], geographical distance [138] or instantaneous channel conditions [9] should be selected in a distributed way. Self-selection has to handle the inherent risk of coordination failures due to communication errors, which may cause duplication of data frames, among others. On the other hand, *transmitter-based relay selection* centralizes the selection responsibility to a single entity, e.g. to the transmitter [65, 85, 136]. This way, the problem of duplicates can be mitigated at the expense of a four-way handshake [65, 85]. The overhead can be further reduced using robust acknowledgements as proposed by Zhong *et al.* [136]. The two additional handshake stages of one transmission are integrated in the following transmission, which can be seen as a special incarnation of passive acknowledgements. In addition, the complexity of the signaling schemes can be reduced by network coding [15, 35, 98, 100, 135]. This way, OR becomes more robust against coordination errors so that the overall efficiency increases.

On the one hand, the identity of the candidate receiver is not relevant for the transmitter of a data frame. What matters is that at least one of them successfully received the frame. On the other hand, the data frame is generally present at several relays of the same routing stage, and a downstream candidate receiver does not care about the actual origin of that frame. In particular, it is not necessary to orthogonalize the transmissions of the candidates in the former case. Instead, advanced PHY techniques enable cooperative transmissions, in which all candidate receivers can take part in. In the latter case, there is no need to decide for only one relay at a particular routing stage. Instead, the data frame can be cooperatively forwarded by some or possibly all relays. Using the described ideas, we have proposed an OR protocol for WMNs based on multiple-input single-output (MISO) and STC as enabling technology for cooperation on the PHY [63].

The objective of the forwarder selection of the MAC centric approaches above is the selection of the *best* relay according to a given metric. The underlying assumption is that the local decision is also beneficial from the global point of view. The metrics used consider geographic locations [138], instantaneous channel conditions [9] or topology and link quality information [23, 67, 136]. However, these approaches split the flow onto the outgoing links without considering the interference relationship between transmitters, so that spatial diversity due to multi-path routing cannot be exploited systematically. Furthermore, they split the flow unaware of their own and foreign traffic. As we will see in section 8, the traffic-unawareness may result in congestion on paths that do not perform as well as the metric suggested. Thus, the end-to-end congestion controller will limit the flow rate on *all* paths regardless that some of them might support even more traffic.

In conclusion, the area of forwarder selection strategies is well explored, and we have highlighted the most important contributions above. However, considering OR from the MAC perspective only is not sufficient in order to achieve desirable end-to-end properties. In particular, it is crucial to consider the question "What are the *best* relays?" from a cross-layer perspective.

The network and cross-layer aspects of OR have been less well explored. Neely *et al.* use control theory to derive a back-pressure scheduling and routing algorithm for OR with perfect TDMA-like scheduling [85, 87]. In special cases, the algorithm supports decentralized and channel-blind operation. Zeng *et al.* cast OR as a linear program and analytically explore bounds on the performance assuming perfect MAC scheduling [134]. In the same line, Zhang *et al.* formulate a convex optimization problem for network-encoded OR and derive a distributed online algorithm using decomposition and duality [135]. In particular, they use an approximation of the OR capacity region within their optimization and assume perfect MAC scheduling as well. The closest to our work is the optimization framework for network-encoded OR by Radunovic *et al.* [98, 100]. As above, they assume perfect MAC scheduling, but they use exact expressions for the OR capacity region. In addition, they derive a protocol called multi-path code casting from their optimization framework [35] that solves the scheduling, opportunistic routing and congestion control problem.

The bottleneck in the contributions above is the attempt to solve the MAC scheduling problem directly. MWS is the optimal strategy [113], which is known to be NP-hard and thus hard to solve even in centralized settings [127]. Thus, most of the approaches above rely on heuristics for random access in order to enable a distributed operation. It is to note that Jiang *et al.* present an extension of UO-CSMA to anycast [47, ch. 9.7.4]. However, they use the term *anycast* to address multi-path routing. In particular, they do not consider the OR capacity region resulting from lossy links and the MUD gain.

3. A Node-Oriented Model of CSMA in WMNs

Recent modeling approaches of CSMA in WMNs employ a link-based CS, i.e. they characterize interference by links (cf. section 2). The contention is processed by the link as a whole, which requires that transmitter and receiver have the same view on the channel [70]. However, the reality with state-of-the-art technologies is different. As we have shown in [62], the CS range with IEEE 802.11b does not significantly extend beyond the receiving range, which gives rise to the well-known *hidden node* problem [114]. Hidden interference is likely the most important source of packet loss in today's WMNs [36]. It renders the network capacity region non-convex [16] and causes severe performance degradation [52]. Thus, it is essential to consider the hidden node problem before any optimization can be applied.

Two approaches can be found in recent literature to mitigate the problem. Jiang *et al.* justify the link-oriented idealized CSMA model using a hidden node free design (HFD) [55], which eliminates hidden nodes by lowering the CS power threshold. However, physical and technological constraints impose a lower limit on the

CS threshold. Thus, for a hidden node free design, large signal to noise ratio (SNR) margins of up to 20 dB have to be introduced [55]. This way, the CS state at the transmitter is sufficient to determine the interference situation at all potential receivers. Nevertheless, the large margins waste a huge amount of spatial resources. Collision avoidance (CA) is a further approach to mitigate the hidden node problem [6], which comes at the expense of exposed nodes. However, the effectiveness of CA based on the RTS/CTS exchange is limited in WMNs [24].

One might ask what the optimal CS range for a CSMA network is. Ven *et al.* answer this question for the protocol model [117]. For high traffic scenarios, the best strategy is to set the CS range just large enough to preclude all hidden collisions. Increasing the CS range further consumes spatial resources only without any benefits in return, as it is the case with HFD.

Our objective in this section is the design of a MAC model for OR in WMNs. The HFD approach does not meet the requirements for OR for several reasons. Due to the large SNR margins with HFD, spatial resources are allocated unnecessarily, which leads to a low spatial reuse. Furthermore, the wireless links will most likely be lossless due to the large SNR margins, so that there are no multi-user gains with anycast. Furthermore, HFD silences *all possible* receivers, so that additional receivers can be included in the MAC transaction at no costs in terms of spatial resources. Thus, HFD does not capture the characteristic tradeoff between opportunistic gain and spatial reuse that arises with OR. Based on our observations about CS in typical WMNs in [62], we develop a *node-oriented* CS model that reconsiders the operation of CA with respect to the characteristics of mesh networks. Furthermore, it supports OR via anycast transmissions and allows for a higher spatial reuse through a selective allocation of necessary spatial resources. Later in section 4, we design a MAC protocol according to this model.

3.1. CSMA/CA as a Markov Random Field

During the MAC transaction, one transmitter tries to send a frame to at least one receiver. We assume that the MAC operates bidirectional, so that transmitter and receiver(s) can exchange signaling information. Furthermore, we assume that the MAC transaction is opaque in terms of interference, i.e. the same interference constraints apply to all involved nodes in the same way. For example, we do not allow that another transmission takes place even if it would interfere with the transmitter of an ongoing transmission only, which is the well-known exposed node situation [55]. This way, an actual MAC protocol remains flexible in the realization of signaling schemes, since there are no temporal dependencies between different links involving the role assignment.

A wireless link can be in one of the following four states: In the contending state (0), the link is idle and tries to get access to the medium. If it wins the contention, it enters one of the probing states (1, 3). Both states are technically equivalent, however, we associate two different values to emphasize the outcome of the probe. If the probe is successful (state 1), the link enters the data transmission state (2), otherwise it proceeds from state 3 to state 0. The reader might associate the RTS of IEEE 802.11 with

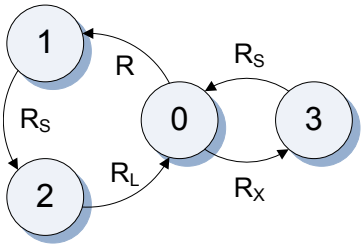
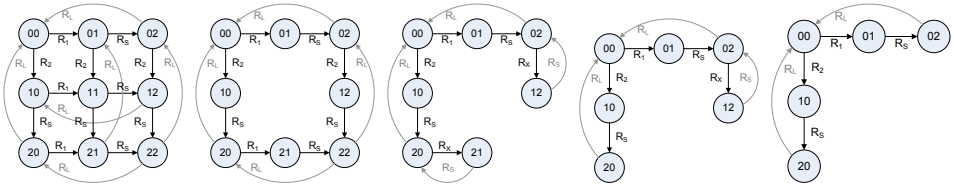


Figure 2: Markov Chain Model of the internal CSMA Operation of a wireless link.



(a) No Conflict (b) Conflict 1 (c) Conflict 2 (d) Conflict 3 (e) Conflict 5

Figure 3: Types of Link Conflicts. The nodes are states in the joint CSMA Markov chain, i.e. the first and second digits correspond to the states of the first and second link, respectively. Arrows mark possible transitions.

the probing states [43]. As shown in Fig. 2, the operation of link k can be modeled as a Markov chain when assigning the medium access rate $R = R_k$, signaling rate R^s and transmission length R^l to the edges. The synchronization rate R^x captures the extent of failed probes that might arise, which we consider in detail shortly. Let C denote a constant that normalizes the stationary distribution P to a probability distribution, then we have

$$P(v) = \frac{1}{C} \begin{cases} 1 & v = 0 \\ R/R^s & v = 1 \\ R/R^l & v = 2 \\ R^x/R^s & v = 3 \end{cases}$$

The link conflict graph (LCG) captures the conflicts between any two links. Each wireless link of the network is represented as a site in the LCG $G = (\mathcal{V}, \mathcal{E})$. The set $\mathcal{V}$ contains K sites, which are both unicast and anycast links. We assign a value $v_k \in \{0, 1, 2, 3\}$ to each link k in $\mathcal{V}$ corresponding to the state of the link. The K-dimensional vector $\boldsymbol{v} = (v_1, v_2 \dots v_K)$ captures the state of all links in the LCG. In the following, we assume that the state space V of $\boldsymbol{v} \in V$ is indexed by j; hence, the notation v^j refers to the j-th vector of V, and v_k^j selects the k-th element of it.

Due to the CS constraints, the space $\{0, 1, 2, 3\}^K$ generally contains infeasible states.

For example, two neighboring transmitters k and l are not allowed to access the medium simultaneously, i.e. if $v_k \neq 0$ then $v_l = 0$ and vice versa. The LCG contains a weighted edge $e_{k,l} \in \mathcal{E}$ between site k and l if the links are in conflict and the weight determines the type of conflict. By exhaustive search over all topologies, we identified 5 structurally different conflict types, which are depicted in Fig. 3. In particular, if both transmitters are separated by more than 3 hops, they can operate simultaneously, i.e. no conflict exists (Fig. 3a). Otherwise the transmitters are within each other's (3-hop) contention neighborhood and conflict with each other. With type 1 conflicts (Fig. 3b), only the probes must be serialized, i.e. the joint state 11 does not exist, but concurrent data transmission is possible. A type 5 conflict results from a topology in which both transmitters are neighbors, for example. In this case, every transmitter has perfect knowledge about the state of the medium at the other node, which prevents them from accessing the medium simultaneously.

With conflict type 2, 3 and 4, which is opposite symmetric to type 3, there is an *information deficit* at either one or both transmitters. Ongoing data transmissions may exist that are hidden from the transmitter but in the vicinity of the receiver. In that case, if the transmitter experiences an unsuccessful probe, we will *block* the link until it is *re-synchronized*. Since the unsuccessful probe indicates that the intended receiver is occupied, the blocking prevents further excessive probing without success. Unsuccessful probes cannot be prevented, but they should be limited while ensuring a fair competition between links. However, the blocked sender needs to know when the receiver becomes available again, so that the competition with hidden nodes remains fair in the next contention period. Note that the IEEE 802.11 CA does not meet this requirement. The transition rate into the state of an unsuccessful probe R^x depends on how often the link is re-synchronized, which in turn depends on the data frame length and the number of active data transmissions in the contention neighborhood of the transmitter. In the following, we make the simplifying assumption that R^x is independent from R. The assumption might not be fulfilled in real systems. However, the model remains very accuracy, as we will point out in the evaluation.

The spatial dependencies of CSMA in WMNs can be captured by a Markov random field [24, 48, 72]. Spatial processes generalize the Markov property of limited dependency from time to space [58]. The extended CSMA Markov chain presented above is a MRF, and we give an intuition for this result in Appendix A. In the following, we substitute $R = \exp(r)$ and we refer to r as transmission aggressives (TA). Furthermore, the indicator function $\delta^I(v_k^j) = \delta_{j,k}^I$ classifies the state j: It evaluates to 1 if $v_k^j \in I$ and 0, otherwise. The stationary distribution p of the CSMA Markov chain

has product form since it is a MRF and is given by

$$\begin{aligned} p(v^j;\boldsymbol{r}) &= \frac{\prod_{k=1}^{K} P(v_k^j)}{C(\boldsymbol{r})} \\ &= \frac{\exp\left(\sum_{k=1}^{K} \delta_{j,k}^1(r_k - r^s) + \delta_{j,k}^2(r_k - r^l) + \delta_{j,k}^3(r^x - r^s)\right)}{C(\boldsymbol{r})} \\ C(\boldsymbol{r}) &= \sum_j \exp\left(\sum_{k=1}^{K} \delta_{j,k}^1(r_k - r^s) + \delta_{j,k}^2(r_k - r^l) + \delta_{j,k}^3(r^x - r^s)\right) \end{aligned} \tag{5}$$

The model operates at two spatial levels. For data transmissions, it considers the *necessary* receivers only and generates a 1-hop exclusion region around transmitter and receiver(s), which is set up during the probing state. However, to protect the probing frames from hidden interference, the transmitter has to establish a larger exclusion region comparable to HFD, which we call (3-hop) *contention neighborhood*. The contention neighborhood is reserved only for a short probing duration. After the probe, the allocated spatial resources are reduced in any case: Either the probe has failed and the exclusion is completely removed, or the probe has been successful and the exclusion is reduced to the (1-hop) neighborhood around transmitter and receiver(s). This way, the spatial reuse is increased compared to HFD, and the model captures the characteristic tradeoff between multi-user gain and spatial reuse.

For unicast, the probing policy is obvious: Only if the CTS has been returned, the sender proceeds with the transmission of the data frame. However, it becomes more involved with anycast. The model demands an *all-or-nothing policy*, i.e. the sender transmits the data frame only if all CTSs have been returned. This approach seems cumbersome on first sight, and one might be tempted to address the data frame to the subset of candidates that have transmitted a CTS. However, the stationary distribution of the CSMA Markov chain would lose its product form this way. On the other hand, if we would ignore missing CTSs and use all receivers in any case, it would be hardly possible to reason about the link quality for data transmissions.

Using the abbreviation $p_j(\boldsymbol{r}) = p(v^j;\boldsymbol{r})$, we define the throughput q_k of a hyperlink k as

$$q_k(\boldsymbol{r}) = \sum_j \delta_2(v_k^j) p(v^j;\boldsymbol{r}) = \sum_j \delta_{j,k}^2 p_j(\boldsymbol{r}).$$

The summation proceeds over the complete state space V. Furthermore, we introduce the superscript i on q_k^i that refers to the state, in which the throughput was generated. For example, $q_k^2 = q_k$ is the data throughput, and q_k^1 and q_k^3 refer to the throughput for successful and unsuccessful probing.

3.2. Operating CSMA/CA in WMNs

In the same line as [48], the TAs $\boldsymbol{r}$ that stabilize the network within a particular steady state $\bar{p}$ of the CSMA Markov chain can be found via maximum likelihood (ML) es-

timation, assuming that the input rates $\lambda_k = \sum_j \delta^2_{j,k} \bar{p}_j$ are feasible. Note that the objective of the MAC is not the maximization of the link throughput. Instead, the maximal feasible flow and link rates are determined by the congestion controller and the routing layer (cf. section 5), and the objective of the MAC is to adapt its operation parameters to achieve the predetermined rates. The ML estimate [11, p. 361] of $\boldsymbol{r}$ is found by maximizing the log-likelihood function $F(\boldsymbol{r})$ with respect to $\boldsymbol{r}$. Plugging (5) into $F(\boldsymbol{r}) = \sum_j \bar{p}_j \log(p_j(\boldsymbol{r}))$,

$$F(\boldsymbol{r}) = \sum_{k=1}^{K} \Big(\lambda_k (r_k - r^l) + \lambda^1_k (r_k - r^s) + \lambda^3_k (r^x - r^s) \Big) - \log(C(\boldsymbol{r}))$$

and maximizing $F(\boldsymbol{r})$ with respect to $\boldsymbol{r}$, we get

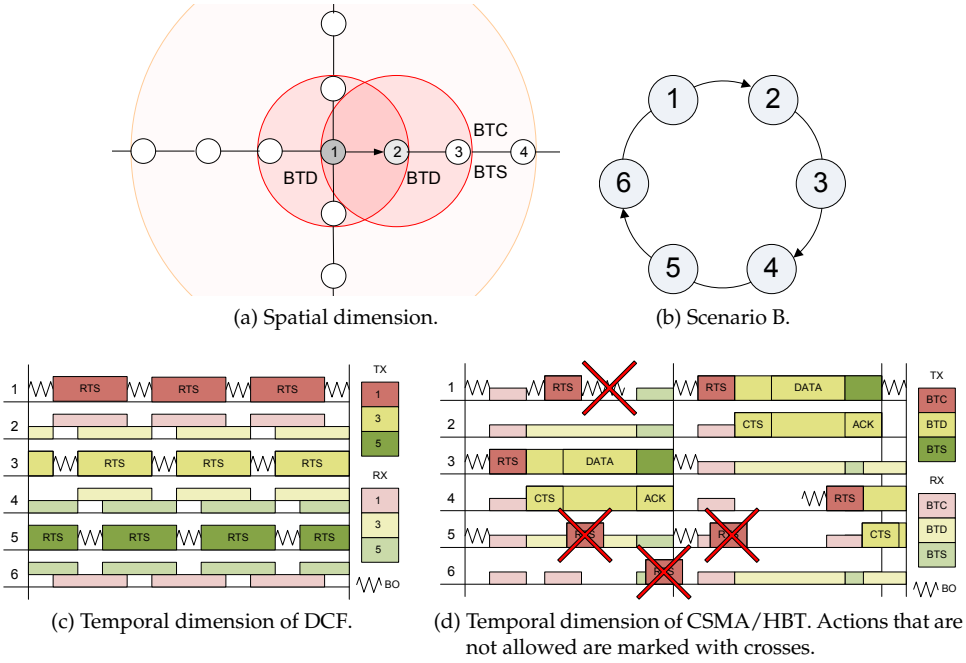
$$\begin{aligned} \frac{\partial F(\boldsymbol{r})}{\partial r_k} &= \lambda_k + \lambda^1_k - \sum_j \delta^{1,2}_{j,k} \\ &\quad \cdot \frac{\exp\left[\sum_{i=1}^{K} \delta^1_{j,i}(r_i - r^s) + \delta^2_{j,i}(r_i - r^l) + \delta^3_{j,i}(r^x - r^s)\right]}{C(\boldsymbol{r})} \\ &= \lambda_k + \lambda^1_k - \sum_j \delta^{1,2}_{j,k} p_j(\boldsymbol{r}) = \lambda_k + \lambda^1_k - q_k - q^1_k, \end{aligned}$$

where λ^1_k and q^1_k are the target and observed signaling costs for successful probing. Due to our assumption about the independence of r^x from any other r_i, the costs for unsuccessful probing are not considered in the estimation of $\boldsymbol{r}$. By construction, the overhead for successful probing is a constant fraction of the data throughput, i.e. q^1_k and q_k differ only by the constant factor R^l/R^s. Hence, we can update the transmission aggressivess (TAs) via the sub-gradient $r_k \leftarrow [r_k + \alpha(\lambda_k - q_k)]_+$, which means that the queue backlog can be used for CSMA scheduling.

In the derivation, we assume that every sender is *saturated*, i.e. the queues are non-empty at all times. This assumption may be less a problem for the flow source, if the arrival process provides sufficient packets. However, especially at the intermediate relays the queues have to be filled via the upstream. As in [48], we may transmit dummy packets if the queue runs empty in order to fulfill the assumption. In section 6.3, we will present an alternative approach to deal with empty queues.

4. CSMA/CA with Hierarchical Busy Tones

In the following, we illustrate how to modify the well-known IEEE 802.11 protocol to conform to the enhanced CSMA model above, and we refer to the refined protocol as CSMA/CA with hierarchical busy tones (CSMA/HBT). The model above is based on the rationale that every link can increase its service rate by using higher TAs, so that it can resolve any unfairness locally. Even with adaptive backoffs, the distributed coordination function (DCF) of IEEE 802.11 cannot ensure this requirement. For example,



(a) Spatial dimension.

(b) Scenario B.

(c) Temporal dimension of DCF.

(d) Temporal dimension of CSMA/HBT. Actions that are not allowed are marked with crosses.

Figure 4: Spatial and temporal dimensions of CSMA/HBT.

consider scenario B in Fig. 4b. Three transmitters are connected to their respective receivers and to one receiver of another link, in addition. If the transmitters 1, 3 and 5 increase their TAs, so that the average backoffs become significantly smaller than the RTS duration, then a life-lock occurs. As shown in Fig. 4c, every sender constantly probes the channel only to find out that its intended receiver is silenced. At the same time, its probe silences the receiver of another link in a circular fashion, so that no link is able to transmit data in the end.

The model in section 3 acquires resources in two stages. The rationale is to use a large exclusion region, the so-called contention neighborhood, during the short probing stage in order to protect the RTS frames from hidden interference. In the following data transmission stage, the exclusion region is reduced to the necessary level in order to allow for higher spatial reuse. In particular, CSMA/HBT relies on physical CS only, whereas virtual CS is not used. In the same line as dual busy tone multiple access (DBTMA) [42], we conceptually separate CS from data transmissions into three exclusive busy tone (BT) channels. We assume that the BT receiver is able to detect activity in each channel with a high probability even if multiple BT signals are superimposed. Interestingly, approaches similar to DBTMA find their way into cellular systems, as well [34].

There are three types of BTs: The contention busy tone (BTC) accompanies the RTS (state 1). It covers the contention neighborhood around the transmitter (cf. Fig. 4a).

It indicates that the medium is busy, i.e. every receiver of the BTC should suspend its contention. On entering state 2, the transmitter turns the BTC off and activates the data busy tone (BTD), which covers and silences its (1-hop) neighborhood (cf. Fig. 4a). If an addressed receiver is not silenced by ongoing transmissions, it also activates its BTD and sends a CTS. All returning CTS packets are properly orthogonalized, e.g. in time, space or frequency. In the case all CTS have returned, the transmitter initiates the actual data delivery, otherwise it goes back to contention (state 0) and marks the link as blocked, i.e. it does not contend on the link until it is re-synchronized.

Every receiver acknowledges the successful reception of a data frame using an ACK packet. At this time, the transmitter has to re-synchronize all blocked links in its contention neighborhood. This way, blocked links in the contention neighborhood become aware of the start of the next contention period, so that the competition among all transmitters remains fair. In particular, the transmitter actives the synchronization busy tone (BTS) to announce the end of its medium access. The BTS covers the (3-hop) contention neighborhood of the transmitter similar to the BTC (cf. Fig. 4a). On finishing the MAC transaction, the transmitter deactivates the BTS, which releases all blocked links and synchronizes all nodes in the contention neighborhood, so that they start to contend for medium access again.

Note that the outcome of the probing stage is already determined at the beginning of the probe. However, the transmitter is not aware of it until the end of the probe. If the outcome is positive, the access to the medium is granted and no other link may interrupt the ongoing communication. All BTs prevent receiving nodes from contending for medium access, i.e. they suspend their backoff on receiving a BT. However, there is an important difference: On sensing the BTD, the node should be completely silent. In particular, it should not answer any incoming RTS, since a data transmission is already active in its direct neighborhood.

There are multiple alternatives for a particular realization of the BTs. For example, they can be allocated in the white spaces of the wireless spectrum. The three-hop coverage can be ensured through higher transmission power, smaller signal bandwidth or lower carrier frequencies. Furthermore, advanced concepts like relaying and multi-carrier technologies like OFDMA may assist in the realization of the BTs. In addition, the costs for an additional BT transceiver should be reasonable in relation to the hardware for time synchronization that would be necessary for collision-free MWS.

Referring back to the topology in Fig. 4b, CSMA/HBT prevents the life-lock since it removes the circular dependency using a larger contention neighborhood. For example, if transmitter 3 sends an RTS as shown in Fig. 4d, all other transmitters are silenced by the BTC. After the probing, only node 2 and node 5 remain in the reduced exclusion region as determined by the BTD. On the other hand, transmitter 1 is not silenced and may send an RTS as shown in the figure. Since the intended receiver is silenced, node 1 will block its link to node 2 until it is re-synchronized by the BTS, or a timeout occurs. However, it will suspend its contention on that link in order to prevent excessive probing. Similar to BTC, the BTS should cover the (three-hop) contention neighborhood, so that all contenders are synchronized to the start of the next contention period. Note that an interleaved operation as shown with transmitter 4 in

Fig. 4b is possible.

As long as one link is in a probing state (1 or 3), all links within the contention neighborhood should not change their state (which can only be 0 and 2) to comply with the proposed model in section 3. To prevent collisions between probing frames, the model has no joint probing state 11 for conflicting links (cf. Fig. 3). However, allowing a transmitting link (state 2) to become idle (state 0) while a conflicting link sends a probe introduces a non-reversible transition in the Markov chain, since the link cannot become transmitting again (state 2) as long as the conflicting link remains in the probing state. Thus, the product form distribution would be lost this way (cf. Appendix A). Hence, the transmission of a link should be prolonged by the time the respective contenders spend in probing in order to comply with the model. From the practical point of view, this is difficult to achieve. Fortunately, through simulations we have gained evidence that the discrepancy to the model is acceptable even when ignoring this issue.

4.1. Link Level Evaluation

In this section, we evaluate the protocol analytically and through simulation in order to derive characteristic properties. Our intention is threefold. At first, we have not said much about the synchronization rate, which determines the amount of probing failures. In the following, we will establish bounds on the synchronization rate for CSMA/HBT and illustrate the probing overhead in typical cases. Secondly, we evaluate CSMA/HBT in non-idealized settings, in which collisions can occur. We estimate the amount of collisions and formulate the tradeoff between collisions and efficiency that arises with adaptive CSMA. Thirdly, we have used the protocol model in the design of CSMA/HBT. However, the signal propagation is more sophisticated in physical environments. We illustrate the degree to which CSMA/HBT is able to achieve orthogonality in the medium access and discuss arising risks for the following optimization problem in section 5.

4.1.1. Synchronization Rate and Probing Overhead

The probing overhead consists of the successful and unsuccessful probes. The former are a fraction of the achieved throughput. The latter are determined by the synchronization rate r^x in the model in section 3. In the following, we characterize CSMA/HBT in terms of synchronization rate. From the technological point of view, the amount of unsuccessful probes within CSMA/HBT depends on how often an unblocked link experiences probing timeouts (*a priori failure*), and on how often false unblocking occurs, i.e. a previously blocked link unnecessarily probes the channel due to an unblocking signal from another node in the contention neighborhood (*unblocking failure*). Remember that the BTs do not carry any additional information like sender addresses.

An exact expression for the *a priori failures* is hard to obtain: It depends among others on the frame sizes, the number of active transmitters in the contention neighborhood and the joint distribution of link states, and it may furthermore vary in time.

Nevertheless, we can upper bound the a priori failures in the following way. Consider a system with two links, which experience an information deficit as shown in Fig. 3c. In the scenario, unblocking failures cannot occur since there are only two links. At the time the primary link decides to probe the channel, the secondary link can be either idle or active, but the actual state is hidden from the primary link due to the information deficit. The probing succeeds if the secondary link is idle, and the probing overhead is a constant fraction of the transmission duration. If the secondary link is active, on the other hand, the probing attempt is not successful. Thus, the primary link remains blocked for the rest of the data transmission on the secondary link. The secondary link synchronizes the other link at the end of its data frame, which enables a fair competition between both links. In the worst case, the secondary links always wins the medium contention. In this case, the primary link sends exactly one probe packet within each frame duration of the secondary link. In summary, the probing overhead for a priori failures is bounded by the overhead of successful probes on the competing link in the considered scenario, i.e. we have $r^x \leq r^l$.

In the same way, we can argue about bounds on *unblocking failures*. If there are multiple transmitters within the contention neighborhood, the primary link may experience unblocking failures. In particular, let n be the number of additional links within the contention neighborhood, which can be concurrently activated. In the worst case, the secondary link is transmitting a data frame, and all n remaining links trigger an unblocking failure within the transmission duration (assuming equal frame sizes). Hence, the worst cast probing overhead increases by a factor of n due to unblocking failures, i.e. we have $r^x = r^l + \log(n)$. Fortunately, the number of links within the contention neighborhood that might cause an unblocking failure is limited due to the propagation properties of the wireless signal.

When transmitting a 1480 Byte data frame at 6 Mbps using the PHY parameters from IEEE 802.11g, the duration of the BTC covers 2.8% of the MAC transaction. Furthermore, we have conducted MATLAB simulations to estimate the effect of probe failures in typical settings. For example, consider the star topology in Fig. 5. The primary flow $F1$ competes with 8 one-hop flows within its contention neighborhood. In particular, the sender and receiver of the other flows are 2-hop and 3-hop neighbors, respectively, of the primary transmitter. All competing flows can transmit data concurrently. They only have to serialize their probes, if necessary. The primary flow, on the other hand, competes with flows $F2 - F4$ about resources for data transmissions. Each flow source transmits 1480 Byte frames at 6 Mbps at a target flow rate of 0.18, 0.64 and 0.82 for $F1$, $F2 - F4$ and $F5 - F9$, respectively. We observed a probing overhead of 0.114 at the primary sender, whereas the overhead for (successful) probing is about 0.018 and 0.022 for the remaining three and five flows, respectively. Thus, the surrounding flows consume at most 16% of the available medium around the primary transmitter via their BTCs. The primary flow uses only 0.5% of the medium for successful probing, and unsuccessful probing accounts for the remaining 10.9% of the probing overhead. Nevertheless, the primary flow is able to realize its flow rate even in the presence of multiple hidden transmitters.

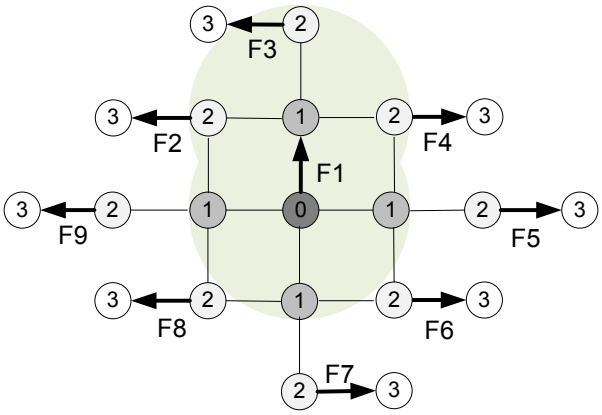


Figure 5: Scenario C: The flow $F1$ competes with 8 flows in its contention neighborhood, and it competes with $F2 - F4$ for data transmissions.

4.1.2. Collisions in CSMA/HBT

In the idealized CSMA model [48], collisions cannot occur due to the assumption about instantaneous propagation and arbitrarily small backoffs. However, these assumptions do not hold in general, so that the transmitters do not have a consistent view on the channel and CSMA-inherent collisions may arise. In particular, the collision problem boils down to the probability that an interferer accesses the channel during the vulnerable period [61].

A recently started transmission is vulnerable for physical and technological reasons. The propagation speed of wireless signals is physically limited by the speed of light. A node separation of 300 m leads to a propagation delay of no less than 1 μs, for example. The imperfectness of today's technology causes additional delays like the clear channel assessment (CCA) delay or the turn-around time to switch the radio from receiving to transmitting. On the other hand, the probability to start a concurrent transmission during the vulnerable period depends on the contention protocol and, in particular, on the average backoff within the collision domain, which further depends on the number of interferers in the neighborhood. In this section, we illustrate how the performance of CSMA/HBT is affected by both the length of the vulnerable period and the interferer activation probability.

The simulation topology consists of two links within a single collision domain, which both carry a single-hop flow. Each flow consists of 1480 Byte data frames, which are transmitted at 6 Mbps using the PHY parameters of IEEE 802.11g. The channel is modeled as AWGN without fading. The receiver uses the SINR-BER relationship to derive the packet error rates (PERs) and supports cumulative interference.

We consider two simulation parameters within our evaluation. We have varied the radio turnaround latency to account for the vulnerable period. Note that the radio

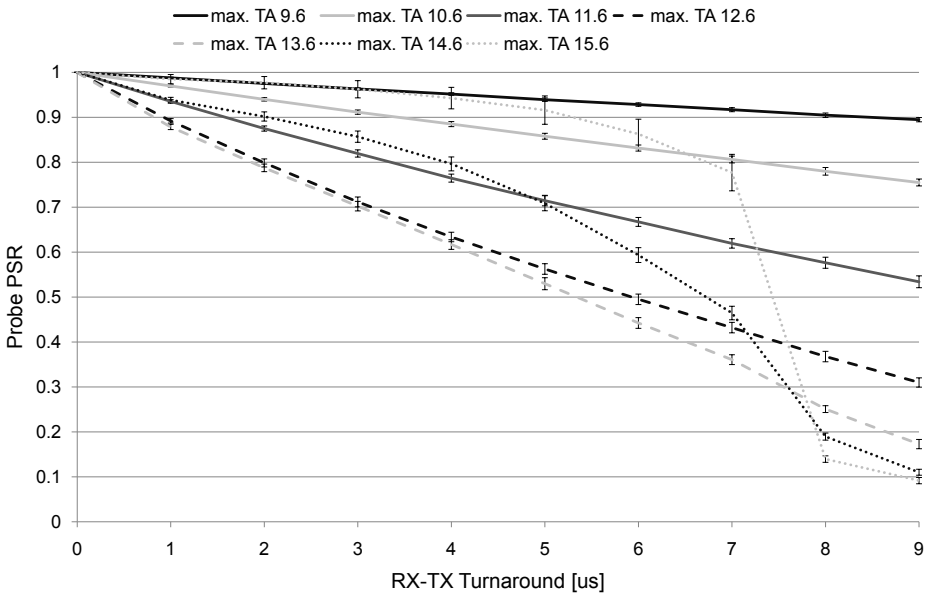


Figure 6: Two links within the same collision domain: Probe success rate in relation to the radio turnaround and the per-link TA ($\pm std.dev.$).

turnaround has the same effect on the vulnerable period as the propagation delays and the CCA duration. Furthermore, we have varied the average backoff via different TA working points for the links. Including additional links has basically the same effect as increasing the TA on each link. Thus, we will consider the case of two links in the following and vary the TA only. Each parameter setting has been averaged over 100 random seeds.

The probe success rate, i.e. the packet success probability (PSR) of the RTS probe and the following CTS frame, is shown in Fig. 6. The data transmissions are almost error-free in our simulation setup, so that the achieved goodput is proportional to the shown probe success rate. Considering the exponential distributed backoffs and a fixed vulnerable period, we would expect to see the complementary cumulative distribution function (CCDF) of the exponential distribution

$$F_c(x) = P(X > x) = \exp(-\lambda x) \tag{6}$$

with vulnerable period $x \geq 0$ and target TA λ. Table 1 shows the estimated λs according to (6) for a vulnerable period of 9 μs. The results for lower target TAs are in good match with our expectation. The deviation with higher TA targets is due to the following implementation detail. Remember that a link is blocked after a failed probe for the rest of the transmission opportunity (TXOP). Thus, both links involved in the probe failure would be unblocked with a separation of the vulnerable period

at most, since they have started their probes therein. Thus, subsequent probe failures are likely if both vulnerable period and target TA are large. However, timers on real hardware have only a limited precision. Hence, we have randomly jittered the unblocking time by $\pm 5\,\mu$s. Note that $-\log(10 \cdot 10^{-6}) \approx 11.5$. We suppose that the deviation from the exponential CCDF is mainly due to the jittering, since it becomes dominant for the contention round after a failed probe in the cases with high target TA and large radio turnarounds.

Max. TA	9.6	10.6	11.6	12.6	13.6	14.6	15.6
Est. TA λ	9.42	10.35	11.15	11.78	12.18	12.41	12.49

Table 1: Estimated TA λ according to (6) for a vulnerable period of 9 μs.

The results for two links cannot be directly extended to multiple contenders. The backoff is exponentially distributed and thus memoryless, but the blocking of links introduces memory. In particular, the two-link topology is a pessimistic estimate for the multi-link case, since both links are blocked after a failed probe. While they are blocked, none of them can generate throughput and the medium time is wasted. With additional contenders, there is a probability that some of them do not take part in the collision. Hence, they remain unblocked and might use the medium time for transmission, while the links are blocked that have been involved in the former collision.

In conclusion, we have reconfirmed that the physical and technological constraints have to be considered in the design of an actual CSMA protocol. The most important factors that influence CSMA collisions are the length of the vulnerable period and the conditional probability that an interferer meanwhile accesses the medium. The former factor is of physical and technological nature, whereas the latter mainly depends on the protocol design. With adaptive CSMA [48], a more efficient working point in terms of scheduling can generally be approached by increasing the per-link TAs. As we have seen, however, this comes at the expense of collisions in a non-idealized model. A tradeoff between the efficiency of the CSMA scheduling and the collision losses arises. Thus, it is necessary to control this tradeoff within a non-idealized world, and we design a protocol for this purpose in section 6.2.

4.1.3. Mutual Exclusion under Physical Interference

We have used the protocol model [125] in the derivation of the model and the protocol above. In contrast, the physical model [125] is considered to have a higher predictive value for the performance that can be encountered in the real world. In this section, we are going to illustrate to which extent the proposed protocol achieves mutual exclusion under a signal to interference and noise ratio (SINR) based reception model within AWGN and fading channels.

We consider a topology with two links within our analytic evaluation in Maple. We are interested in the performance of the *primary* link in relation to the *secondary* link, which acts as an interferer in our case. In particular, we use the PSR of the primary link as performance metric and evaluate the system with varying separation between

Parameter	Value
Radio frequency	2.4 GHz
Path loss	Log-distance, exponent 3, ref. dist. 1 m
Channel	AWGN & Rayleigh fading
Reception	Physical model (SINR-based)
PHY	IEEE 802.11ag
Signal bandwidth B	20 MHz
MCS	6 Mbps
SINR threshold γ_t^{dB}	5.4 dBm
TX power	19 dBm
Noise floor N	-92.965 dBm

Table 2: System Parameters for the Analytic Evaluation

the primary receiver and the secondary transmitter. The parameters of our evaluation are summarized in Table 2. Most importantly, we compare the system within two different channels models, which are the AWGN channel and the Rayleigh fading channel.

In AWGN channels, the PSR of a link is determined by the bit error rate (BER), which is i.i.d. in time. According to [37, 64], the BER P_b of uncoded binary phase shift keying (BPSK) can be estimated as

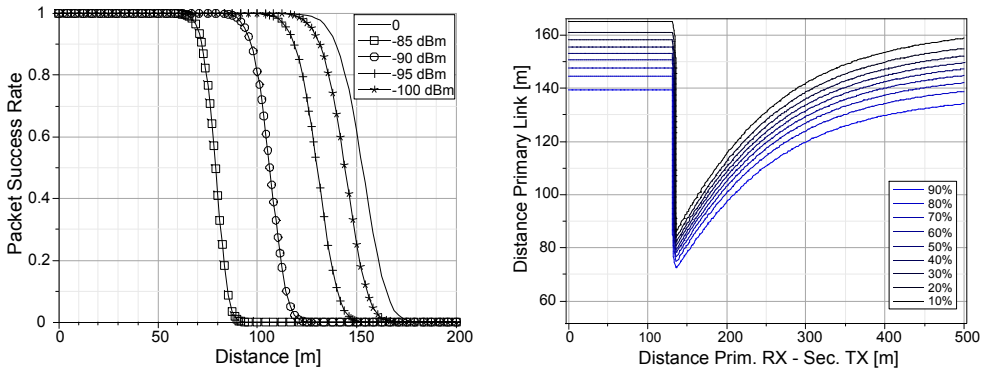
$$P_b(\gamma) = \mathrm{Q}\left(\sqrt{\frac{2E_b}{N_0}}\right) = \mathrm{Q}\left(\sqrt{\gamma\frac{2B}{R}}\right) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\gamma\frac{B}{R}}\right)$$

with energy per bit E_b, noise power spectral density N_0, signal bandwidth B, data rate R and SINR γ. In particular, we have set $B = 20\,\mathrm{MHz}$ and $R = 6\,\mathrm{Mbps}$. We assume uncoded operation. Thus, a frame is received correctly only if all bits are free from errors. Due to the i.i.d. nature of the BER in the AWGN channel, the PSR P_f becomes

$$P_f(\gamma) = (1 - P_b(\gamma))^n, \tag{7}$$

where n is the number of bytes per frame. Furthermore, we add the constant term $1.2 - \gamma_t^{dB}$ to the SINR γ^{dB} in logarithmic scale in order to shift the SINR-PSR curve, so that the PSR at the reception threshold γ_t^{dB} for a 1500 Byte frame is about 90%. The resulting PSR is shown in Fig. 7a for different levels of interference. On first sight, one might wonder why an interfering signal with a power of -100 dBm, which is far below the noise floor, causes such a degradation of the receiver performance. However, we have to become aware that for an already noise-limited reception process, even small amounts of additional interference lead to further SINR losses, and the PSR-SINR curve is rather steep in the AWGN.

If the interferer is within CS range of the primary receiver, it will be silenced by the



(a) Without CCA: PSR vs. distance according to (7).

(b) With CCA: PSR vs. distance according to (8) with $C = -85$ dBm.

Figure 7: Analytic PSR in the AWGN channel under interference.

BTs. Otherwise, we have to consider the contribution of the secondary transmitter within the SINR. Thus, the combined PSR P can be written as the following piecewise function.

$$P = P_f\left(\frac{P_{W1}}{N + P_{W2} \cdot \mathbf{1}_{P_{W2}<C}}\right) \tag{8}$$

In the equation above, P_W denotes the received power according to the log-distance path-loss model [37]. N is the noise power and C is the CCA power threshold. Furthermore, the indicator function $\mathbf{1}_z$ evaluates to 1 if z is true and is 0, otherwise. Fig. 7b shows the resulting PSR for a CCA threshold $C = -85$ dBm. If the interferer is sufficiently close to the primary receiver, then the CCA procedure is able to orthogonalize the transmissions. However, if the distance to the interferer is increased, the CCA will eventually become ineffective. With a CCA threshold of $C = -85$ dBm as in the figure, this point is reached at about 135 m. By lowering the CCA threshold, we can move the critical distance farther away. For example, the CCA threshold distance becomes 199 m (292 m, 428 m) with $C = -90$ dBm (-95 dBm, -100 dBm). Thereafter, we observe a severe PSR degradation especially for distances immediately beyond the CCA threshold, which can be mitigated to some extent by lowering the CCA threshold. Nevertheless, the protocol model has a limited reliability when additionally considering the physical impairments.

In slow fading channels, the instantaneous signal power is a random variable. It evolves slowly in comparison to the symbol duration, so that deep fades may corrupt multiple consecutive symbols. An *outage* occurs if the system is not able to recover from these error bursts, so that no reliable communication is possible during that time. In slow fading channels, the system performance is dominated by outages. The outage probability P_{out} captures the probability that the instantaneous signal power

drops below a threshold, so that the performance becomes unacceptable [37].

In the following, we use the outage probability as performance metric. In particular, we assume that the channel is slowly Rayleigh block-fading, i.e. the instantaneous signal power p_W is exponentially distributed according to the probability density function (PDF) (9) with average signal power P_W. Furthermore, the instantaneous fading realization captures the whole frame and it is i.i.d. between frames.

$$f(p_W; P_W) = \frac{1}{P_W} \exp\left(-\frac{p_W}{P_W}\right) \tag{9}$$

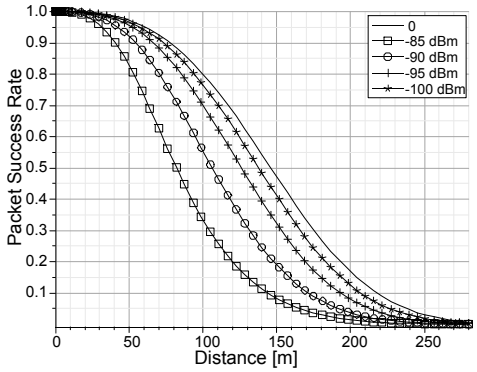
At the primary receiver, the instantaneous signal strength of both the primary signal as well as the interfering signal is random. In order to keep the analysis tractable, we assume independence between both random processes. Hence, the joint PDF is the product of the individual PDFs, i.e. $f(p_{W1}, p_{W2}; P_{W1}, P_{W2}) = f(p_{W1}; P_{W1}) \cdot f(p_{W2}; P_{W2})$. If the interference channel from the primary receiver to the secondary transmitter is above the CCA threshold, i.e. the instantaneous fading realization is in a way that the received signal strength is above the threshold, the CCA procedure is effective and we have to consider the outage probability of the primary link only. Otherwise, the additional interference power raises the required power level in the outage calculation of the primary link. In summary, we get the following outage probability for a Rayleigh fading channel with CCA and Rayleigh faded interference.

$$\begin{aligned} P_{out} = &\int_0^C \int_0^{(N+p_{W2})\gamma_t} f(p_{W1}, p_{W2}; P_{W1}, P_{W2})\, dp_{W1}\, dp_{W2} \\ &+ \int_C^\infty \int_0^{N\gamma_t} f(p_{W1}, p_{W2}; P_{W1}, P_{W2})\, dp_{W1}\, dp_{W2} \end{aligned} \tag{10}$$

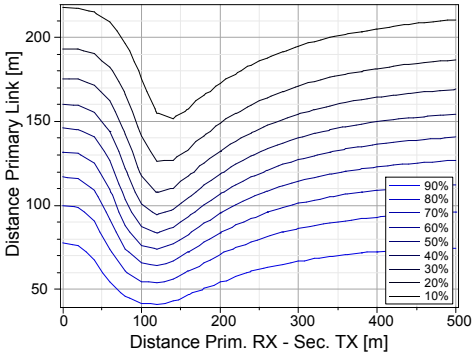
Note that (10) contains perfect and absent CCA as special cases with $C = 0$ and $C = \infty$, respectively. The latter special case is shown in Fig. 8a. Fig. 8 shows the resulting PSR for different CCA thresholds. Interestingly, the degradation of the PSR is slightly less severe with Rayleigh fading, if we compare Fig. 8c and Fig. 7b, for example. Nevertheless, a degradation of the PSR cannot be prevented for some link constellations with reasonable CCA thresholds.

In conclusion, we have illustrated the limitations of the protocol model within a physical environment, which apply to CSMA/HBT as well. The accumulative nature of the interference power gives rise to hidden nodes to a certain extent. However, other physical effects like shadowing or multi-slope propagation, which has been empirically observed, may render the problem less severe [37, 120]. The problem may be mitigated using larger exclusion regions via a higher transmission power for the BTs, which comes at the expense of lower spatial reuse.

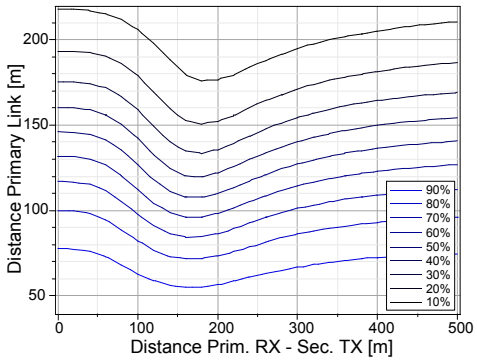
We have decided to ignore the problem in the further design. In this way, the medium access becomes be a mixture of CSMA and ALOHA to a varying degree,



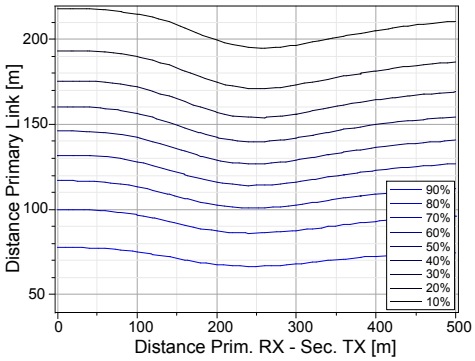
(a) Without CCA: $C = \infty$



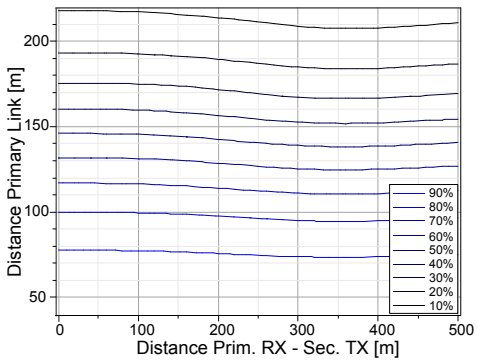
(b) With CCA: $C = -80$ dBm.



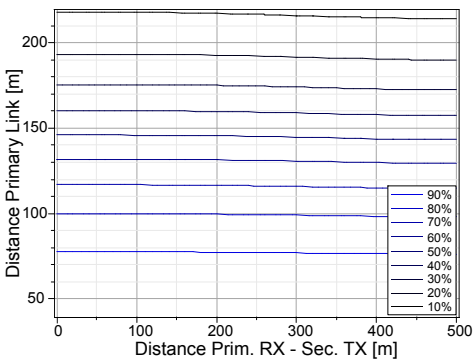
(c) With CCA: $C = -85$ dBm.



(d) With CCA: $C = -90$ dBm.



(e) With CCA: $C = -95$ dBm.



(f) With CCA: $C = -100$ dBm.

Figure 8: Analytic PSR in the Rayleigh fading channel under interference according to (7).

and the PSR of a link becomes dependent on the behavior of hidden nodes. As we will see in the next section, we are using the PSRs in the optimization to characterize the capacity of a link. Thus, the throughput along a link depends not only on its activation rate. Instead, we would have to optimize over the PSRs, as well, which leads most likely to a non-convex problem and we cannot approach the optimum via greedy algorithms.

In the following, we assume that the PSR is constant and independent from hidden nodes. As we have illustrated above, the assumption does not hold in some cases, and the PSR will degrade. Nevertheless, the nodes conduct PSR measurements, which will capture the effect of hidden nodes, so that they can react to severe degradations. However, we have to face the risk that the system does not approach the most efficient working point or even worse, that the system is not stable. Looking ahead to our evaluation in section 7 and section 8, we have observed that the risk is low when using a reasonable CCA threshold.

5. Optimization Framework for Opportunistic Routing with Adaptive CSMA/CA

In this section, we derive a cross-layer algorithm for opportunistic routing (OR) in WMNs based on adaptive carrier sense multiple access (CSMA). Before we go into details about the optimization problem, we introduce the concepts of the multi-user gain and the anycast goodput capacity region first.

In typical WMNs, *unreliable* links of intermediate quality are common [1] and the frame errors are *conditionally independent* across different wireless channels [141]. Thus, the anycast communication primitive may achieve a *multi-user gain*, i.e. using multiple receivers with anycast within an error-prone environment yields a higher PSR in comparison to any of the underlying unicast links. In particular, let $p_{i,j}$ denote the PSR of the unicast link $(i, j),\ i, j \in N$. The PSR of the anycast link $(i, J),\ i \in N, J \in \mathcal{P}(N)$[2] is defined as follows.

$$p_{i,J} = 1 - \prod_{j \in J} (1 - p_{i,j})$$

In particular, we assume that the frame error process is stationary and i.i.d. in time, which resutls in a memoryless channel. A memoryless channel is the best case for OR since it has the highest uncertainty about the transmission success. The better the channel can be predicted in advance, the lower the benefits of OR become in comparison to opportunistic scheduling.

In the following, we distinguish between throughput q and goodput x. The former captures the consumed medium time. The latter additionally accounts for bit-rates (modulation and coding schemes) and channel errors and thus refers to the rate, at which innovative information is transported. We select the bit-rate R per anycast link

[2] $\mathcal{P}(N)$ refers to the power set of set of nodes N.

that offers the highest goodput. In particular, we assume that the communication system provides a finite set of bit-rates per link, and the PSR per link and per bit-rate can be estimated. We use the qualification $p_{i,J}(R)$ to refer to the PSR of link (i, J) when using bit-rate R. The bitrate $R_{i,J}$ of link (i, J) is selected as follows, breaking ties arbitrarily.

$$R_{i,J} = \arg\max_{R} p_{i,J}(R) \cdot R$$

In addition, we use the shorthand notation $p_{i,J}^{i,K} = p_{i,J}(R_{i,K})$ to refer to the PSR of link (i, J) when using the selected bit-rate of link (i, K).

In a multi-user system, the capacity region is the union of achievable rate vectors under all multi-user transmission strategies [37]. The *capacity region of anycast goodput* (in short: anycast goodput region) has to consider all multi-user transmission strategies. The goodput $x_{i,j}$ between node i and j can be realized via any hyperlink $q_{i,J}$ with transmitter i and receiver set $J \ni j$ containing node j, which leads to the inner summation in (11). Furthermore, the anycast goodput region has to account for *non-innovative* throughput. If multiple candidates have successfully received a packet, then the packet is innovative only for the further forwarder. It is non-innovative and thus does not generate goodput for all other receivers, since relaying the packet would result in duplication and hence a waste of wireless resources. Thus, we have to introduce one constraint per anycast link (i, J) in (11).

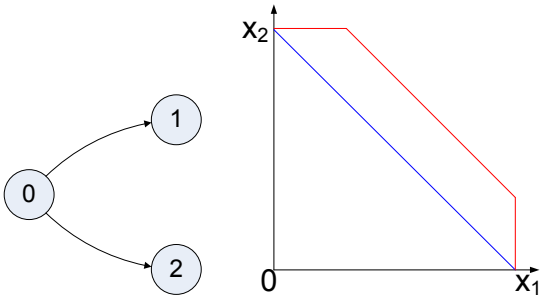
$$\sum_{j \in J} x_{i,j} \leq \sum_{L \in \mathcal{P}(J)} \sum_{K \in \mathcal{P}(N \setminus J)} p_{i,L}^{i,L \cup K} \cdot R_{i,L \cup K} \cdot q_{i,L \cup K}, \quad \forall i \in N, \forall J \in \mathcal{P}(N) \tag{11}$$

In the general case, the number of necessary constraints to describe the anycast capacity region is exponential in the number of network nodes. However, we can reduce the amount of constraints to a polynomial number if we upper bound the maximum number of (simultaneous usable) candidates (cf. Appendix C). Furthermore, the number of active constraints is upper bounded by the number of candidates (cf. Appendix D). Due to the multi-user gain, the anycast goodput region is a superset of the unicast capacity region.

For example, consider the topology in Fig. 9a. Transmitter 0 may use any non-empty subset of the node set $\{1, 2\}$ as candidates. The following constraints describe the anycast goodput region.

$$\begin{aligned}
x_{0,1} &\leq p_{0,1}^{0,1} \cdot R_{0,1} \cdot q_{0,1} &&+ p_{0,1}^{0,\{1,2\}} \cdot R_{0,\{1,2\}} \cdot q_{0,\{1,2\}} \\
x_{0,2} &\leq && p_{0,2}^{0,2} \cdot R_{0,2} \cdot q_{0,2} + p_{0,2}^{0,\{1,2\}} \cdot R_{0,\{1,2\}} \cdot q_{0,\{1,2\}} \\
x_{0,1} + x_{0,2} &\leq p_{0,1}^{0,1} \cdot R_{0,1} \cdot q_{0,1} &&+ p_{0,2}^{0,2} \cdot R_{0,2} \cdot q_{0,2} + p_{0,\{1,2\}}^{0,\{1,2\}} \cdot R_{0,\{1,2\}} \cdot q_{0,\{1,2\}}
\end{aligned}$$

Fig. 9b illustrates the bounds of the capacity region for time sharing (blue) and anycast (red). The anycast link is not used with time sharing ($q_{0,\{1,2\}} = 0$), so that one link can be traded for the other in terms of goodput. With error-prone links, the anycast



(a) Topology.

(b) Anycast goodput region with time sharing (red) and anycast (blue).

Figure 9: Illustration of the anycast goodput region in error-prone environments.

goodput region is a proper superset of the time share region. However, the multi-user gain changes the shape of the region. In particular, even if the goodput of one link is at its maximum, we can achieve goodput on the other link. Compared to time sharing, the additional goodput is for free since the first link does not have to make sacrifices.

5.1. Optimization Problem

Given the set of nodes N, the link conflict graph (LCG) and the CSMA Markov chain, the PSRs p, the set of flows $f \in F$ and the concave and increasing utility function U, the optimization problem of congestion control, opportunistic routing and CSMA scheduling in WMNs is as follows.

$$\max_{y,x,q,u} -\sum_l u_l \log(u_l) + K \sum_f U(y^f) \tag{12}$$

$$\text{s.t.} \sum_j x^f_{j,i} + y^f \mathbf{1}_{i=\sigma(f)} \leq \sum_j x^f_{i,j}, \quad \forall f, i \in N \backslash \delta(f) \tag{13}$$

$$\sum_{j \in J} x^f_{i,j} \leq \sum_{L \in \mathcal{P}(J)} \sum_{K \in \mathcal{P}(N \backslash J)} p^{i,L \cup K}_{i,L} \cdot R_{i,L \cup K} \cdot q^f_{i,L \cup K}, \quad \forall f, \forall i \in N, \forall J \in \mathcal{P}(N) \tag{14}$$

$$0 \leq x^f_{i,j} \quad \forall f, i, j \tag{15}$$

$$0 \leq q^f_{i,J} \quad \forall f, i, J \tag{16}$$

$$\sum_f q^f_{i,J} \leq \sum_l \delta^2_{l,(i,J)} u_l, \quad \forall i, J \tag{17}$$

$$\sum_l u_l = 1, \quad 0 \leq u_l \; \forall l \tag{18}$$

The notation $\sigma(f)$ and $\delta(f)$ denotes the source and destination of flow f, respectively. The indicator function $\mathbf{1}_z$ evaluates to 1 if z is true and is 0, otherwise. Our objective is to maximize the sum-utility of all flows. The flow rates are constrained by the flow conservation, anycast goodput region and non-negativity constraints (13), (14), (15) and (16). As shown in Appendix B, adaptive CSMA introduces the constraints (17), (18) and an entropy term into the objective function.

Based on the optimization problem above, we have derived Algorithm 1 in the following way. At first, let us relax the constraints (13), (14) and (15) and get the Lagrangian (19). The problem separates vertically in its primal variables, and the Lagrangian will get separated into $\mathcal{L} = \mathcal{L}_y + \mathcal{L}_x + \mathcal{L}_{q,u}$ in the following.

$$\begin{aligned}\mathcal{L}(y,x,q,u;\alpha,\beta,\gamma) = &-\sum_l u_l \log(u_l) + K\sum_f U(y^f) \\ &+ \sum_{f,i} \alpha_i^f \left(\sum_j x_{i,j}^f - \sum_j x_{j,i}^f - y^f \mathbf{1}_{i=\sigma(f)} \right) \\ &+ \sum_{f,i,J} \beta_{i,J}^f \left(\sum_{L\in\mathcal{P}(J)} \sum_{K\in\mathcal{P}(N\setminus J)} p_{i,L}^{i,L\cup K} \cdot R_{i,L\cup K} \cdot q_{i,L\cup K}^f - \sum_{j\in J} x_{i,j}^f \right) \\ &+ \sum_{f,i,j} \gamma_{i,j}^f x_{i,j}^f \end{aligned} \tag{19}$$

The newly introduced $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are dual variables for the constraints (13), (14) and (15). Intuitively, a physical meaning may be assigned to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, depending on their function in routing and forwarding. For incoming traffic, routing is the operation of allocating the packets to outgoing links. On the other hand, forwarding eventually transports traffic to the next hop(s). A simple gradient algorithm to solve for $\arg\min_\alpha \mathcal{L}$ is

$$\alpha_i(t+1) = \left[\alpha_i(t) - s_\alpha(t) \left(\sum_j (x'_{i,j}(t) - x'_{j,i}(t)) - y'(t) 1_{i=\sigma(f)} \right) \right]_+ \tag{20}$$

where $s_\alpha(t)$ is some small positive step size, $[.]_+$ denotes the projection onto the non-negative orthant, and x', y' are the average arrival and service rates, respectively. Taking a deeper look at the formulation, we notice that the dynamics of α_i and the queue evolution at node i are very similar [48]. If the start conditions for the queue length and α are set to 0, then α is proportional to the packet queue. The value of a queued packet with respect to α_i is determined by the step size $s_\alpha(t)$. Similar to Radunovic *et al.* [98], we associate one unit of α_i with a *node credit (C)*. Thus, α_i can be seen as a virtual queue of node credits.

Algorithm 1 Cross-Layer Opportunistic CSMA.

Require: $t_{up} > 0$ ▷ Update interval
Require: $s > 0$ ▷ Step size

procedure UPDATETA(Link (i, J))
  $f_{i,J} \leftarrow \arg\max_f K_f \sum_{L \in \mathcal{P}(J)} p_{i,L}^{i,J} \sum_{K \in \mathcal{P}(N_i \setminus J)} \beta_{i,L \cup K}^{f}$
  $r_{i,J} \leftarrow R_{i,J} K_{f_{i,J}} \sum_{L \in \mathcal{P}(J)} p_{i,L}^{i,J} \sum_{K \in \mathcal{P}(N_i \setminus J)} \beta_{i,L \cup K}^{f_{i,J}}$
  SETAVGBACKOFF($\exp r_{i,J}$) ▷ Adapt CSMA
  SLEEP(t_{up})
  UPDATETA(i, J)
end procedure

procedure UPDATEC(Node i, Flow f)
  $\beta_{i,J}^{f} \leftarrow 0, \forall J$ ▷ Zero otherwise
  $J = [1, \ldots, j] \leftarrow \text{SORT}(C_{N_i}^{f})$ ▷ Least cost neighbors
  $C_i^f \leftarrow \min_{j \in J} C_j^f + TC_i^f$
  for all $j \in J : C_j^f < C_i^f$ **do** ▷ Allocate transport credits
    $TC_{i,j}^{f} = \beta_{i,\{1,\ldots,j\}}^{f} \leftarrow \min(C_i^f, C_{j+1}^f) - C_j^f$
  end for
  SLEEP(t_{up})
  UPDATEC(i)
end procedure

procedure UPDATEY(Flow f) ▷ At source of f
  $y^f \leftarrow 1/C_{s(f)}^{f}$
  SLEEP(t_{up})
  UPDATEY(f)
end procedure

Require: $L \subseteq J$ ▷ Receiver set
procedure PACKETFORWARDED(Link (i, J), L)
  $f \leftarrow f_{i,J}, j = \arg\min_{j \in L} C_j^f$
  $TC_i^f \leftarrow TC_i^f - s$
  $TC_j^f \leftarrow TC_j^f + s$
end procedure

procedure PACKETARRIVED(Flow f) ▷ Exogenous arrival
  $TC_{s(f)}^{f} \leftarrow TC_{s(f)}^{f} + s$
  $t \leftarrow \text{EXPRND}(1/y^f)$
  SLEEP(t)
  PACKETARRIVED(f)
end procedure

Using similar argumentation, the dynamics of β is given by

$$\beta_{i,J}(t+1) = \left[\beta_{i,J}(t) - s_\beta(t) \left(\sum_{L \in \mathcal{P}(J)} \sum_{K \in \mathcal{P}(N \setminus J)} p_{i,L}^{i,L \cup K} R_{i,L \cup K} q'_{i,L \cup K}(t) - \sum_{j \in J} x'_{i,j}(t) \right) \right]_+ \tag{21}$$

In particular, β is the difference between supply and demand of link capacity. Intuitively, if the supply exceeds the demand, it will decrease to zero. In the opposite case, the capacity demand cannot be satisfied and β increases. Thus, it can be interpreted as differential back-pressure, i.e. the number of credits that wait to be served on the link under consideration. In the following, we use the term *transmission credit (TC)* with β according to the terminology of Radunovic *et al.* [98].

In Algorithm 1, both node credits (Cs) and transmission credits (TCs) are used relative to the K in (12). Later in section 6, we will highlight the benefits of the relative notation. The dynamics of TC can be found on line 23 et seq. of Algorithm 1 and the dynamics of C is derived from the TCs as described in section 5.3.

5.2. Congestion Control

Separating the flow variables $\boldsymbol{y}$ in the Lagrangian (19), we get

$$\mathcal{L}_y(y; \alpha) = \sum_f \left(KU(y^f) - \alpha^f_{\sigma(f)} y^f \right).$$

The congestion control problem is to find the flow variables $\boldsymbol{y} = \arg\max \mathcal{L}_y$ that maximize the Lagrangian. For proportional fairness, in particular, the utility function $U(y) = \log(y)$ applies. Solving $\partial \mathcal{L}_y / \partial y^f = 0$ provides us the following solution.

$$y^f = K / \alpha^f_{\sigma(f)} \tag{22}$$

Hence, knowing its credits α, each source is able solve the congestion control problem locally. Remember that C is used relative to K at line 19 of Algorithm 1.

5.3. Opportunistic Routing

After some rearranging, we get the partial routing Lagrangian (23).

$$\mathcal{L}_x(x; \alpha, \beta, \gamma) = \sum_{f,i,j} x^f_{i,j} \left(\alpha^f_i - \alpha^f_j - \sum_{J \ni j} \beta^f_{i,J} + \gamma^f_{i,j} \right) \tag{23}$$

Hence, $x^f_{i,j}$ can be interpreted as a dual variable for the Lagrange dual problem (24). In particular, we have used the relation $0 \leq \gamma^f_{i,j}$ to derive the inequality constraints of

the feasibility problem (24).

$$\begin{aligned} &\max 0 \\ &\text{s.t. } 0 \leq \alpha_i^f,\ 0 \leq \beta_{i,J}^f \ \ \forall f,i,J \\ &\quad 0 \geq \alpha_i^f - \alpha_j^f - \sum_{J \ni j} \beta_{i,J}^f \ \ \forall f,i,j \end{aligned} \tag{24}$$

At the Karush-Kuhn-Tucker (KKT) points, it holds $\alpha_i^f - \alpha_j^f = \sum_{J \ni j} \beta_{i,J}^f$ if $x_{i,j}^f > 0$ due to complementary slackness, i.e. the credit difference between neighboring nodes can be interpreted as back-pressure. In the next paragraph, we describe how the opportunistic routing layer determines the back-pressure and hence solves the arising system of linear equations.

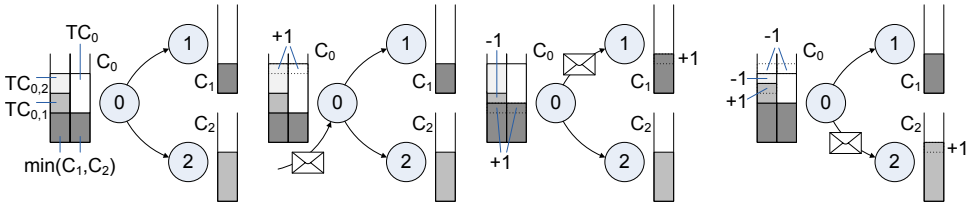
Let the cumulated transmission credits TC_i^f be the number of *queued packets* of flow f at node i. Routing solves the problem of how the queued packets (or equivalently, the cumulated transmission credits) are allocated to the outgoing links in a way that it holds $TC_i^f = \sum_j TC_{i,j}^f$, where $TC_{i,j}^f$ is the number of queued packets per flow f and link (i,j). Furthermore, let the node credits C_j^f be the *accumulated costs* to route traffic to the destination of flow f via node j, which is zero at the destination of the flow. Let N be the set of neighbors of node i. Then node i solves the routing problem and determines its costs in terms of node credits as follows.

$$TC_{i,j}^f = \begin{cases} C_{j+1}^f - C_j^f, & C_0 \leq \ldots \leq C_j^f \leq C_{j+1}^f \leq \ldots < C_i^f \\ 0, & C_j^f \geq C_i^f \end{cases} \tag{25}$$

$$C_i^f = \min_j C_j^f + TC_i^f = \min_j C_j^f + \sum_{j \in N} TC_{i,j} \tag{26}$$

The approach is straightforward, as shown in line 8 et seq. of Algorithm 1. For flow f, node i sorts its neighbors according to ascending node credits, so that it holds $C_0^f \leq \ldots \leq C_j^f$ WLOG. Node i starts with allocating transmission credits to the least cost neighbor 0 according to (25) until the cumulated credits $C_0^f + TC_{i,0}^f$ reach the costs of the neighbor having the next higher credits, which is node 1 in our case. The sender proceeds in the same way with node 1 and afterwards with the following neighbors. The allocation process terminates when all transmission credits have been allocated, i.e. it holds $TC_i^f = \sum_j TC_{i,j}^f$. The sender determines its costs C_i^f within the routing of flow f according to (26).

After the routing process, the costs of the transmitter C_i^f in terms of node credits are higher than the costs of any neighbor with transmission credits assigned. Furthermore, all neighbors that will not carry traffic coming from node i do not have lower costs. However, if two neighbors j and k have equal costs and node i assigns transmission credits to them, then only one queue will receive credits, say $TC_{i,j} > 0$, and the



(a) Estimation of the TCs. (b) Packet arrived at 0. (c) Packet forwarded from 0 to 1. (d) Packet forwarded from 0 to 2.

Figure 10: Estimation and evolution of the node and transmission credits on topology Fig. 9a.

other remains empty by design ($TC_{i,k} = 0$). Note that the proposed approach is completely distributed, since it involves information from neighboring nodes only, which can be obtained via periodic link probing or via piggy-backing during the MAC transaction.

The presented approach is based on the idea to reduce the queueing efforts from quadratic to linear complexity [12, 99], which is also called *floating queues* [83]. A formal justification can be found in Appendix E. Instead, we will illustrate the routing decisions using the topology in Fig. 10 in the following. The transmitter 0 starts to allocate TCs to the least cost neighbor, which is node 1 in Fig. 10a. According to (26), the resulting credits of node 0 are $C_0 = C_1 + TC_0$, as shown in the picture. In this example, node 2 is included in the routing decision since it has lower costs ($C_2 < C_0$), i.e. node 0 assigns TCs to both neighbors according to (25). We have $TC_{0,1} = \beta_{0,1} = C_2 - C_1$ and $TC_{0,2} = \beta_{0,\{1,2\}} = C_0 - C_2$ and all other βs are zero.

When a packet arrives at node 0 either exogenously or from upstream nodes as shown in Fig. 10b, we increase the cumulated transmission credits TC_0 at node 0. This way, the credits C_0 of node 0 increase as well as $TC_{0,2}$, i.e. the transmitter allocates the newly arrived credit to neighbor 2. After a packet has been forwarded to a receiver set J, we determine the receiver with the least costs $j = \arg\min_{j \in J} C_j^f$ from J and transfer one credit from 0 to j. In particular, the credit is transferred from TC_0 to TC_j (cf. line 23 et seq. of Algorithm 1), so that C_j increases. If the receiver j is the least cost neighbor as shown in Fig. 10c, then the node credits of the transmitter C_0 remain unchanged since C_1 increases in the same way as TC_0 decreases. Furthermore, the transmitter allocates fewer TCs to neighbor 1 due to its higher costs ($TC_{0,1} = C_2 - C_1$). On the other hand, if the receiver j is not the least cost candidate, the credit dynamics is slightly different as shown in Fig. 10d. In this case, the node credits of the receiver C_2 increase as before. Nevertheless, the node credits of the transmitter C_0 are reduced at the same time. This means that $TC_{0,1}$ increases, whereas the transmission credits of the receiver $TC_{0,2}$ decrease by two units.

5.4. CSMA Scheduling

Separating the variables $\boldsymbol{u}$ and $\boldsymbol{q}$ in the Lagrangian, we get

$$\mathcal{L}_{q,u}(\boldsymbol{q},\boldsymbol{u};\boldsymbol{\beta}) = -\sum_l u_l \log(u_l) + \sum_{f,i,J} R_{i,J} \cdot q_{i,J}^f \left(\sum_{L\in\mathcal{P}(J)} p_{i,L}^{i,J} \sum_{K\in\mathcal{P}(N\setminus J)} \beta_{i,L\cup K}^f \right)$$

subject to constraint (17), (18) and (16). We fix $\boldsymbol{u}$ and $\boldsymbol{\beta}$ in $\mathcal{L}_{q,u}$ and solve for $\boldsymbol{q}$, which corresponds to the following optimization problem.

$$\begin{aligned} \max_q & \sum_{f,i,J} R_{i,J} q_{i,J}^f \left(\sum_{L\in\mathcal{P}(J)} p_{i,L}^{i,J} \sum_{K\in\mathcal{P}(N\setminus J)} \beta_{i,L\cup K}^f \right) \\ s.t.\ & 0 \le q_{i,J}^f \quad \forall f,i,J \\ & \sum_f q_{i,J}^f \le \sum_l \delta_{l,(i,J)}^2 u_l, \quad \forall i,J \end{aligned}$$

The solution is to schedule the flow $f_{i,J}$ which each hyperlink (i,J) that maximizes

$$r_{i,J} = R_{i,J} \max_f \left(\sum_{L\in\mathcal{P}(J)} p_{i,L}^{i,J} \sum_{K\in\mathcal{P}(N\setminus J)} \beta_{i,L\cup K}^f \right) \tag{27}$$

The remaining $q_{i,J}^f$ are set to zero. Thus, $\mathcal{L}_{q,u}$ simplifies to

$$\mathcal{L}_{q,u}(\boldsymbol{u};\boldsymbol{\beta}) = -\sum_l u_l \log(u_l) + \sum_{i,J} r_{i,J} \left(\sum_l \delta_{l,(i,J)}^2 u_l \right),$$

subject to the constraint (18), i.e. that $\boldsymbol{u}$ is a distribution. Thus, the maximum is achieved if $\boldsymbol{u}$ is the stationary distribution of the CSMA Markov Chain with TA $r_{i,J}$. Note that node i is able to solve the medium access problem for each of its outgoing links with local information only, as shown on line 1 et seq. of Algorithm 1.

6. Design of a Cross-Layer Protocol

The derived cross-layer algorithm in section 5 has several drawbacks in a straightforward realization due to the formulation of the optimization problem and the underlying tradeoffs. In particular, the parameter K in (12) determines the working point of the system in terms of efficiency. However, increasing the efficiency via higher Ks results in higher TAs and thus smaller backoffs. As we have illustrated in section 4.1, the backoff cannot be arbitrarily decreased for physical and technological reasons. In the following, we design an adaptation mechanism for K that maximizes the ef-

ficiency while operating the TAs in a feasible regime and maintaining proportional fairness among competing flows.

The optimization problem (12) considers neither convergence nor end-to-end delay. Theoretical results suggest that adaptive CSMA sacrifices delay in order to achieve high throughput with low complexity [127]. From a practical point of view, the tradeoff between throughput and delay should be controlled in order to provide a satisfactory service to the end user. In particular, a newly started flow should get end-to-end service as soon as possible, whereas a less efficient working point is acceptable. While the flow persists, the system should incrementally increase the throughput efficiency until a working point is reached that offers an appropriate tradeoff between throughput and delay.

The vanilla approach to back-pressure routing is topology-blind, i.e. it considers neighboring information only. Thus, all possible paths have to be taken into account, even if they carry marginal or no traffic at all. However, each additional path increases the optimality gap with UO-CSMA (cf. section 2.2). From the practical point of view, the traffic should be concentrated to the important paths only. For that purpose, we design a route pre-selection that uses topology information from WMN routing metrics in the following.

6.1. Packet Transition and Forwarder Selection

The area of forwarder selection policies on the MAC layer is well explored, as already pointed out in section 2.3. Our intention is not to design another selection procedure. Instead, we will decide for an existing one based on the requirements of the cross-layer algorithm presented in the previous section. Within our cross-layer algorithm, the forwarder having the fewest credits should be selected (cf. Algorithm 1). The influence of the forwarder selection procedure on the performance results should be as low as possible. Thus, reliability and robustness against coordination errors are the main requirements in our design. On the other hand, the delay of the selection policy is not a primary issue as long as the back-pressure related delays dominate.

In our cross-layer design and related ones [35, 85, 100], the transmitter should be aware of the selected forwarder. The successful transmission of a packet causes credit transitions as illustrated in section 5.3. Without the knowledge of the actual forwarder at the transmitter, the routing decision has to be delayed until the next credit update from downstream nodes. Delayed feedback does not necessarily result in instability. Nevertheless, it may slow down the convergence or raise the variability.

From the point of view of the presented cross-layer algorithm, it is not necessary that the transmitter gets feedback from all candidates that have successfully received the data frame. However, the perceived link qualities for forwarding data traffic may differ from the link probe results due to the hidden node problem (cf. section 4.1.3). Thus, the forwarder selection procedure should provide reception feedback from all candidates to improve the PSR estimation at the transmitter.

Considering all requirements above, we have decided for the robust ACK policy according to Zhong *et al.* [136]. In particular, the underlying slotted ACK provides the source node with the necessary PSR feedback. The approach is more robust against

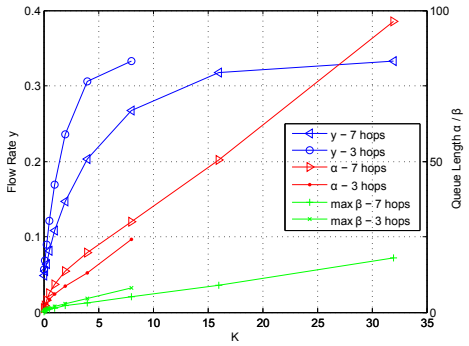


Figure 11: Flow rate and queue length vs. the efficiency factor K for a chain of 3 and 7 links. The results are obtained from numerically solving (12) under the idealized CSMA model with error-free links and a minimal TA of -3.

communication errors than the vanilla slotted ACK since it relies on transmitter-based forwarder selection instead of self-selection. The additional delay that incurs with the robust ACK approach is acceptable in our case: It coarsely corresponds to an increase in queue length by one.

6.2. Efficiency-Collisions Tradeoff in Multi-Hop Routing

The throughput efficiency of the algorithm in section 5 depends on the selection of the tradeoff factor K in the objective function (12) and in Algorithm 1. Using a larger K, the source of the flow pumps more aggressively packets into the network, resulting in longer queues and higher TAs. Thus, the efficiency of the protocol approaches the optimum at the expense of longer credit queues $\boldsymbol{\alpha}$. With multi-hop routing, there are packet queues on all intermediate nodes. From section 5 we know that the back-pressure is expressed in terms of TCs. Hence, reducing K at the source results in a smaller node credit queue, so that there are fewer TCs available along the route and the TAs decrease. However, this means that the system approaches a less efficient working point having a lower flow rate. To verify our explanation above, we have solved the optimization problem (12) for the idealized CSMA model and a logarithmic utility function using an optimizer in Maple for a chain of 3 and 7 links.[3] The resulting flow rates and queue lengths are plotted in Fig. 11. Both flows approach the optimal flow rate 0.333 with increasing K. However, the longer flow is less efficient at lower Ks, e.g. at $K = 4$ the flow rates is 0.306 and 0.203 for the short and the long flow, respectively. Thus, a fixed K will not meet our design requirements. If it is too high, then the TA of a short flow may exceed the technological limits. On the other hand, the throughput and efficiency of long flows will be low with small Ks.

The tradeoff can also be seen in the utility gap estimation in [47]. Consider a very

[3]More precisely, we added the term $-3\sum q$ to the objective function in order to realize a minimal TA below zero.

large network with a (single-path) multi-hop flow of n hops. Since all nodes, which are not on the route, do not participate in the forwarding process, we can ignore them and focus on the remaining n nodes on the route. The final destination can be ignored, too, since it is not involved in the forwarding process. The difference of the achievable and the optimal utility is bounded by $n \log(2)/K$ [48]. If we increase the length of the route, the bound on the optimality gap will linearly increase with the length of the route.

From the practical point of view, the TAs are upper bounded in real systems (cf. section 4.1). On the one hand, there are physical and technological reasons like the speed of light and the duration of the CCA procedure, which lead to inconsistencies within the channel state between different nodes at the affected timescale. On the other hand, the collision probability of CSMA crucially depends on the number of contenders within the same collision domain. Hence, an upper limit for the TA on each link should be established to accommodate for the above mentioned impairments. In addition, a more advanced approach may try to adapt to the instantaneous conditions within the collision domains in order to prevent a collision breakdown in dense networks.

In previous work, it is assumed that the requested input rates are supported within the reduced capacity region of TA-limited UO-CSMA [51, 78]. In particular, if the requested input rates are outside of the limited UO-CSMA capacity region for a particular TA, but within it for a higher TA, then we have observed in our simulations that the not supported back-pressure accumulates at the bottleneck links. Hence, these links become unable to control their queue length via adaptive CSMA.

The determination of a working point for the considered tradeoff is a matter of design, of course. However, we argue that fixed Ks lead to unexpected working points, which do not efficiently utilize precious resources in scenarios with longer routes and few users. A more intuitive working point would be to provide proportional fairness and the highest throughput efficiency while ensuring that the TAs do not exceed their technological limits. In the following, we propose a combination of *intra-flow* and *inter-flow adaptation* of the efficiency parameters to achieve both objectives.

6.2.1. Achieving High Efficiency with Limited TAs

In order to limit the TAs per link on the route while maintaining high throughput efficiency, we perform *intra-flow adaptation* of the efficiency parameter K. In particular, we embed an end-to-end feedback loop into the forwarding process. Our objective is to upper limit the aggregated TA per flow on each node. From the previous section we know that the aggregated TA per node and flow is non-decreasing in the enqueued TCs. Each forwarding node determines its aggregated TA corresponding to the considered flow, and feeds back this information in the upstream direction (e.g. via the ACK or piggy-backed via data packets). Meanwhile, it may clip the excess in order to enforce the technological limits on the TAs. The upstream node relays the maximum out of its own and the feedback received from downstream neighbors. Eventually, the maximum aggregated TA along all paths arrives at the source of the flow. The source takes the feedback as input, and adapts the efficiency parameter K of the flow

accordingly.

There are multiple alternatives for adapting K. For example, the source may use a proportional-integral-derivative (PID) controller to update K based on the difference between the feedback and the *target TA*, i.e. the target value of the aggregated TA per node. However, the PID controller is difficult to parameterize because of the varying response time of the dynamic system. Instead, we are using a model predictive approach. If we consider Fig. 11, it is evident that the queue length at the source increases with K. In the low efficiency regime, the increase in queue length is steep. As an interpretation, it can be seen as overcoming the entropy of the CSMA Markov chain. Afterwards, the slope decreases and becomes almost constant in the high efficiency regime, i.e. the queue becomes proportional to K. Considering the congestion control (22), we observe that the relation between K and C approaches proportionality the closer the flow rate gets to its optimum. As simplification, we assume that the flow already operates in the high efficiency regime and proportionality between the credit queue C and the parameter K holds. We estimate the efficiency parameter K that meets the target TA at the bottleneck link(s), i.e. the links exposing the highest TA along the route. During the update process, we filter our estimate of K using an exponentially weighted moving average (EWMA) to smooth the convergence.

For proportional fairness, the estimation of the optimal K is as follows. In the high efficiency regime, the flow rate y is (near) maximal and (almost) fixed so that K is proportional to the queue length C at the source. As shown in Fig. 11, the additional packet ingress will accumulate at the bottleneck links, thereby increasing their TAs approximately proportional to K. Thus, our new estimate for the efficiency parameter is $K(t+1) = \bar{r}/r_{\max}K(t)$, where $\bar{r}$ is the target TA and $r_{\max}$ is the maximum aggregated TA along the path, which is fed back to the source of the flow.

On adapting K, the packet influx into the system is regulated, which will be reflected in the credit queues. However, the convergence of the credits will take some time. Furthermore, the response time may vary with the length of the routes and the number of involved nodes. In order to minimize the response time, we adapt the credit queues along all paths according to K. This way, the association between packets and credits becomes weakened, as we will point out in the next section. In the realization, the adapted value of K might be piggy-backed to the downstream nodes during the MAC transaction. From the implementation point of view, it is beneficial to not actually modify the queue counter, but to store the updated K value in addition and multiply them on demand to prevent numerical inaccuracies. The complete adaption algorithm is summarized in Algorithm 2.

For the above scenarios with 3 and 7 links, we have conducted MATLAB simulations to demonstrate the effectiveness of our approach. The results are shown in Fig. 12. We observe that the system converges into a small neighborhood of the optimal efficiency, for which the TAs are at the technological limit. Solving the optimization numerically yields slightly larger Ks of 4.7 and 7.9 for the short and long flow, respectively. We have observed that this loss of efficiency is related to the step size of the TA updates. In particular, the burstiness of the packet dynamics is independent of the step size. On the other hand, smaller step sizes lead to longer queues, since more packets have to be enqueued to achieve the same TA level. Hence, the overall vari-

Algorithm 2 Intra-Flow Efficiency Adaptation.

procedure FEEDBACKUPSTREAM(Flow f, Node i)
  $r_{sum} \leftarrow \log \sum_J \exp r^f_{i,J}$
  $K_{pref} \leftarrow K^f_i (\bar{r} - r_{min})/(r_{sum} - r_{min})$ ▷ Target TA $\bar{r}$
  for all $j \in \mathrm{Nb}_{DS}(i)$ **do** ▷ Downstream Neighbors
    $K_{pref} \leftarrow \min(K_{pref}, \text{FEEDBACKUPSTREAM}(f, j))$
  end for
  return K_{pref}
end procedure

Require: $\gamma \in (0, 1)$ ▷ EWMA smoothing factor
procedure ADAPTEFFICIENCY(Flow f) ▷ At the source node s of f
  $K^f \leftarrow (1-\gamma)K^f + \gamma\,\text{FEEDBACKUPSTREAM}(f, s)$
  PROPAGATEDOWNSTREAM(f, s, K^f)
end procedure

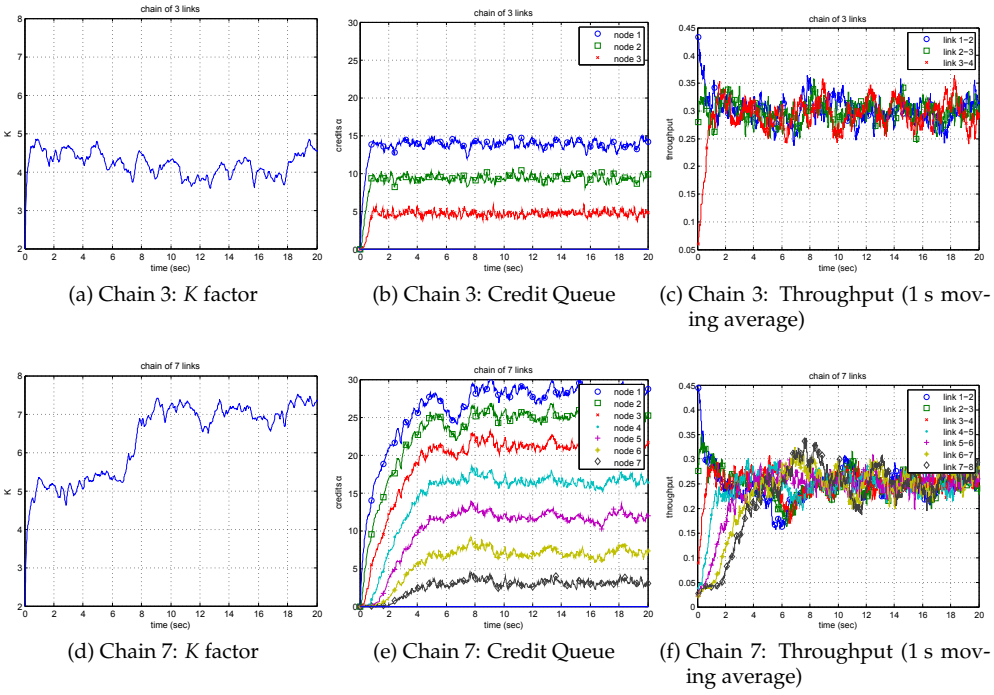
procedure PROPAGATEDOWNSTREAM(Flow f, Node i, K)
  $C^f_i \leftarrow C^f_i \cdot K/K^f_i$ ▷ Adapt credits
  for all j **do** ▷ Adapt transmission credits
    $TC^f_{i,j} \leftarrow TC^f_{i,j} \cdot K/K^f_i$
  end for
  $K^f_i \leftarrow K$
  for all $j \in \mathrm{Nb}_{DS}(i)$ **do** ▷ Downstream Neighbors
    PROPAGATEDOWNSTREAM(f, j, K)
  end for
end procedure

ability of the C and TC queues decreases with smaller step sizes, so that the accuracy of the estimated K increases. We consider the step size in more detail in section 6.3.

In the results, we observe that the credit queues do not fill evenly across the transmitters along the route. With 7 links, for example, the queue of the first link is filled first, and only afterwards the downstream queues fill one by one. As shown in Fig. 12f, it takes about 5 s until the throughput on the seventh link reaches its operation point. The convergence time depends among others on the step size, and we will consider this point in more detail in the section section 6.3.

6.2.2. Enforcing Proportional Fairness

In the vanilla approach of UO-CSMA [48], each link is able to increase its TA at all times in order to take its fair share of the wireless resources. Thus, if a transmitter experiences unfairness, it can resolve this situation locally by taking the appropriate actions itself. However, the effectiveness of this approach is limited in our case due to



(a) Chain 3: K factor (b) Chain 3: Credit Queue (c) Chain 3: Throughput (1 s moving average)

(d) Chain 7: K factor (e) Chain 7: Credit Queue (f) Chain 7: Throughput (1 s moving average)

Figure 12: Simulation of a chain of 3 and 7 links, which are assumed error-free. The PHY and MAC layer properties are chosen according to IEEE 802.11g operating at 6 Mbps. The upper and lower TA limits are 2 and -3, respectively. Furthermore, we used our extended CSMA model, a fixed step size of 20 packets per credit (see section 6.3), an EWMA smoothing factor of 0.1 and an update interval of 20 ms for the adaptation.

the technologically limited TAs. Hence, instead of increasing the TA on the affected links, the TAs of the competing links have to be decreased, which requires cooperation between all involved flows. In the following, we propose an *inter-flow adaptation* approach in order to meet the specified objective.

In the previous section, we have shown how to adapt the efficiency parameter K in order to find a working point where the bottleneck TAs are at the technological limit. However, using different K parameters with interfering flows results in weighted proportional fairness [59]. Note that there is still an ongoing discussion about the appropriate fairness objective in multi-hop mesh networks. In particular, Gambiroza *et al.* propose a fairness definition that does not penalize long multi-hop flows in favor of short ones [29]. However, we will consider equally weighted proportional fairness in the following, since its effectiveness in wireless networks has been established [96].

From the remarks above, we conclude that the efficiency parameters of different flows should be equal if the flows share resources in the WMN. The mesh nodes de-

tect resource sharing flows by monitoring the wireless medium in their vicinity. In particular, if the node carriers traffic for the flow f_1, either as transmitter, relay or receiver, and observes data or signaling frames belonging to a different flow f_2, then it identifies f_2 as a competitor to f_1. It is to note that there is a tradeoff between accuracy and efficiency in a multi bit-rate environment. With using adaptive MCSs for data and signaling frames, the MAC transaction becomes more efficient. However, the competition between flows may not be recognized due to the higher SNR requirements, which may render the decoding of these frames impossible.

In addition, the relay nodes determine the efficiency parameter K of competing flows, which may be embedded in the MAC frame as suggested in section 6.2.1. Through feedback and in-network aggregation, the source of the flow is able to obtain the minimal K parameter across all competing flows. In turn, it determines whether the considered flow is *technologically* or *fairness limited*. In the former case, it proceeds as lined out in the previous section. Otherwise, its anticipated efficiency parameter exceeds the minimal K of the competing flows, so that it would unfairly take wireless resources away from its competitors. Hence, the source adapts its efficiency according to the minimal K it has obtained from the feedback, so that unweighted proportional fairness is enforced between the involved flows. In the realization, the fed back value of K can be used as new prediction in the adaption framework of section 6.2.1.

The proposed approach introduces a form of global coupling between all flows in the WMN. In particular, a problem arises that is similar to starvation in scheduling problems. For example, consider two competing flows, where the first one is technically and the other is fairness limited. The system will adapt the efficiency parameter of the second flow according to the first one's. However, when they eventually converge, the technological limited flow will become fairness limited as soon as it increases its efficiency. Hence, it will become impossible for both to increase their efficiency regardless whether the efficiency may be unnecessarily low.

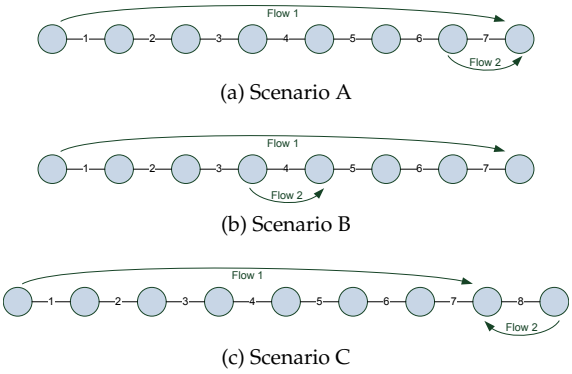
We address this problem in the following way. A flow is considered fairness limited only if it exceeds competing flows in the efficiency parameter K by more than a fixed ratio $\delta > 1$. In the example above, both flows are now allowed to increase their efficiency up to a ratio of δ of the other one's, so that one of them will reach again its technological limit. The downside is that the other flow may still increase its efficiency over the now technological limited flow by the give factor, so that the resulting flow rates are generally not proportional fair. However, the deviation from proportional fairness (PF) increases with δ. For a reasonable small δ, we observed that the inter-flow adaptation remains effective while the deviation from PF is small. The algorithm for inter-flow fairness adaptation is summarized in Algorithm 3.

For the scenarios in Fig. 13, we conducted MATLAB simulations to demonstrate the effect of inter-flow adaptation. The setup is identical to the simulations presented in Fig. 12 except that we have given the system 30 s to settle. The presented results in Table 3 are averages over the following 30 s. All scenarios consist of two flows having seven hops and a single hop, respectively. However, they differ in the way the flows interact with each other. In scenario A and B, they share a wireless link at different positions within the longer flow. In scenario C, there is no shared link. Instead, certain links of both flows are within the same collision domains and com-

Algorithm 3 Inter-Flow Fairness Adaptation.

1: **procedure** PROMICRECEIVED(Packet p, Node i)
2: $K_i \leftarrow \min(K_i, p.K)$ ▷ +Timeout for aged entries
3: **end procedure**

Require: $\delta > 1$ ▷ Hysteresis factor
4: **procedure** FEEDBACKUPSTREAM(Flow f, Node i)
5: ... ▷ As in Algorithm 2
6: **return** $\min(K_{pref}, \delta \cdot K_i)$
7: **end procedure**



(a) Scenario A

(b) Scenario B

(c) Scenario C

Figure 13: Three example scenarios. All links are considered error-free.

pete for wireless resources. For comparison, the PF rates achievable with TDMA are $(y_1, y_2) = (0.166, 0.5)$ and $(0.25, 0.5)$ in scenario A and B and in scenario C, respectively. In the analytic model, we estimated the parameters K, for which the resulting TA matched the target value (cp. section 6.2.1) within the idealized CSMA model. The simulated flow rates match the analytical results within reasonable accuracy. Without inter-flow adaptation, the resulting rates deviate from PF, which can be seen in Table 3 in the rows *WPF*. We observed that longer flows benefit in this situation, in particular, due to their longer response time of the feedback loops.

6.3. Convergence Period and Queueing Delays

In the previous sections, we have focused on the operating point to which the system converges and its properties. Now, we will consider the convergence process to reach the operating point. With convergence period, we refer to the duration from the start of the flow until it reaches its operating point within a small proximity. We have already mentioned the problem of the *delayed start* (cf. Fig. 12). During the convergence period, the incoming packets are used to fill the queues first, so that the associated TAs are raised to a target level. The queues do not fill evenly but one after

Scenario	Method	Fairness	K_1	y_1	K_2	y_2
A	Ana.		2.58	0.133	2.58	0.515
	Sim.	PF	2.5	0.136	2.5	0.519
	Sim.	WPF	4.9	0.200	1.8	0.375
B	Ana.		2.26	0.118	2.26	0.451
	Sim.	PF	2.3	0.122	2.3	0.462
	Sim.	WPF	5.0	0.205	1.0	0.198
C	Ana.		3.01	0.158	3.01	0.602
	Sim.	PF	3.0	0.163	3.0	0.611
	Sim.	WPF	5.0	0.205	2.6	0.528

Table 3: Inter-flow fairness adaptation in the scenarios A-C in Fig. 13. In contrast to the fairness model *PF*, we turned off the inter-flow adaptation in the model *WPF*. The methods *Ana.* and *Sim.* refer to analytical results in Maple (for the idealized CSMA model) and simulations in MATLAB, respectively.

another. Thus, the system cannot achieve end-to-end throughput until a queue on the last hop has sufficiently filled. For example, it takes 5 s for the flow in Fig. 12f to approximately achieve its target service level. On the other hand, *queueing delay* is an inherent problem in back-pressure based approaches. For example, in the scenario of Fig. 12 we have assigned one credit to 20 packets. Thus, the system enqueues about 300 and 600 packets at its operating point in the 3 and 7 link scenario, which would lead to excessive delays.

Within this section, we address the problem of the delayed start and the queueing delay from the practical point of view. As our first objective, the system should provide service in terms of end-to-end throughput as soon as possible, whereas its initial starting point may have a lower throughput efficiency. During the convergence period, the system should incrementally improve its efficiency until it approaches its final operating point. The rationale behind is as follows. Traffic flows can be coarsely divided into short- and long-lived flows, originating, for example, from interactive web browsing and FTP file download, respectively. For the former, the reactivity becomes more important whereas a lower throughput efficiency is more likely to be tolerated. However, the classification of newly arriving flows into short and long-lived is generally not possible. Thus, an adaptive strategy that classifies and treats flows according to their age should offer an acceptable tradeoff between reactivity and efficiency. As our second objective, the system should avoid excessive queueing that result in large end-to-end delays. In the scenarios of Fig. 12, we observe that the temporal variability of the credit queues is small after convergence. The system builds up queues mainly to generate back-pressure, and the risk of draining the queues below a certain level is small. Hence, our objective is to reduce the queueing efforts to a level of practical relevance.

We address both problems by decoupling the physical packet queues from the credit queues as suggested by Jiang *et al.* [47]. We propose an algorithm that han-

dles the above mentioned problems and reduces the end-to-end delays to a level of practical relevance. With constant step sizes in the cross-layer framework of section 5, one credit can be associated with a fixed number of packets, so that packet and credit queues are proportional. By using adaptive step sizes, the physical and credit queues generally evolve differently. When starting with a large step size, a coarse neighborhood of the target operating point is located in shorter time, and the system incrementally improves it by reducing the step size. Furthermore, the queueing delay can be reduced by limiting the amount of in-flight packets in the network. In the following sections we consider both approaches in detail.

6.3.1. Providing Service during Convergence

We consider the step sizes in (20) and (21) that express the value of a packet in terms of credits in the following. The step size controls the tradeoff between convergence speed and variability at the operating point. With a smaller step size, the convergence is slower but the credit queues vary less. We use the interquartile range (IQR) to measure the variation of the credit queues after convergence. Furthermore, we define the convergence time t_c as the duration from the start of the flow until the credit queue at the source of the flow exceeds the first quartile of the steady state queue distribution. For the above mentioned example of a 7 link chain (cf. Fig. 12), we have determined the variability and convergence time for several step sizes s in Table 4 through simulations. In the results, the tradeoff is evident, which supports our argumentation above.

Step size s	K	y	C_1	**IQR C_1**	t_c
0.50	5.05	0.215	23.6	4.11	0.45
0.10	6.77	0.247	27.4	1.55	1.95
0.05	7.07	0.253	27.9	1.10	5.40
0.01	7.46	0.256	29.1	0.55	25.9

Table 4: Convergence time t_c and variability (interquartile range (IQR)) of the credit queue C_1 in a chain of 7 links for several step sizes s, as observed in MATLAB simulations. The presented values of efficiency parameter K, flow rate y and credits C_1 are 30 s averages taken after convergence.

In the design of an algorithm for the adaptation of the step size, we have to consider the initial and target step size, the adaptation rate and the way the changes are propagated through the network. For the initial step size, we have to take two observations into account. At first, the TAs grow in relation to the bit-rate of the underlying links according to (27). An initial value that is designed for a low bit-rate may not be appropriate for higher bit-rates. Hence, the initial step size should depend on the bit-rate of the involved links. However, since multiple different bit-rates may be involved, the initial step size should be node dependent. On first sight, a link dependent initial value may seem more natural, since bit-rates are associated with links instead of nodes. However, the step size is used to update the credit queues

when a packet arrives, whereas the assocation to outgoing links is done afterwards within the routing decision (cf. Algorithm 1 on page 38).

The efficiency parameter K may vary rapidly during convergence, and these changes quickly propagate. In simulations with long flows, for example, we have made the typical observation that the convergence slows down with the distance from the source of the flow. The relative step size of downstream nodes that are involved by the flow a little later may be different from upstream nodes because the credit queues are scaled according to K in the intra-flow efficiency adaptation (cf. section 6.2.1). As shown in Fig. 12d, the parameter K quickly increases at the source in the scenario with long flows, so that the initial step size for the transmitter on the last hop is smaller compared to its upstream nodes, which explains the slowdown of the convergence. Thus, the initial step size should be used relative to K.

In consideration of the specified design goals above, the source node of a flow f maintains the *current step size* s^f, which is initialized with an empirical value s_0 that has been determined for the technologically lowest bit rate R_{base} and the actual system parameters (e.g. upper and lower TA bounds). When a packet arrives exogenously, the source node embeds its current step size into the packet. Thus, the value of a packet is readily available whenever a packet enters or leaves a node. In either case, the node and transport credits are updated according to the embedded step size that is down-scaled according to the maximal bit-rate among its downstream links. Furthermore, the credit update is independent from the current efficiency parameter K, since the node and transmission credit queues are independent from K, as well.

The *target step size* determines the operating point in terms of step size after convergence. Thus, it affects both the reactivity and the variability after convergence. Environmental changes like newly arriving flows or node and link breakdowns may alter the operating point for the considered flow. The duration of the necessary re-convergence process mainly depends on the number of packets necessary to drive the credit queues to their new value. In that case, a high reactivity is desirable. However, it comes at the expense of higher variability at the operating point. An advanced approach may solve this tradeoff by detecting and reacting on changes, e.g. it enlarges the step size if another operating point should be approached. However, our intention is to verify the effectiveness of the step size adaptation in the initial convergence period. Using an appropriate means of change detection, it is straightforward to extend our approach to the re-convergence case.

Nevertheless, it is not sufficient to use a common target step size for all flows across the network. For example, let us consider two flows. The first flow achieves significant higher throughput than the second, because it uses higher bit-rates, for example. Hence, at their respective sources the credit queue of the first flow is smaller compared to the second (cf. section 5.2). A common target step size for both flows would result in a higher variability for the first flow. Thus, a common target step size is not able to provide equal variability across all flows. We introduce a flow rate dependent target step size s_∞^f. It is determined as product of an empirical determined relative step size s_∞ and the credits at the flow source C_i^f. Remember that C_i^f is inversely proportional to the flow rate according to (22).

Algorithm 4 Step Size Adaptation.

Require: $s^f = s_0 > 0$ ▷ Initial step size w.r.t. R_{base}
Require: $s_\infty > 0$ ▷ Target step size relative to 1 credit
Require: $t_{up} > 0$ ▷ Update interval
Require: $t_{hl} > 0$ ▷ Half-life of the step size
procedure UPDATESTEPSIZE(Flow f) ▷ At the source node i of f
    $s^f_\infty \leftarrow C^f_i \cdot s_\infty$
    $s^f \leftarrow s^f_\infty + (s^f - s^f_\infty) \cdot 2^{-t_{up}/t_{hl}}$
    SLEEP(t_{up})
    UPDATESTEPSIZE(f)
end procedure

procedure PACKETARRIVAL(Flow f, Packet p) ▷ At the source node i of f
    $p.s \leftarrow s^f$
    PACKETRECEIVED(f, p, i)
end procedure

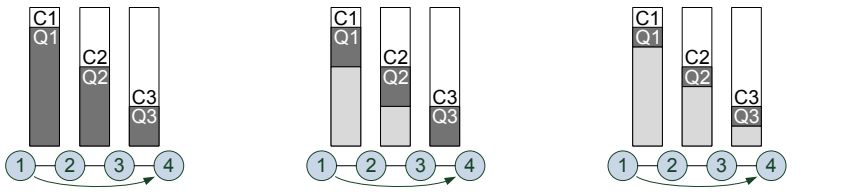
Require: $R_{base} > 0$ ▷ Technologically lowest bit-rate
procedure PACKETRECEIVED(Flow f, Packet p, Receiver j)
    $R \leftarrow \max(R_{j,k}), \forall k \in \mathrm{Nb}_{DS}(j)$ ▷ Downstream Neighbors
    $C^f_j \leftarrow C^f_j + p.s \cdot R_{base}/R$
    $TC^f_j \leftarrow TC^f_j + p.s \cdot R_{base}/R$
end procedure

Require: $R_{base} > 0$ ▷ Technologically lowest bit-rate
procedure PACKETTRANSMITTED(Flow f, Packet p, Transmitter i)
    $R \leftarrow \max(R_{j,k}), \forall k \in \mathrm{Nb}_{DS}(i)$ ▷ Downstream Neighbors
    $C^f_i \leftarrow C^f_i - p.s \cdot R_{base}/R$
    $TC^f_i \leftarrow TC^f_i - p.s \cdot R_{base}/R$
end procedure

We update the current step size using an exponential decay. The source of a flow updates its step size in fixed intervals towards the target value. The magnitude of change is controlled by a given half-life. Algorithm 4 summarizes our above described approach.

6.3.2. Back-Pressure with Reduced End-To-End Delays

With the UO-CSMA approach presented so far, our focus has been throughput efficiency, and it remains our main objective throughout the paper. However, an inherent problem in back-pressure based routing is the end-to-end delay. Within this section,



(a) Vanilla back-pressure. (b) According to Horizon [99]. (c) Proposed delay-limiting approach.

Figure 14: Physical (dark gray) and virtual queues (also called shadow queues, light gray) for a chain topology with 3 link and a single flow.

we review this problem and show how the delays can be reduced to a level of practical relevance at the expense of throughput efficiency.

The causes of end-to-end delay are manifold. There is the actual radio transmission that consumes medium time. The acknowledgement and forwarder selection process consumes time in addition. However, the back-pressure approach introduces two further sources of delay. At first, the links have to generate back-pressure for scheduling purposes with UO-CSMA, which translate to queueing delay for the involved packets. Furthermore, there is no notation of delay in back-pressure based routing. In fact, the objective is throughput maximization and the minimization of resource usage is not considered. For example, in the local perspective of back-pressure routing, two different paths offering the same throughput are considered equal regardless whether one of them is substantially longer and has higher resource consumptions. In this section, we will focus on the delays caused by back-pressure scheduling and discuss routing related issues in section 6.5. Note that the robust ACK scheme presented in section 6.1 can be seen as an enlargement of the link queues by one hidden space. Thus, its contribution to the end-to-end delay is covered within the following discussion.

For the scheduling problem, we have already introduced an important technique to reduce the delays in section 5.3 that originates from *Horizon* by Radunovic *et al.* [99]. In back-pressure routing, the differential back-log between two nodes is used for scheduling. In order to provide positive back-log with increasing distance from the destination, the queues have to grow (cf. Fig. 14a). In general, a node has to generate a sufficient amount of back-pressure before it is allowed to generate throughput, and this threshold increases with the distance from the destination. However, by replacing the queue up to the threshold by shadow queues, which are actually counters, the amount of queued packets per flow in the network is reduced (cf. Fig. 14b). In particular, the growth of the queues is reduced from quadratic to linear efforts in the number of hops [12].

The throughput-delay-complexity tradeoff is characteristic for MWS (cf. section 2.2). In order to limit the end-to-end delay of UO-CSMA, we generally have to sacrifice throughput. Nevertheless, the precise delay characterization of MWS in wireless multi-hop networks with general interference constraints is still an open prob-

lem [38, 39, 127]. Neely *et al.* characterize the tradeoff between utility and end-to-end delay for NUM problems [32, 86, 89]. In particular, for a given control parameter V[4] they present an algorithm that achieves a total network utility within $O(1/V)$ of the optimum while the average network delay is within $O(\log V)$, which is the best possible tradeoff [86]. However, the stochastic upper bounds are order results, and they may be of limited value in characterizing real systems. Large-deviations analysis is an alternative approach, which tries to estimate the probability of queue overflows. As an interesting result, Venkataramanan *et al.* have observed that a modified back-log that favors links closer to the destination may lead to lower end-to-end delays [119]. As a further approach, Gupta *et al.* propose to analyze a reduced system capturing the bottlenecks only [38, 39]. They derive a lower bound on the end-to-end delay performance and propose a similar modification of the back-log definition. The heavy traffic approximation is another approach to address network delay. Using this approach, Yi *et al.* provide an exact characterization of the average delay [131].

As lined out in [85, 89], Little's law is useful in deriving bounds on the network delay. For a stable system, Little's law can be stated as

$$L = \lambda \cdot W,$$

where L, λ and W are the average queue length, arrival rate and delay, respectively [7]. Thus, for each path from source to destination we can upper bound the average delay by controlling the number of in-flight packets in relation to the achieved rate on the path. Thus, the source may keep track of the number of in-flight packets and limit the packet ingress accordingly. However, the actual aggregated queue length across all nodes is volatile in relation the time necessary for the feedback. Furthermore, with a centralization of the logic solely at the source the approach is unable to control the delays across different paths in the network. Hence, we favor a distributed solution, in which each transmitter dynamically determines an *upper limit for its queue* $\hat{Q}$ depending on the average arrival rate λ and a given upper limit for the per-hop delay t_{max} (cf. Algorithm 5). Thus, the physical queues Q per flow and node have a lower and an upper limit $0 \dots \hat{Q}$, so that the amount of physical packets per hop remains small (cf. Fig. 14c). In the best case, the back-pressure should be primarily generated in the virtual queues C, and the remaining physical queue space Q should account for all variations due to the underlying network dynamics. However, we have to sacrifice throughput efficiency if the network dynamics cannot be outweighed via the remaining physical queue.

According to Jiang *et al.*, the delays could be reduced by injecting slightly more (virtual) credits than (physical) packets [47]. This way, the actual arrival rate of physical packets becomes smaller than the (physical) service rate of the link, so that the (physical) queues tend to zero. Using this approach, the problem is that it achieves order results only. If the injection rate for additional credits is small, the delays still remain substantial. If it is large, the throughput seriously suffers. The advantage of the pro-

[4] Our efficiency parameter K from section 6.2 is comparable to V. However, it is important to remember that CSMA imposes technological limits on K in our case.

posed delay limiting approach is that the queue limits are adapted according to the specified delay targets, and the injection rate of additional credits is calculated implicitly. However, the saturation assumption is violated due to the decoupling of the packet and credit queues. For example, the queue may be empty at the time the link gets a TXOP. In this case, it is suggested to sends *dummy packets* that alter the credits only while the packet queues remain unchanged [12, 47, 85]. The credit transition within dummy packets is important for the convergence of the CSMA Markov chain since the relationship between the TAs and the consumed medium time is conserved this way. If the transmitter would remain idle instead until its queue gets non-empty again, the TAs become unable to control the node's throughput. However, the sole purpose of dummy packets is to transport transmission credits, whereas they consume precious resources of the WMN.

Empirically, we observed that *virtual transmissions* are able to mitigate dummy packets to a large extend while preserving the stability and enhancing the throughput efficiency. In contrast to dummy packets, a virtual transmission is not a physical radio transmission that consumes wireless resources. Instead, the sender simulates the transmission, i.e. after contention for a TXOP it enters the transmission state and remains there for the (anticipated) transmission duration. The difference is that neighboring nodes are not affected by the virtual transmission, so that precious resources are preserved. On the other hand, a virtual transmission includes a credit transition, whereas the actual credit delivery is postponed to the next physical packet exchange. Furthermore, the simulation of the transmission also includes the reception process at the receivers. The transmitter draws a random sample from its PSR statistics to determine potential receivers.

Virtual transmissions can be used in both cases, either if the transmitter queue is empty, or if the receiver queue is full. For an anycast link with multiple destinations, a virtual transmission is initiated whenever there is at least one queue overflow. In addition, the congestion controller has to consider the queue limits, too. It only injects packets if the queue at the source can accommodate it. Otherwise, it only injects the associated credit and the packet may be dropped or buffered for future release.

In our approach, the nodes have to exchange the information about queue sizes and limits within its neighborhood, so that upstream transmitters are aware of the queue utilization of potential relays. In a straightforward way, we put this information into RTS and CTS signaling frames. On overhearing them, neighboring nodes can extract the necessary information. Furthermore, a timeout mechanism takes care that this information is periodically propagated even if the considered node has not accessed the medium for a longer period, so that possible stalls are resolved. The algorithm is sketched in Algorithm 5.

In our prototype, we estimate the arrival rate using an EWMA rate estimator. Furthermore, we have introduced a lower limit for the per-hop queues $\hat{Q}_{min}$ in Algorithm 5 to ensure that the node remains operational. Table 5 shows simulation results for a 7-link chain (Scenario A) and a double chain of 2x7 links (Scenario B) for different per-hop delay limits. Considering scenario A, it is evident that smaller per-hop delay limits are able to decrease the amount of in-flight packets $\sum_i Q_i$ and the end-to-end delays t_{E2E}. At the same time, the risk of queue under- and over-runs increases. Thus,

Algorithm 5 Delay Limiting.

Require: $t_{max} > 0$ ▷ Hop-wise delay limit
Require: $\hat{Q}_{min} > 0$ ▷ Min. queue limit

1: **procedure** UPDATEQUEUELIMIT(Flow f, Node i, Arrival rate λ_i^f)
2: $\hat{Q}_i^f \leftarrow \max(\hat{Q}_{min}, \lambda_i^f \cdot t_{max})$ ▷ According to Little's law
3: **end procedure**

4: **procedure** PACKETARRIVAL(Flow f, Packet p) ▷ At the source node i of f
5: **if** $Q_i^f >= \hat{Q}_i^f$ **then**
6: $p \leftarrow \emptyset$ ▷ Drop incoming packet
7: **end if**
8: PACKETRECEIVED(f, p, i)
9: **end procedure**

10: **procedure** BACKOFFEXPIRED(Flow f, Node i, Link l)
11: **if** $0 = Q_i^f$ **or** $\exists Q_j^f >= \hat{Q}_j^f, \forall j \in l.dst$ **then** ▷ Receiver set $l.dst$
12: VIRTUALTRANSMIT(f, i, l) ▷ Physical queue length Q
13: **else**
14: TRANSMIT(f, i, l)
15: **end if**
16: **end procedure**

transmissions are more often virtual, which can be observed at the virtual goodput at the source $\sum_j x_{0,j}^{virt}$. In turn, the flow rate y deviates from the actual goodput, e.g. the goodput at the flow source $\sum_j x_{0,j}$. In addition, the limited amount of in-flight packets leads to less efficient schedules and the goodput slightly decreases.

The same observations apply to scenario B, as well. In addition, it demonstrates the effect of multi-path on the end-to-end delays. Despite additional relays have been added, the achievable physical goodput is almost unchanged. Thus, the flow splits up between both chains and halves the per-hop queue limits at the relay nodes, so that the aggregated queue limit remains comparable to scenario A. With smaller per-hop limits, the risk of virtual transmissions increases. Furthermore, we make two interesting observations with 400 ms and 50 ms in scenario B. With 400 ms, the amount of in-flight packets and the end-to-end delays substantially increase in scenario B. Hence, we argue that in this case, the per-hop queue limits are rather conservative and do not play a significant role in the queueing process. With 50 ms on the other hand, the aggregated queue limit is higher than in scenario A. This, in turn, causes a higher end-to-end delay as compared to the previous scenario. The observation is caused by the lower queue limit of 4 packets, i.e. $\hat{Q}_{min} = 0.009$ in the given notation. In summary, multi-path introduces additional relay nodes that may split up the traffic flow, so that the individual per-hop queue limits decrease. Thus, the risk of queue over- and under-runs grows and virtual transmissions become more likely. In

Scenario	Hop-Delay	t_{E2E}	y	$\sum_j x_{0,j}$	$\sum_j x_{0,j}^{virt}$	$\sum_i Q_i$	$\sum_i \hat{Q}_i$
A	400 ms	921 ms	0.265	0.264	0	0.225	0.728
	200 ms	835 ms	0.275	0.261	0.003	0.200	0.363
	100 ms	420 ms	0.291	0.259	0.035	0.092	0.179
	50 ms	216 ms	0.349	0.248	0.097	0.037	0.086
B	400 ms	1820 ms	0.276	0.252	0.019	0.422	0.714
	200 ms	825 ms	0.336	0.256	0.070	0.171	0.353
	100 ms	449 ms	0.432	0.247	0.189	0.072	0.172
	50 ms	354 ms	0.478	0.241	0.239	0.046	0.121

Table 5: End-to-end delay (t_{E2E}) and flow rate (y) results obtained from the Brn.Sim network simulator for a 7-link chain (A) and a double chain of 2x7 links (B). The results are 20 s averages excluding a 5 s warm-up phase. The last two columns show the amount of in-flight packets across all nodes and their respective upper limit. The two columns to the left refer to the physical and virtual goodput at the source. For a better comparability, the rate and queue values are given as fraction of the maximal link throughput of 442 pps.

addition, the end-to-end delay increases if the per-hop limits reach the lower bound and cannot be decreased further.

Virtual transmissions do not compete with neighboring links within the contention area, which may negatively affect the convergence of the CSMA Markov chain. With increasing virtual throughput in the example above, the physical goodput has slightly decreased, i.e. the system converges to a less efficient operating point. On the other hand, the virtual goodput contributes to the traffic rate of the flow. The higher flow rate has potentially negative effects on the fairness between competing flows, since the congestion controller provides PF to the joint flow rate of virtual and physical goodput. Thus, the hop-delay parameter should be chosen in order to keep the virtual flow sufficiently small, so that the above mentioned issues do not become critical. However, a tradeoff arises with multi-path routing. Additional paths and relays lead to further splitting of the flow. The end-to-end delays increase if our delay limiting approach becomes ineffective, i.e. if the rate of the sub-flows becomes too small and the lower bounds for the per-hop queue limits apply. In these cases, concentrating the traffic flow while dropping low-throughput paths may reduce the end-to-end delay.

With UO-CSMA, the queueing behavior differs from TCP. The intention of TCP is to fully utilize all intermediate channels in order to maximize the end-to-end throughput. In particular, TCP tries to adapt its window size, so that the maximum number of concurrent transmitters can be activated at each time instance. Due to the overlapping collision domains of WMNs, the optimal window size is generally smaller than the number of relay nodes. Increasing the window size further does not change the throughput but affects the end-to-end delay according to Little's law. However, it may even reduce the throughput when considering hidden nodes [27]. On the other hand, UO-CSMA relies on back-pressure and thus queueing information to adapt the MAC persistency in order to converge to an efficient operating point of the CSMA

Markov chain. In particular, each link is assumed to be saturated. In the vanilla approach, this assumption is justified since back-pressure is expressed in terms of queued packets. Furthermore, the assumption allows for an independent operation of CSMA and the queueing process. The presented approach to limit the end-to-end delay violates the saturation assumption to varying degrees, and the resulting trade-off can be controlled via the per-hop delay limit parameter. An integrated approach which considers both CSMA and queueing in non-saturated conditions seems to be a promising approach for the future work. However, the required theoretic foundations still have to be developed.

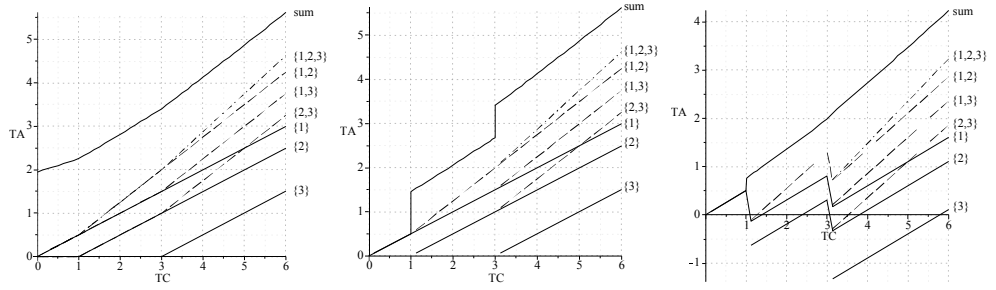
6.4. Compensation for Candidate Set Dynamics

The intra- and inter-flow adaptation presented in section 6.2 try to find a working point in terms of TA that offers the highest efficiency while maintaining the physical constraints. However, the characteristics of anycast links may cause oscillations in the feedback loop. For example, consider the transmitter i having the neighbors $\{1, 2, 3\}$. Let us suppose that nodes 2 and 3 have longer credit queues than the transmitter ($C3 > C_2 > C_i$), and only node 1 has fewer credits ($C_i > C_1$). Thus, the transmitter assigns TCs to neighbor 1 only. According to the CSMA scheduling in section 5.4, the transmitter increases the TA on *all* hyperlinks containing neighbor 1. In particular, the transmitter also activates the hyperlinks to node 1 and combinations of neighbor 2 and 3, although it does not intend to transmit a packet to neighbor 2 or 3.

On the one hand, the unnecessary receivers cause additional MAC overhead, since the number of signaling frames and the frame size increases. On the other hand, the cumulated TA increases with the number of activated hyperlinks, since *all* of them have to be taken into account at all times. For example, if the minimal TA is 0 for a single link, then the cumulated TA[5] cannot be smaller than $\log(7) \approx 2$ for 3 neighbors, as shown in Fig. 15a. The increased minimum for the cumulated TA is problematic because it leaves less room for the per-link TA adaptation in the presence of technological upper limits.

The straightforward solution would be not to activate a hyperlink if it contains unnecessary receivers. However, this leads to oscillations. For example, if the system evolves and the credit queue at the transmitter grows, then neighbor 2 might have a lower credit queue ($C_i > C_2$) , so that it is included in the routing decision and the transmitter assigns TCs to it. Thus, the transmitter activates the link to neighbor 2. In addition, it has to activate the hyperlink to neighbors 1 and 2, which is one of the hyperlinks we decided not to activate in advance. The TA of that hyperlink is equal or larger than the TA of the link to neighbor 1 by construction (cf. section 5.4). Thus, the cumulated TA at the transmitter abruptly changes, as shown in Fig. 15b. The intra-flow efficiency adaptation uses the cumulated TAs to determine the working point in terms of K, so that the system may converge to a different working point after neighbor 2 has been included. Furthermore, the system may start to oscillate if the credit dynamics causes regular crossings of the credit levels of both the transmitter

[5] Remember that the TAs sum up in the exponential domain.



(a) According to (27). (b) Selective activation of links without compensation. (c) According to (28).

Figure 15: TA vs. TC for three candidate receivers having $C_1 = C_2 - 1 = C_3 - 3$. The PSR of single, two and three receiver links are 0.50, 0.75 and 0.875, respectively. All links use the same bit-rate.

and node 2.

We address the problem of abrupt changes of the aggregated TA in the following way. Let us reconsider the example above: Only neighbor 1 is included in the routing decision of node i, i.e. $C_1 < C_i \leq C_j \; \forall j > 1$. Since all other receivers are unnecessary in this case, node i activates link $(i, 1)$ only using the TA $r_{i,1}$, as in the straightforward approach. Let us assume the system evolves in a way that receiver 2 is included in the routing decision ($C_1 \leq C_2 < C_i$). Now, node i activates the links $(i, \{1, 2\})$ and $(i, 2)$ in addition. Note that the former link would already have been activated according to the traditional approach. Thus, on activating link $(i, \{1, 2\})$, we split the TA $r_{i,1}$, which link $(i, 1)$ has had in isolation, into the TAs of both links $(i, 1)$ and $(i, \{1, 2\})$ in a way that the cumulated TA is preserved. Since the aggregation is done in the exponential domain, we can accomplish this task by subtracting a compensation term $e_{i,\{1,2\}}$ from both TAs. In particular, if both links operate with equal bit-rates, the compensation term is $e_{i,\{1,2\}} = \log(2)$, i.e. we subtract $\log(2)$ from $r_{i,1}$ and $r_{i,\{1,2\}}$.

In the following, the uncompensated TAs have a subscript only (e.g. $r_{i,M}$), whereas TAs with an additional superscript are compensated according to the superscript (e.g. $r^J_{i,M}$). For the general case, let $J' = \{1, \ldots, j-1\}$ be the set of neighbors of node i that the routing decision has currently chosen ($C_1 \leq \ldots \leq C_{j-1} < C_i$). Thus, node i has used the TAs $r^{J'}_{i,M'}$ adapted to J' for all $M' \in \mathcal{P}(J')\backslash\emptyset$. Now, node i chooses the candidate receiver j, in addition, with $C_{j-1} \leq C_j < C_i$. For link (i, M'), we adapt the TA as $r^J_{i,M'} = r^{J'}_{i,M'} - e_{i,M}$ with $J = J' \cup \{j\}$, $M = M' \cup \{j\}$, $m = \max(M')$ and e is defined as

$$e_M = \log\left(1 + \exp\left[(R_{i,M} - R_{i,M'}) \sum_{K \in \mathcal{P}(M')} p_{i,K} \sum_{L \in \mathcal{P}(\{1,\ldots,m-1\}\backslash M')} \beta_{i,K\cup L}\right]\right) \quad (28)$$

In addition, the link (i, M) is activated using the compensation terms of link (i, M'), i.e. $r^J_{i,M} = r_{i,M} - (r_{i,M'} - r^J_{i,M'})$. In this way, all links except (i, j) have been considered so far. As a heuristics for the newly activated link (i, j), we use the lowest compensation terms that any other link has experienced so far, i.e. $r^J_{i,j} = r_{i,j} - \min_J(r_{i,M'} - r^J_{i,M'})$. With multiple flows per link, we first apply the compensation to the TAs, and afterwards we select the flow per link (cf. section 5.4).

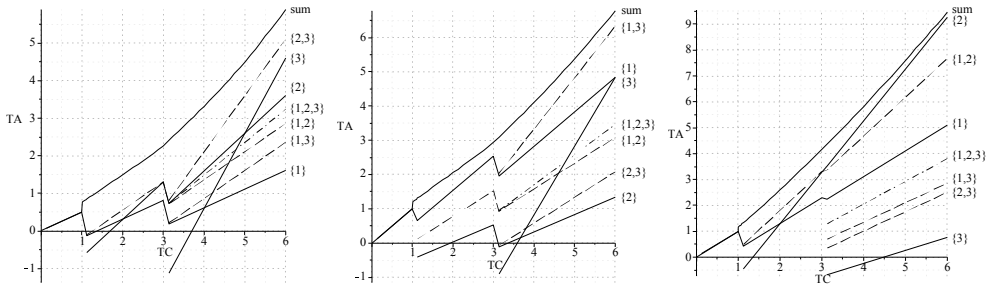
The resulting cumulated TA of node i will still be discontinuous due to the heuristics for the newly introduced TAs. Fortunately, the jumps in the cumulated TA will vanish in the high TA regime. Note that the aggregated TA is calculated via a log-sum-exp expression, which is an approximation of the max function [11], and the influence of the newly introduced term vanishes if the difference to the already contained TAs increases. As shown in Fig. 15c, our heuristics mitigates the discontinuity at $TC_i = 1$ to some extend and the discontinuity at $TC_i = 3$ is not noticeable anymore. Let us consider the example in Fig. 15 in detail. In the case $C_1 < C_i$ and the routing decision chooses neighbor 1 only, then the transmitter activates link $(i, 1)$ only using $r_{i,1} = R_{i,1} p_{i,1} \beta_{i,1}$. If the credits of the transmitter increase in a way that $C_1 \leq C_2 < C_i$ and receiver 2 is included in the routing decision, then we have the following.

$$
\begin{aligned}
e_{\{1,2\}} &= \log\left(1 + \exp\left[(R_{i,\{1,2\}} - R_{i,1}) p_{i,1} \beta_{i,1}\right]\right) \\
r^{\{1,2\}}_{i,1} &= r_{i,1} - e_{\{1,2\}} = R_{i,1} p_{i,1} (\beta_{i,1} + \beta_{i,\{1,2\}}) - e_{\{1,2\}} \\
r^{\{1,2\}}_{i,\{1,2\}} &= r_{i,\{1,2\}} - e_{\{1,2\}} = R_{i,\{1,2\}} (p_{i,1} \beta_{i,1} + p_{i,\{1,2\}} \beta_{i,\{1,2\}}) - e_{\{1,2\}} \\
r^{\{1,2\}}_{i,2} &= r_{i,2} - e_{\{1,2\}} = R_{i,2} p_{i,2} \beta_{i,\{1,2\}} - e_{\{1,2\}}
\end{aligned}
$$

The explicit expressions for three and more receivers can be derived in the same way.

Note that the sum expression in the calculation of the compensation terms in (28) looks very similar to the estimation of the TAs in section 5.4. However, an important difference can be found in the second summation, where we do not consider the whole complement of M, but only the subset having lower costs. Hence, the above compensation term $e_{\{1,2\}}$ does not contain any other TCs except $\beta_{i,1}$, so that it vanishes as the difference between the costs C_1 and C_2 decreases. Furthermore, $e_{\{1,2\}}$ stays the same regardless whether we include further candidate receivers, since it is independent from all TCs except $\beta_{i,1}$.

We have already presented the resulting TAs for equal bit-rates in Fig. 15c. The results for different bit-rate combinations are plotted in Fig. 16. By using different bit-rates across the candidate receivers, the compensation terms change slightly. For example, in Fig. 16c receiver 3 has the lowest bit-rate so that the compensation terms are generally smaller then $\log 2$. In particular, they are almost not noticeable for the links $(i, 2)$ and $(i, \{1, 2\})$.



(a) Rate vector $R = (1, 2, 4)$. (b) Rate vector $R = (2, 1, 4)$. (c) Rate vector $R = (2, 4, 1)$.

Figure 16: TA vs. TC for three candidate receivers having $C_1 = C_2 - 1 = C_3 - 3$. The PSR of single, two and three receiver links are 0.50, 0.75 and 0.875, respectively. The bit-rate of the single receiver links is given in the rate vectors. The bit-rate of multi-receiver links is the minimum of the involved single-receiver links.

6.5. Route Pre-Selection via Routing Metrics

The rationale of back-pressure routing is to find the throughput-optimal path(s) without relying on topology information. However, the straightforward solution is agnostic to network delay and resource usage. Individual packets may experience excessive delays. Furthermore, the system does not favor shorter routes (in terms of resource usage), which may increase the end-to-end delay and energy consumption further. Neely *et al.* address the mentioned issues using a shortest path bias in the routing decision. In particular, they introduce the routing metric of the associated paths in the calculation of the back-pressure, so that shorter paths are favored [32, 88].

As another approach, Gupta *et al.* target at bounding the expected delay [40]. Continued by Naghshvar *et al.*, they define a back-pressure and a forwarder selection policy that is provable throughput-optimal [84]. The basic idea is to dynamically determine the expected delay for all neighbor relays, and select the neighbor with the lowest delay as forwarder. The draining time based scheduling as proposed by Subramanian *et al.* uses a similar approach to address the delay performance in back-pressure routing [112].

In recent works, Bui *et al.* consider the resource consumption in the objective function. In summary, they introduce a tradeoff term into the routing decision, which controls the extent to which additional resources can be utilized [12]. Ying *et al.* introduce the shortest path information directly into the optimization problem. In particular, they define a factor, for which back-pressure is traded off against the routing metrics. Thus, in low traffic scenarios the nodes are encouraged to select short routes. With increasing traffic, the allowed deviation from the shorted path increases in the same way as the intermediate queues grow [132]. Moeller *et al.* presents a very similar approach that introduces the ETX routing metric as penalty function into the objective function [83].

In the following, our objective for the routing decision is the exploitation of MUD while ensuring an acceptable end-to-end delay. The system may use topology information obtained from state-of-the-art protocols for link quality estimation. This way, we can eliminate all routes that lead to dead ends and focus on feasible routes. We differentiate between flows with short and long routes, i.e. whether the shortest path contains few or many relays.[6] In the former case, spatial (multi-path) diversity is generally not available since the transmitters compete for the same radio resources. On the other hand, MUD can be obtained if the involved links are unreliable and the environment allows for that. For long routes, both MUD and spatial diversity are available. In order to limit the resource usage and the end-to-end delay, only the most promising paths should be taken and the number of involved relays should be kept reasonable small. A broad diversification of the traffic within the network should be prevented, so that the traffic flow per link remains sufficiently high. Otherwise, the arising virtual transmissions may have detrimental effects on end-to-end delay, throughput and fairness as described in section 6.3.2.

Our route pre-selection heuristics aims at supporting the dynamic routing decision by a proactive and traffic-independent routing metrics in a way comparable to Yuan *et al.* [133]. Routing paths are pruned according to their anticipated effectiveness, expressed in terms of a cumulative metric like ETX or expected transmission time (ETT) [20, 22]. Before continuing with the presentation of our approach, we introduce the concept of routing stretch and relay count stretch. The routing metric stretch factor, or simply *routing stretch*, is defined as the ratio of the maximum end-to-end metric of all considered paths with respect to the shortest path. The *relay count stretch* is analogously defined in terms of the number of considered relay nodes.

In the route pre-selection, we prune the network topology according to upper limits for both routing and relay count stretch. Thus, the route pre-selection only considers the network topology in the same way as state-of-the-art WMN routing protocols like optimized link state routing (OLSR) [45] and it is not traffic-adaptive. In particular, for every network node i we estimate the shortest path between source and destination of the packet that includes the considered node, which can be obtained from two passes of the Dijkstra algorithm. We refer to the resulting route metrics as m_i, and the shortest path among them has the metric m^s. Given a maximal routing stretch of f_r, the largest admissible metric for a relay is $m^r = f_r m^s$. In addition, we determine the largest shortest path metric m^h, for which the limit of the relay count stretch is still satisfied. A neighbor i will not become a relay if its shortest path metric m_i exceeds the given routing or relay count stretch limits, i.e. it holds $m_i > m^h$ or $m_i > m^r = f_r m^s$. Algorithm 6 summarizes the described procedure.

7. Evaluation in Illustrative Scenarios

In this section, we evaluate the proposed cross-layer protocol within small and synthesized scenarios in order to illustrate the operation of the system. Our objective is

[6] Note that a numerical differentiator, e.g. in terms of hop count, has to consider the actual propagation environment and the MCSs used.

Algorithm 6 Route Pre-Selection.

```
1: procedure DIJKSTRA(Source n)
2:     ...
3:     return (H, D)                          ▷ Shortest path hop count and metric
4: end procedure

Require: f_r ≥ 1                                          ▷ Routing stretch limit
Require: f_h ≥ 1                                          ▷ Relay count stretch limit
5: procedure SELECTCANDIDATES(Packet p)                   ▷ At node i
6:     (H_d, D_d) ← DIJKSTRA(p.dst)
7:     m^r ← f_r · D_d(p.src)                             ▷ Max. E2E metric
8:     h ← f_h · H_d(p.src)                               ▷ Max. relay count
9:     (H_i, D_i) ← DIJKSTRA(i)
10:    for all k do
11:        D(k) ← D_i(k) + D_d(k)
12:    end for
13:    I ← SORTASC(D)                                     ▷ Sorted index array
14:    m^h ← 0
15:    for l ← n downto 1 do
16:        if l = n or D(I(l)) ≠ D(I(l + 1)) then
17:            k ← l                                      ▷ Handle nodes with identical metric
18:        end if
19:        if k ≤ h or l = 1 then
20:            m^h ← max(m^h, D(I(l)))
21:        end if
22:    end for
23:    C ← ∅,
24:    for all k do                                       ▷ Match against routing and relay count stretch
25:        if k ∈ Nb(i) and D(k) ≤ m^r and D(k) ≤ m^h then
26:            C ← C ∪ {l}
27:        end if
28:    end for
29:    return C
30: end procedure
```

to validate the effectiveness of the algorithm (cf. section 5) and our design decisions (cf. section 6) under more realistic conditions using a detailed packet level simulator[7] based on JiST/SWANS [5]. Thereby, we evaluate several facets of the protocol, one at a time, in isolated scenarios and characterize the static and dynamic properties of the system. Due to the limited scope of the scenarios, we are able to solve the underlying optimization problem numerically. We compare the simulation outcomes to the numerical results in order to illustrate the gaps between theory and practice.

[7] See `http://sarwiki.informatik.hu-berlin.de/Brn.Sim` for details on the Brn.Sim simulator.

Parameter	Value
No. links	3, 7
Path loss	Log-distance, exponent 6
Channel	AWGN
Fading	None
Receiver	BER & cuml. interference
PHY	IEEE 802.11ag, 6 Mbps
PSR	100%
Propagation delay	0 ms
Radio turn-around	0 ms
TA limits	$4 \ldots 8.8, 10.8, 12.8$
Inter-flow hysteresis	1.1
Step size (initial/target)	$0.1 \cdot K^{-1}$ / $0.002 \cdot K^{-1} \cdot C^{-1}$
Step size update interval	200 ms
Flow duration	20 s evaluated (excl. warm-up)
Packet size	1500 Byte
Seeds	20

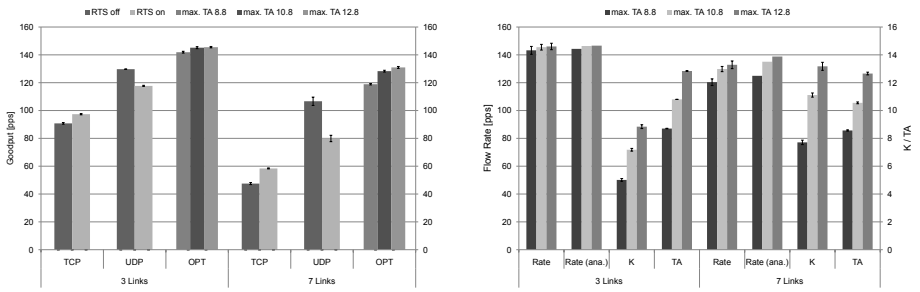
Table 6: Simulation Parameters

7.1. Efficiency Adaptation with Limited TAs

To start with, we illustrate the convergence to the predetermined tradeoff point between throughput efficiency and collisions. We consider an equispaced chain of nodes consisting of 3 and 7 links, as already introduced in section 6.2. Remember that only direct neighbors in the chain can exchange packets. Non-neighboring nodes are not able to carrier-sense each other, so that a hidden node situation might emerge depending on the protocol.

The simulation parameters are summarized in Table 6. In particular, fading has been absent in the simulations since we wanted the wireless links to be almost error-free. In the same line, the bit-rate was fixed and we disabled both propagation delay and the radio turn-around delays. In addition, we used the log-distance path loss model [37]. Throughout section 7, we have parameterized the model with a high path loss exponent to be able to generate the LCG of interest. Since it is not our focus in this section, we used a sufficiently large per-hop delay limit and step size in order to minimize their influence on the results. Thus, virtual packets should not occur often (cf. section 6.3.2). In addition, we varied the target TA the system should converge to. Note that the transmission rate is approximately $R^l \approx 440$, i.e. $r^l \approx 6.1$.

In Fig. 17a, we have plotted the resulting goodput along with the standard deviation across all repetitions. For comparison, we additionally plotted the results for TCP and genie-aided saturating UDP flows using the Berlin RoofNet (BRN) variant of the dynamic source routing (DSR) [56] protocol, either with or without using the RTS/CTS prologue. As supposed, Fig. 11 shows that higher target TAs result in higher goodput. Furthermore, it is remarkable that the proposed protocol (OPT) per-



(a) Goodput of the proposed protocol in relation to DSR with TCP and UDP on top ($\pm std.dev.$).

(b) Flow rate, efficiency parameter K and TA at the bottleneck ($\pm std.dev.$).

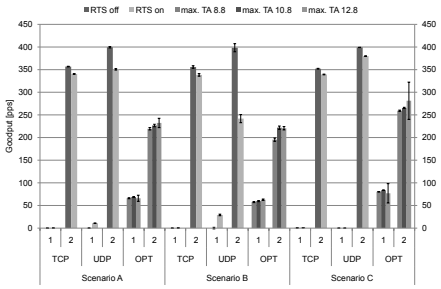
Figure 17: Simulation results for the 3 and 7 link chains.

forms better than UDP-DSR. Remember that the simulation scenario does not offer multi-user diversity, since the wireless links are almost free from errors. On the other hand, the proposed protocol relies on CSMA/HBT, so that the RTS/CTS exchange is mandatory. The IEEE 802.11 standard is not designed for the operation in a WMN, which is the main reason for the lower goodput. Using IEEE 802.11, the upstream nodes are unable to determine their transmission opportunities and probe the channel while their respective receivers are silenced. Thus, the medium access fails, i.e. there is no CTS or ACK, and the transmitters excessively back off. Note further that the RTS/CTS exchange introduces additional overhead, but does not solve the problem. The performance of the TCP variant suffers from the additional ACKs and from the specific problems if TCP in wireless networks [27], i.e. the implicit assumption of TCP that segment losses are due to congestion instead of channel impairments.

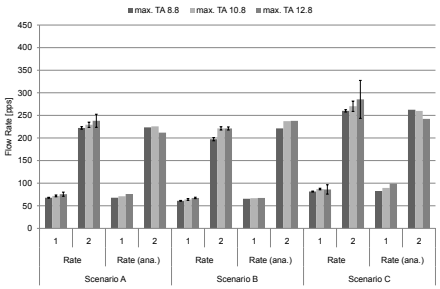
In Fig. 17b, we have plotted the efficiency factor K, the maximal TA and the flow rates, both estimated during the simulations, in comparison to the analytical flow rate for the same K. As the figure suggests, the system is able to operate the TAs near the anticipated working point specified by the target TA. In particular, the achieved TA at the bottleneck node becomes slightly smaller and more variable when increasing the length of the chain. Using the achieved Ks, we have calculated the flow rate in Maple using the idealized MAC model without probing. The analytical flow rates closely match the simulation outcomes. Thus, we conclude that even the idealized model has a high predictive value.

7.2. Fairness with Limited TAs

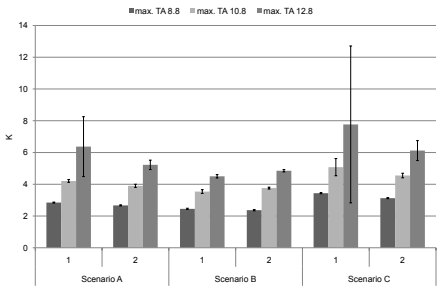
In this section, we illustrate the effectiveness of the inter-flow adaptation in a detailed network simulator. The inter-flow adaptation is designed to ensure proportional fairness in throughput between flows. In particular, we have simulated the two-flow scenarios in Fig. 13 (cf. section 6.2.2) using the simulation parameters from the previous section (cf. Table 6).



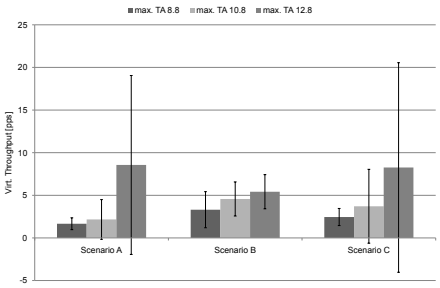
(a) Goodput of the proposed protocol in relation to DSR/TCP and DSR/UDP for flow 1 and 2 in scenario A, B and C ($\pm std.dev.$).



(b) Flow rate of flow 1 and 2 in relation to the analytical flow rate for the efficiency parameter K in the diagram below ($\pm std.dev.$).



(c) Efficiency parameter K as estimated by the inter-flow fairness adaptation ($\pm std.dev.$).



(d) Virtual throughput at the source of the longer flow ($\pm std.dev.$).

Figure 18: Simulation results for the three scenarios in Fig. 13 with 2 flows each.

From the goodput comparison in Fig. 18a, we conclude that the longer flow starves in most cases with DSR/UDP and DSR/TCP. On the other hand, the proposed protocol converges close to the proportional fair point, i.e. the efficiency parameters K of both flows match each other closely in Fig. 18c. Remember that we use a hysteresis to prevent deadlock situations (cf. Algorithm 3), so that small imbalances for the Ks of both flows are inevitable. In addition, the analytic results of the idealized model for the dynamically estimated Ks are close to the flow rates (cf. Fig. 18b).

However, the variability in the results grows when increasing the target TA to approach more efficient working points. This problem can be observed in scenario C for K and the goodput, for example, which exhibit a considerable standard deviation for a target TA of 12.8. This can be explained by the dynamic properties of the CSMA Markov chain. There is a tradeoff between efficiency and short-time fairness [77]. When requesting higher efficiency, the short-time fairness decreases and the links have to wait longer to access the medium. However, due to the delay and queue limits the link may not hold the channel sufficiently long to be able to achieve the requested throughput burst. Thus, the system compensates for the limited queue sizes

by transmitting virtual packets (cf. section 6.3.2), which we have observed in the simulation results. In particular, Fig. 18d illustrates the growing virtual throughput in relation to the requested efficiency.

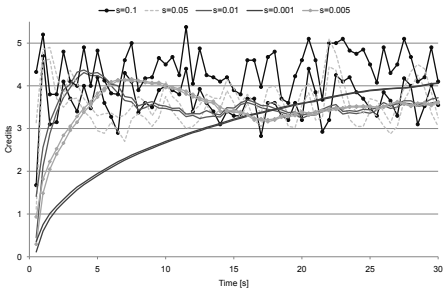
7.3. Step Size Adaption during and after Convergence

In the following, we evaluate the dynamic step size adaptation and its impact on the system during and after convergence. In particular, we illustrate the dynamics of the system during convergence using a detailed network simulator. Furthermore, we consider the length of the convergence and illustrate the influence of the step size adaptation in terms of variability and throughput efficiency after the system has converged. The simulation setup is comparable to the previous sections, and we highlight the differences only. The simulation topology is a chain of nodes with 2 and 7 links, respectively, which is traversed by a single flow for about 70 s. In the evaluation, we assume that the system converges within the first 40 s, and we use the remaining 30 s for the calculation of the steady state values ("after convergence"). The working point for the efficiency adaptation is set to $TA = 8.8$. We used an additional bit-rate of 12 Mbps.

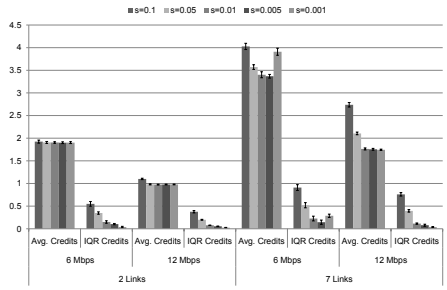
To start with, we investigate the dynamics of the system using fixed step sizes. In the following, we report the credits C and the step size s relative to the efficiency parameter K. Thus, the relative credits C/K can be directly compared across different step sizes. Furthermore, the relative credits are reciprocal to the flow rates and they are directly proportional to the queue lengths, where the step size is the proportionality constant. In Fig. 19a, we have plotted the evolution of the credit queues at the source for five individual simulations. We grouped the values into 0.5 s bins with about 50 entries each and plotted the first and third quartile of each bin in the diagram. As supposed, the speed at which the credit queues are filled heavily depends on the step size. For example, it takes more than 10 s for the credit queue to reach the working point with the smallest step size $s = 0.001$. On the other hand, the working point is reached within tens of milliseconds using the largest step size. However, the drawback is the higher variability of the credits queues at the working point, i.e. we observe a higher IQR with larger step sizes in the figure.

Considering the credit level after convergence in Fig. 19b, we observe that the credit queue shrinks and thus the flow rate tends to grow with decreasing step size, whereas the variability in terms of credit IQR declines at the same time. The reasons for this observation can be found in Fig. 19d. Due to the higher variability in the credit queues, it becomes more difficult for the system to meet the target TA in the intra-flow adaptation. Using the largest step size on the 7-hop chain, for example, the average TA after convergence is about 6 for 6 Mbps and even lower for 12 Mbps, whereas the target TA of 8.8 is almost exactly achieved with smaller step sizes. In the same line, the achieved efficiency parameters K tend to increase with smaller step sizes in Fig. 19d. One exception to the observation above is the smallest step size $s = 0.001$ with 7 links. However, the problem is that the 70 s duration of the flow is too short in relation to the necessary convergence time, which introduces a bias in the experimentation results.

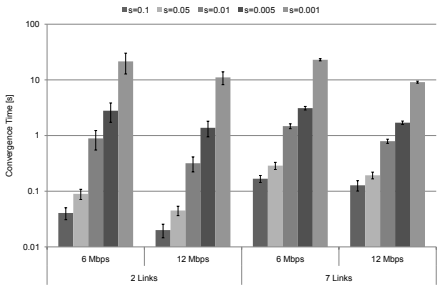
As introduced in section 6.3.1, the convergence time is the time from the start of



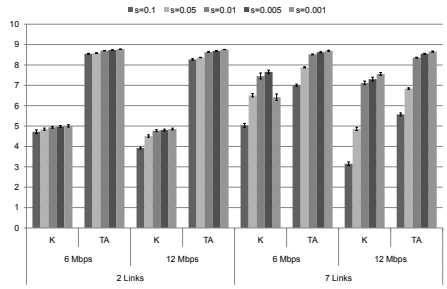
(a) Evolution of the relative credits (C/K) at the source of the flow depending on the step size (5 individual simulations with 7 links and 6 Mbps). The values are grouped into 0.5 s slices with about 50 values each. The lines show the $25th$ and $75th$ percentile.



(b) Credits at the source of the flow (Average of C/K after convergence and IQR of C/K, $\pm std.dev.$). Remember that the flow rate is reciprocal to the credits at the source.



(c) Convergence time ($\pm std.dev.$).



(d) Estimated K and max. TA (Average after convergence, $\pm std.dev.$).

Figure 19: Fixed step size: Simulation results for the chain scenario and a single flow.

the flow until the credit queue at the source breaks through the first quartile for the first time. As shown in Fig. 19c, the convergence time is increasing with smaller step sizes, and the results suggest that there is an inverse relationship between them. Additionally, we observe the impact of the bit-rate. Since the TAs are weighted by the bit-rate (cf. section 5.4), fewer credits are necessary to reach the efficiency working point with higher bit-rates (cf. Fig. 19b). Hence, the convergence time is reduced (cf. Fig. 19c). On the other hand, more TAs are associated with a single packet at higher bit-rates. Thus, the effects of variability are more pronounced (cf. Fig. 19d).

Let us consider the dynamic step size adaption in the following. In the simulations, the step size adaptation is controlled by three parameters: the initial step size s_0, the target step size s_∞ and the half-life t_{hl} (cf. Algorithm 4). Note that the target step size is denoted relative to K and relative to C, in addition. Thus, a target step size of $s_\infty = 0.01$ means that a packet is worth 1% of the amount of (relative) credits at the source. In contrast, the initial step size is given in relation to K only, i.e. it is worth

a fixed amount of (relative) credits not depending on the number of credits at the source.

We can think of the adaptation as a transition of the system from one step size configuration to another. In particular, the properties of the system after convergence are solely determined by the target step size. For example, refer to Fig. 20b and Fig. 20d. Both figures illustrate the characteristic tradeoff between efficiency and variability at the working point, which we have already encountered in the previous section. In contrast, we did not find a significant influence of the half-life and the initial step size on the system after convergence.

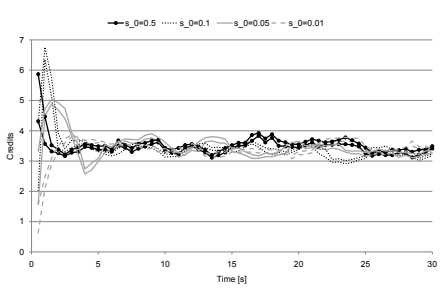
The initial step size and the half-life determine the dynamics of the system during convergence. For illustration purposes, we have picked four simulation instances and plotted the credit queue dynamics in Fig. 20a. Depending on the selection of s_0, the convergence time differs. The higher the initial step size, the faster the credit queues fill. Thereafter, the dynamics of all simulation instances becomes essentially equal, since it is dominated by an identical target step size and the influence of the initial step size vanishes. However, we also notice that the system may overshoot the working point during convergence for certain initial values. Note that this is not dramatic with respect to the TAs, since the system will outweigh this over-run by reducing the efficiency parameter K in a way that the aggregated TA on each hop does not exceed the technological limits. After the over-run has dissolved, the system raises K again to approach the requested efficiency. Looking at Fig. 20c, the convergence time for different initial step sizes is comparable to the results with fixed step sizes in Fig. 19c.

The influence of the half-life is harder to quantify. Using the above presented indicators, the half-life did not have a significant impact. However, in the simulations we observed that a small half-life may lead to an uneven filling of the credit queues. The queues close to the source fill quickly, but the step size is reduced too fast so that the queues close to the destination need substantially more time to converge. From the simulation results, we found that a half-life $t_{hl} = 0.2\,\mathrm{s}$ leads to an acceptable dynamics across all queues in the scenarios of interest.

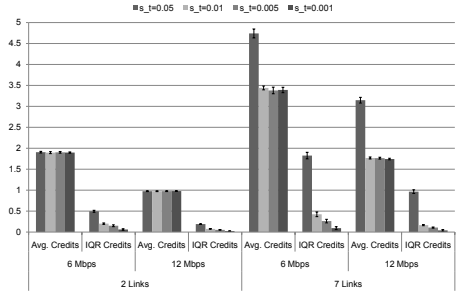
In summary, we have seen that the dynamic step size adaptation combines the advantages of both small and large step sizes, i.e. the convergence time is considerably reduced and the variability at the working point is low and controllable. In particular, we have investigated the impact of the individual parameters during and after the initial convergence. On the other hand, re-adaptation was out of scope for our investigation above. For the implementation of a system under real-world conditions, re-adaptation may be necessary in case of arriving and departing flows or significant environmental changes.

7.4. End-To-End Delay and Short Time Fairness

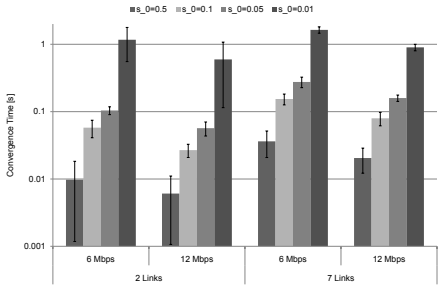
In the following, we investigate the delay properties and the short time fairness of the proposed system. In particular, we illustrate the fundamental tradeoff between throughput efficiency and short time fairness [76, 77], which translates to a tradeoff between end-to-end delay and throughput efficiency. We show the extent to which our concept of virtual transmissions is able to mediate between delay and efficiency,



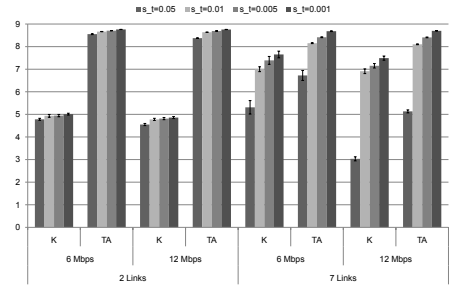
(a) Evolution of the credits (C/K) at the source of the flow depending on the initial step size s_0 (7 links, 6 Mbps, $s_\infty = 0.005$, $t_{hl} = 0.2$ s). The values are grouped into 0.5 s slices. The lines show the $25th$ and $75th$ percentile.



(b) Credits at the source of the flow in relation to the target step size s_∞ (Average of C/K after convergence and IQR of C/K, $s_0 = 0.1$, $t_{hl} = 0.2$ s, $\pm std.dev.$).



(c) Convergence time in relation the initial step size s_0 ($s_\infty = 0.005$, $t_{hl} = 0.2$ s, $\pm std.dev.$).



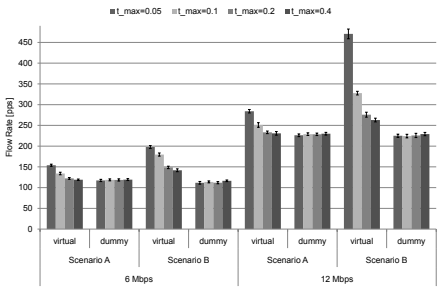
(d) Estimated K and max. TA in relation to the target step size s_∞ (Average after convergence, $s_0 = 0.1$, $t_{hl} = 0.2$ s, $\pm std.dev.$).

Figure 20: Adaptive step size: Simulation results for the chain scenario and a single flow.

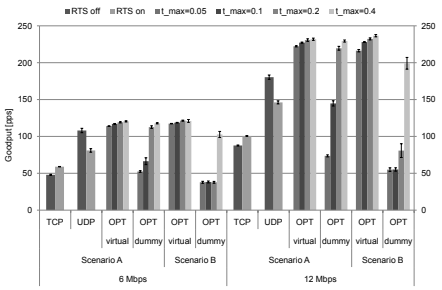
and we take a closer look at the arising tradeoff between multi-path (spatial) diversity and delay.

In our simulative investigation, we explore the scenarios introduced in section 6.3.2 in detail: Scenario A and B consists of a 7-link chain and a double chain of 2x7 links, respectively. The displacements are set in way that one-hop neighbors experience error-free links and two-hop neighbors do not directly interact with each other. We have set the lower queue limit to four packets and varied the per-hop delay limits (cf. Algorithm 5) across all simulations. In addition, we repeated the simulations with dummy packets [48] instead of virtual transmissions. The remaining simulation parameters are comparable to the above sections (cf. Table 6). The results are shown in Fig. 21, and we will go through them in the following.

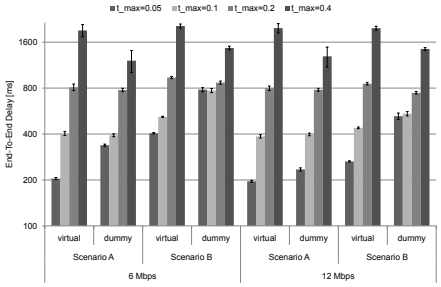
To start with, let us consider the influence of the per-hop delay limits. As supposed, a reduction of the delay limits leads to a smaller amount of in-flight packets



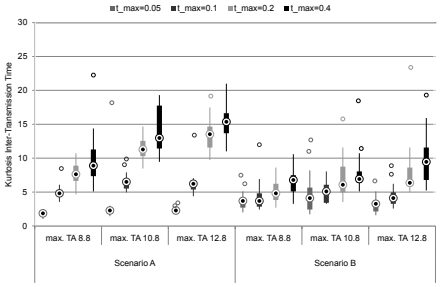
(a) Flow rate for different bit-rates and both virtual and dummy transmissions (max. TA 8.8, $\pm std.dev.$).



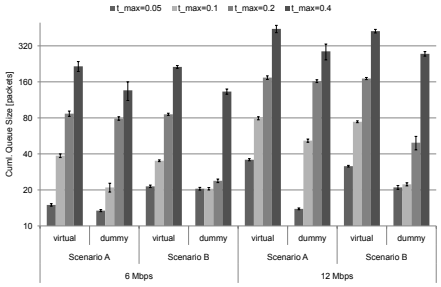
(b) Goodput in relation to DSR/TCP and DSR/UDP (max. TA 8.8, $\pm std.dev.$).



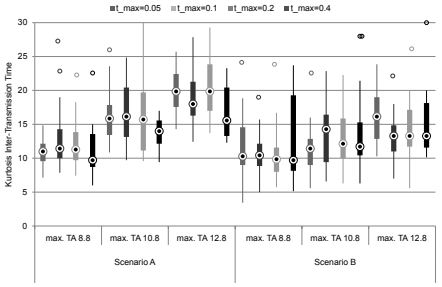
(c) End-to-end delay (max. TA 8.8, $\pm std.dev.$).



(d) Kurtosis of the inter-transmit time distribution on (one of) the last hop(s) for 6 Mbps and virtual transmissions only (Box plot across all seeds).



(e) Average queue size, cumulated across all nodes (corresponds to the average number of in-flight packets, max. TA 8.8, $\pm std.dev.$).



(f) Kurtosis of the inter-transmit time distribution on (one of) the last hop(s) for 6 Mbps and dummy transmissions only (Box plot across all seeds).

Figure 21: Simulation results for a single flow in scenarios A and B from section 6.3.2.

(cf. Fig. 21e), and the related end-to-end delays are reduced (cf. Fig. 21c). Remember that Algorithm 5 calculates the queue limits in order to meet an average per-hop queueing delay. Thus, for the same per-hop delay, the queue limits for 12 Mbps are generally twice the limits for 6 Mbps in terms of packets. In Fig. 21b, we notice that the goodput efficiency increases with higher delay limits, whereas the end-to-end delays rise at the same time (cf. Fig. 21c). For comparison purposes, we have plotted the goodput results for DSR/UDP and DSR/TCP in Fig. 21b. However, their delay results are meaningful only to a limited extent. In the simulations, we have set the maximum window size for TCP to 10 packets. The resulting delays are 70.7 ms (73.5 ms) in scenario A and 48.7 ms (44.8 ms) in scenario B, respectively, without (and with) RTS. For UDP, we have limited the number of in-flight packets to 600. The average delays are 540 ms, 3800 ms, 370 ms and 3200 ms in the above ordering.

In the following, we consider the influence of virtual and dummy transmissions. Fig. 21a and Fig. 21b illustrate the estimated flow rate and the achieved goodput, respectively. Remember that the flow rate and the (relative) credits at the source are reciprocal. Furthermore, the difference between flow rate and goodput is the amount of virtual (or dummy) traffic. From the figures, we conclude that the goodput increases with higher per-hop delay limits, and the amount of virtual and dummy traffic is reduced at the same time. As supposed, with very large per-hop delay limits the behavior of the system is essentially equal for both virtual and dummy transmissions, since the risk of queue over- and under-runs is low. However, with virtual transmissions the system is able to maintain a high goodput even with small per-hop delays, which comes at the expense of a higher deviation from the working point of the flow rate. With dummy transmissions, on the other hand, the system accurately maintains the working point of the flow rate at the expense of severe goodput degradations for small per-hop delay limits.

At first sight, the effect of virtual and dummy transmission on the end-to-end delay seems contradictory. With high per-hop delay limits, the number of in-flight packets is lower with dummy transmissions (cf. Fig. 21e), and as Little's law suggest, the delay is lower at the same time (cf. Fig. 21c). However, with small per-hop delays, virtual transmissions achieve a lower end-to-end delay with no less or even more in-flight packets. This behavior can be explained with the tradeoff between throughput efficiency and short-term fairness [24, 76, 77]: The more efficient the working point of UO-CSMA is, the longer the nodes have to wait to gain access to the channel and the longer they have to hold the channel or, in other words, the higher the burstiness of the inter-access times. For (one of) the last hop(s), we have estimated the distribution of the inter-transmission times and plotted the excess kurtosis in Fig. 21d and Fig. 21f for virtual and dummy transmissions, respectively. Remember that the higher the kurtosis, the more weight the tail of the distribution has in relation to the normal distribution. In particular, the normal and the exponential distributions have an excess kurtosis of 0 and 6, respectively. With dummy packets, the kurtosis increases with the target TA, i.e. the traffic gets more bursty, as supposed. But the kurtosis is virtually independent from the per-hop delay limits (cf. Fig. 21f). The per-hop delay imposes limits on each queue, and if the queue cannot sustain the required burst length, the node has to emit dummy packets. With virtual transmissions, on the other

hand, the kurtosis decreases with smaller per-hop delay limits. If the queue cannot sustain the required burst length due to the imposed hop-delay limits, only a virtual transmission is initiated and the node releases the channel. Thus, the per-hop delay limits control the burstiness and the short-term fairness, which comes at the expense of throughput efficiency according to the underlying tradeoff (cf. Fig. 21b). This way, the contradiction above can be resolved: With virtual transmissions, the smaller per-hop delay limits reduce both the length of the packet bursts and the waiting time for the medium access due to the higher short-term fairness, so that the individual packet traverses the network faster.

In summary, we have illustrated how the per-hop delay limits are able to control the amount of in-flight packets and the related end-to-end delay. However, the resulting queue limits may not be sufficient to sustain the required burstiness caused by the throughput-delay tradeoff. In this case, we have to sacrifice throughput with either dummy or virtual packets. Using dummy packets, the working point of the tradeoff and thus the requested short-term fairness is *maintained*, and the system excessively emits dummy packets in case its queues cannot sustain the requested burst lengths. With virtual transmissions, on the other hand, the system actively controls the short-term fairness in order to reduce the burstiness to a sustainable level for the imposed per-hop delay limits, i.e. the system approaches *another working point* of the tradeoff. From our simulation results, we conclude that virtual transmissions cope better with small per-hop delay limits. In particular, the loss in goodput due to reduced delay limits is gradually, and it is significantly smaller than with dummy packets, whereas the delay performance is even better. On the other hand, virtual traffic introduces a bias in the estimation of proportional fair flow rates, which may be detrimental to the fairness between flows (cf. section 6.3.2).

In the following, let us consider the difference between scenario A and B in order to illustrate the effect of multi-path routing. Note that both scenarios are identical in terms of spatial diversity, i.e. the collision domains are identical whereas only the number of nodes and links differs. In particular, we focus our observations to virtual transmissions only. As shown in Fig. 21b, the difference in goodput is only marginal between both scenarios. On the other hand, the flow rates in Fig. 21a show a larger bias in scenario B especially for small delay limits. In the same line, the end-to-end delays for small delay limits are generally higher with multi-path routing (cf. Fig. 21c), whereas the number of in-flight packets is of the same order in both scenarios (cf. Fig. 21e). With a per-hop delay limit of 50 ms, there is a difference between scenario A and B for the in-flight packets at 6 Mbps. During the simulations, we ensured that the maximal queue size does not under-run the limit of 4 packets in order to keep the nodes operational, which is most likely the reason for the mentioned exception.

In both scenarios, the spatial diversity is equal. Thus, the achievable end-to-end goodput is essentially the same. However, due to the additional chain of links in scenario B, the flow splits up between them evenly and halves the rate of incoming traffic. Thus, the maximal queue size per relay is also halved according to Algorithm 5, or the minimum of 4 packets applies. Thus, the risk of virtual transmissions increases with smaller delay limits, which explains the larger deviation of the flow rates from

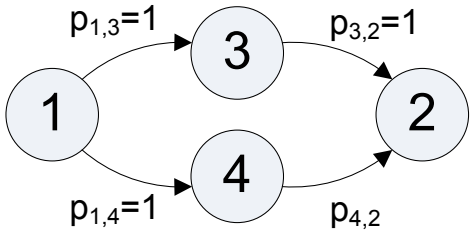


Figure 22: Simulation topology.

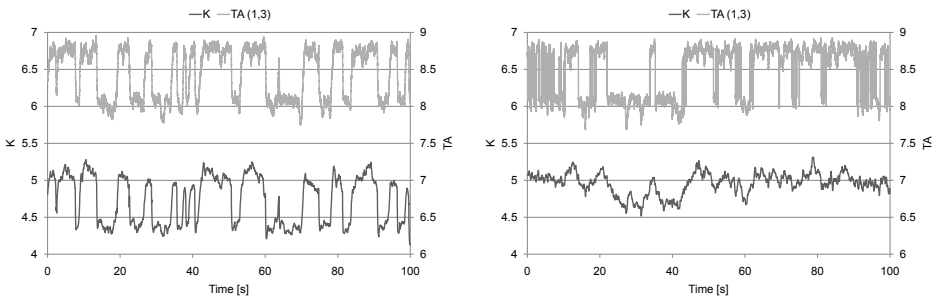
the actual working point with multi-path routing in Fig. 21a.

To summarize, we have observed a tradeoff between multi-path usage and throughput and delay. By using multiple paths, the system may achieve additional spatial diversity to mitigate potential bottlenecks. On the other hand, the traffic has to split up for additional routing paths. Thus, the per-hop queue limits decrease, so that the risk of queue over- and under-runs grows and virtual transmissions become more likely. Furthermore, if the rate of a sub-flow becomes too small and the lower bounds for the per-hop queue limits apply, our delay limiting approach becomes ineffective and the delay on that path cannot be controlled anymore. In section 6.5, we have presented a heuristics that considers the tradeoff by limiting the relay count stretch. However, a system that controls both the routing sub-graph and the delay limits at the same time should be able to achieve an even broader range of throughput-delay working points, and we plan to consider this in our future work. For example, low-rate flows should be focused on a small amount of essential routes only. On the other hand, high-rate flows may extend their routing stretch to achieve additional spatial diversity or to limit the interaction with neighboring flows.

7.5. Candidate Set Dynamics

In this section, we illustrate the oscillation problems within the intra-flow adaptation that arise with candidate set dynamics. We demonstrate our solution to that problem, in which we compensate the TAs for the dynamics in the candidate set as described in section 6.4. Fig. 22 shows a network topology, in which oscillations may arise with any-casting if the PSR on link $(4, 2)$ is small. In particular, we vary $p_{4,2}$ from $0.1 \ldots 0.3$, whereas the PSRs of the remaining links are 1. A traffic flow is set up between node 1 and 2. The flow persists for 200 s, and we present results from the latter 100 s only in the following. The remaining simulation parameters are similar to the previous sections (cf. Table 6). Since all nodes are within the same collision domain, we can trade retransmissions on link $(4, 2)$ against transmissions on link $(3, 2)$ that are successful in any case. Thus, only the upper route involving node 3 should be used at the optimum.

The results for one particular simulation seed are plotted in Fig. 23. In the left diagram, the compensation is turned off. As supposed, the credits of node 4 and 1 almost alike, so that node 1 alternates between inclusion and exclusion of node 4 in



(a) Without compensation.

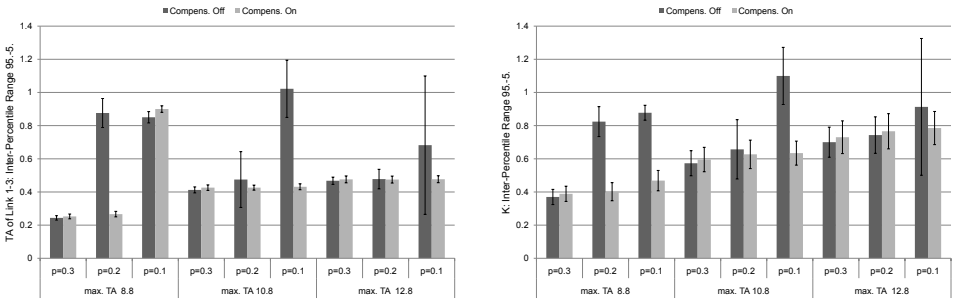
(b) With compensation.

Figure 23: Efficiency parameter K and the TA of link $(1,3)$ over time for one particular seed ($p_{4,2} = 0.1$, max. sum-TA 8.8).

its routing decision. On including node 4, it additionally activates the links $(1,4)$ and $(1,\{3,4\})$. The latter hyperlink and link $(1,3)$ have almost equal TAs by construction, so that the cumulated TA significantly exceeds the specified limit. Thus, the intra-flow adaptation has to decrease K in order to drive the cumulated TA of node 1 to its target value. In the same line, the efficiency parameter K has to be re-raised to its former level when node 4 is excluded from the routing decision, which leads to the oscillation of K in Fig. 23a. The rationale of the compensation heuristics is to distribute the already available TA among the affected hyperlinks, so that the intra-flow adaptation remains almost unaffected. Thus, the TA of compensated link $(1,3)$ in Fig. 23b oscillates similar to the uncompensated link. However, the variability of K in the diagram is significantly reduced.

In the following, we measure the variability using the difference between the 95. and the 5. percentile. The inter-percentile ranges of the efficiency parameter K and the TA of link $(1,3)$ across all seeds are shown in Fig. 24. The results indicate that the example above is characteristic for the given parameters. However, the variability of the TA of link $(1,3)$ is considerably reduced when increasing the target sum-TA or when increasing the PSR on link $(4,2)$. In particular, we have observed that the routing decision of node 1 becomes more stable in both cases, i.e. the TCs assigned to node 4 run empty less often and the oscillation problem is less severe.

Nevertheless, there is a further difference between the compensated and uncompensated versions. As shown in Fig. 23, the intra-flow efficiency adaptation ensures that the compensated and the uncompensated TAs of the link $(1,3)$ (and $(1,\{3,4\})$, of course) are approximately equal in average. However, we apply the compensation to link $(1,4)$ in addition, which increases the TC of node 4 and thus makes the routing decision with compensation more stable. In consequence, the routing decision in the compensated version alternates only for the parameter set $p_{4,2} = 0.1$ and maximum sum-TA 8.8, as shown in Fig. 24a. In summary, the compensation approach significantly reduces the oscillation within the intra-flow adaptation, as indicated by the lower variability of the Ks in Fig. 24b.



(a) Inter-Percentile Range (95.-5.) of the TA of link $1-3$ ($\pm std.dev.$).

(b) Inter-Percentile Range (95.-5.) of the efficiency parameter K ($\pm std.dev.$).

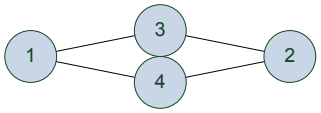
Figure 24: Variability of TA and K for the diamond topology in Fig. 22.

7.6. Multi-User Diversity

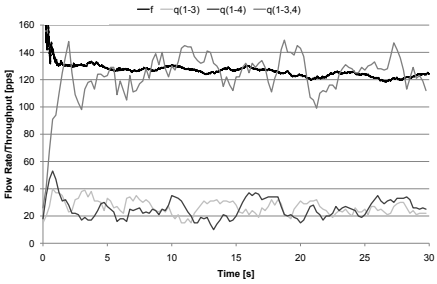
In the following, we illustrate how the system detects and utilizes MUD in the form of anycast transmissions to enhance the end-to-end goodput in fading environments. In particular, we use a block-fading channel model, in which the signal undergoes a parameterized Rician fading [37, 94] without temporal correlation. For the ease of illustration, fading is not applied to signaling traffic, so that the underlying LCGs are stable in time. Due to the missing temporal correlations in the considered channel model, an informed scheduling strategy that considers the instantaneous channel conditions has no advantages over channel-blind strategies relying on statistical channel knowledge only. For example, the considered channel model may be encountered in systems with short channel coherence times or in systems without transmitter channel state information (CSI) like IEEE 802.11.

The simulation topology consists of 4 nodes forming two hops for a single flow from node 1 to node 2. The first hop has two possible receivers (cf. Fig. 25a). Hence, the topology offers MUD but no spatial diversity, since at most one link can be activated at each time instance. The PSR is about 50% (75%) for unicast (anycast) links. The target step size and the per-hop delay limit are 0.001 and 0.6 s, respectively. The remaining parameters can be found in Table 6.

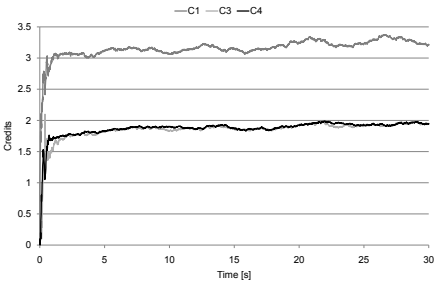
In Fig. 25, we have illustrated the evolution of the system for a particular simulation instance. The credits reach the proximity of the working point quickly (cf. Fig. 25c). Since both routes offer the same throughput, the credits of the relay nodes are almost equal. Thus, the TC for both receivers $\beta_{1-3,4} = \alpha_1 - \max(\alpha_3, \alpha_4)$ is nonzero whereas the TCs of the individual receivers $\beta_{1-3} = [\alpha_4 - \alpha_3]_+$ and β_{1-4} vary close to zero. The system weights the TCs with the PSR to get the TA in Fig. 25d. Remember that the PSRs have to be determined from measurements and thus, they are subject to measurement noise. We conclude that the system prefers the anycast link, since it has the highest TA. Furthermore, we observe that the anycast link achieves substantially more throughput on the first hop compared to the unicast links (cf. Fig. 25b).



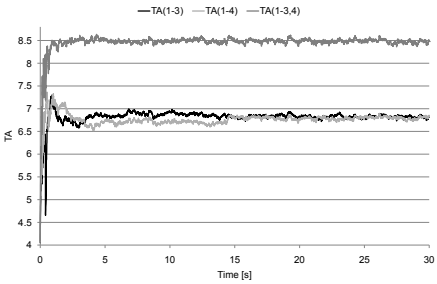
(a) Simulation topology with a flow from 1 to 2. The PSR is 50% (75%) for unicast (anycast) links.



(b) Flow rate and throughput on the first hop (moving average over 1 s).



(c) Evolution of the (relative) credits.

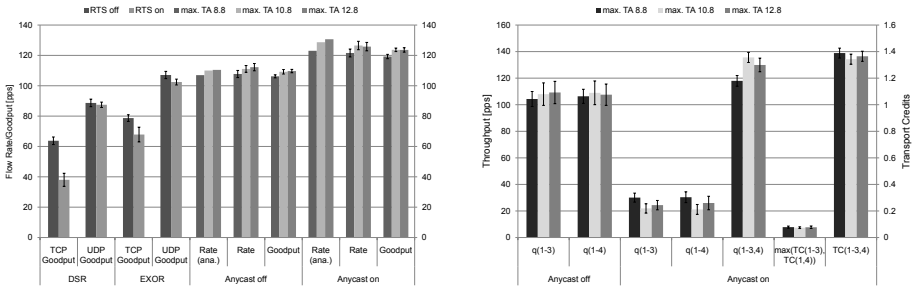


(d) Evolution of the TAs.

Figure 25: Simulation topology and evolution of the system for an individual simulation instance (max. TA 8.8).

And in fact, the analytical model suggests that using the anycast link only on the first hop is the best solutions. However, the system approaches the optimum only asymptotically, and in the finite regime it cannot completely prevent that unicast links are used on the first hop. Solving the problem for the idealized CSMA model in Maple for the same system parameters and $K = 4.5$ as estimated in the simulation, the average throughput on the anycast and unicast link is $q_{1-3,4} = 131$ pps and $q_{1-3} = q_{1-4} = 26.5$ pps, respectively, and the resulting flow rate is $f = 125$ pps. Hence, the local decision of the system to prefer the anycast link has turned out right in the end.

Furthermore, the scenario illustrates how the system detects opportunities to benefit from MUD. From the point of view of the network layer, there is no difference in using the anycast link or the unicast links, since the credits of both downstream nodes are equal. However, while determining the TAs, the MAC layer locally adjusts its preferences according to the PSR of the considered links (cf. Fig. 25d). In particular, the MAC does not specify the best (hyper)-links in advance, since this would involve global knowledge. Instead, *all links* on the first hop have to *compete against each other* for medium access. The structure of the problem ensures that the resulting



(a) Flow rate and goodput ($\pm std.dev.$).

(b) Throughput and transport credits on the first hop ($\pm std.dev.$).

Figure 26: Simulation results for the network in Fig. 25a.

solution only gradually deviates from the optimum in the underlying tradeoff. On the other hand, the optimum can be approached by increasing the efficiency parameter K, which linearly scales the TAs in the same way. Hence, the TA is increased on all links, but the absolute difference between anycast and unicast TA increases, so that the preference of the system for the anycast link becomes more pronounced.[8] Thus, throughput is reallocated from the unicast links to the anycast link. For example, if we increase the efficiency to $K = 6.2$ in the analytic model, so that the cumulated TA is 10.8 on the first hop, the throughput changes to $q_{1-3,4} = 149$ pps and $q_{1-3} = q_{1-4} = 17.2$ pps, as supposed. The simulated system along with its imperfectness still achieves 138 pps and 21 pps, respectively. Fortunately, in our case the *local* decision coincides with the *global* solution. In the next section, we will illustrate what happens if local and global views disagree.

We now generalize our view from a particular instance to a set of simulations, in which we varied the target TA for intra-flow efficiency adaptation. In addition, we turned off the anycast operation for a subset of simulation runs and we compare the performance against UDP and TCP on top of DSR and extremely opportunistic routing (ExOR) [8, 140]. In Fig. 26a, we have plotted the flow rate and goodput for the considered protocols. DSR achieves the lowest goodput. Nevertheless, remember that it is a single-path routing protocol, which does not benefit from MUD. ExOR on the other hand performs better. However, it uses anycast on top of a slightly modified IEEE 802.11. Without anycast, the proposed protocol only uses multi-path, which has no advantages in the considered scenario. In particular, without anycast the goodput of the protocol with single-path (not shown) and multi-path routing is almost identical. And with anycast turned on, the protocol increases its rate and goodput due to the available MUD. As expected, requesting higher efficiency via larger target TAs results in higher goodput. In either case, the simulation results show a good match with the analytic predictions of the idealized CSMA model in Maple. In the next paragraph, we will consider why the deviations from the model

[8] Remember that the TAs are the transition rates of the CSMA Markov chain in *exponential* scale.

are higher with anycast and high target TA. In the analytical model, we estimated an optimal flow rate of $f = 132.6$ pps in the very high efficiency regime. However, the benefits of anycast MUD seem to lag behind the expectations at first sight. The problem is that only the first hop can benefit from anycast in our case. Due to the transmissions saved on the first hop, we can inject additional traffic into the network, but we have to use the costly and less efficient transmissions on the second hop.

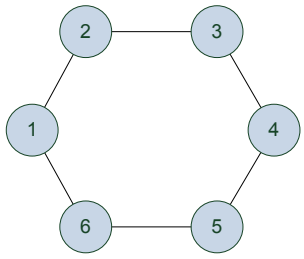
In the paragraphs above, we gained the insight that the efficiency of the candidate selection improves when we increase the target TA. However, when using the highest target TA of 12.8, the system is not able to reallocate further throughput from the unicast to the anycast link (cf. Fig. 26b). In addition, in the above example, we have seen that the throughput of the anycast link is generally lower than the analytic predictions, whereas the throughput on the unicast links is higher. We identified several factors that contribute to the described behavior. Virtual transmissions are one of them: The risk of a virtual transmission is higher for anycast links, since a queue overflow at a single receiver will cause a virtual transmission for the whole anycast link. Note further that virtual transmissions[9] are rare due to the conservative selection of the per-hop delay limits. Nevertheless, they are present. Furthermore, the transmitters have to estimate the PSRs from active measurements. Thus, the resulting TAs are subject to the PSR measurement noise, in addition. Since fading is a random process, there is uncertainty within the channel. In addition, the channel may corrupt signaling frames (RTS, CTS, ACK). As with virtual transmissions, the risk of signaling errors is higher for anycast links, so that the anycast link may not be able to take all its TXOPs.

A further reason is the variability of the TCs and the TAs, which lead to short-term imbalances between paths. The analytic model suggests that on the first hop, only the anycast link should have non-zero TCs. In Fig. 26b, we have plotted the TC of the anycast link and the maximum TC of the unicast links over time. In the simulations, the unicast TCs are small, but they are not zero due to the dynamics of the credit queues. Note that this issue is closely related to both the intra-flow and the step size adaptation. In particular, requesting higher throughput efficiency increases the burstiness of the traffic at the same time. Hence, the imbalance problem between paths will get more serious. On the other hand, a smaller step size will reduce the variability and the imbalances to a certain degree.

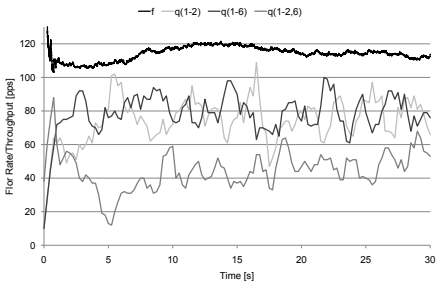
7.7. Spatial Diversity

Depending on the topology, spatial and multi-user diversity may not be usable at the same time. In this section, we illustrate how the system detects the form of diversity that performs best. The considered topology in the last section provides multi-user but no spatial diversity. On the first hop, the transmitter has only considered the PSR of the involved links and decides to prefer the anycast link. In the end, this *local* decision has turned out to be correct for the *global* performance, as well. In

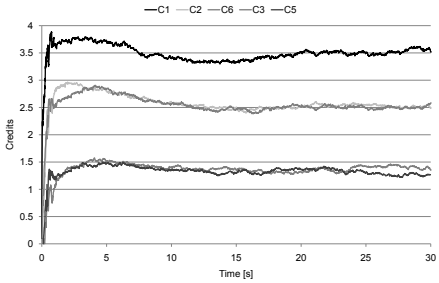
[9] Virtual transmissions can be roughly estimated by considering the difference between flow rate and end-to-end goodput.



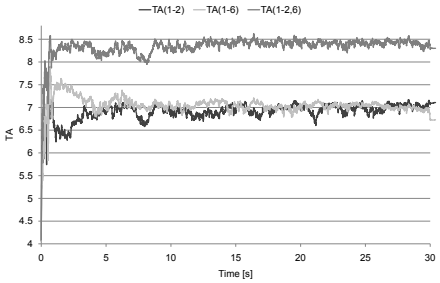
(a) Simulation topology with a flow from 1 to 4.



(b) Flow rate and throughput on the first hop (moving average over 1 s).



(c) Evolution of the (relative) credits.

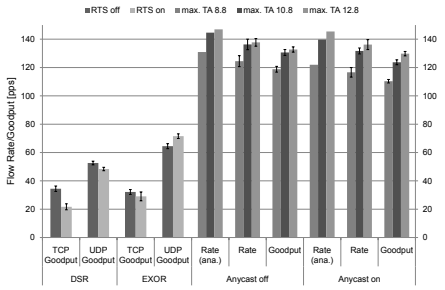


(d) Evolution of the TAs at the first hop.

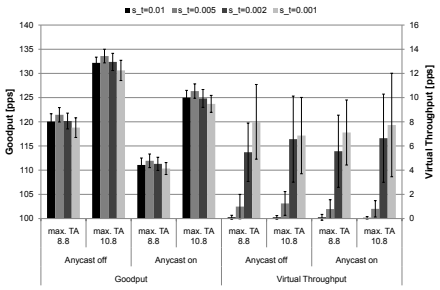
Figure 27: Simulation topology and evolution of the system for an individual simulation instance (max. TA 8.8, $s_\infty = 0.001$).

the following, we consider the hexagonal topology in Fig. 27a for a single flow from node 1 to node 4. From the perspective of node 1, which is the transmitter on the first hop, the topology is equivalent to the topology in the last section. However, spatial diversity is available since there are multiple collision domains. For each unicast link there is another unicast link that can be activated at the same time (e.g. link $(2, 3)$ and $(6, 5)$). In contrast, the anycast link $(1, \{2, 6\})$ can only operate in isolation. Assuming a PSR of 50% (75%) on a unicast (anycast) link as above, the local decision of node 1 to prefer the anycast link is *in opposition* to the globally optimal strategy, which activates the unicast links only. The remaining simulation parameters are comparable to the previous section (cf. Table 6).

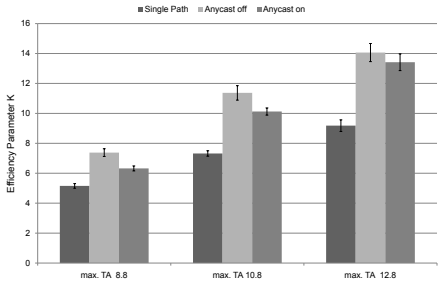
In the considered simulation instance in Fig. 27, both credits and TAs evolve comparable to the rhomboid network in the previous section (cf. Fig. 27c, Fig. 27d). However, the throughput distribution on the first hop is different. The anycast link still achieves a considerable amount of throughput, but the unicast links dominate (cf. Fig. 27b). On first sight, this seems to be contradictory since the TAs are almost identical to the rhomboid network and thus, node 1 still prefers the anycast link in terms of TA. The difference is within the topology. For example, if link $(3, 4)$ is currently



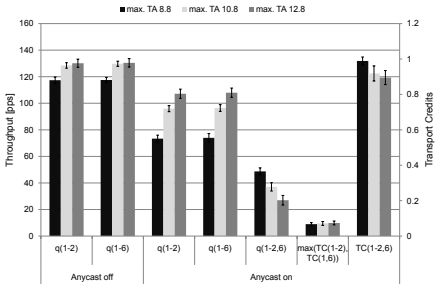
(a) Flow rate and goodput ($\pm std.dev.$).



(b) Goodput and virtual throughput in relation to s_∞ ($\pm std.dev.$).



(c) Efficiency Parameter K ($s_\infty = 0.001$, $\pm std.dev.$).



(d) Throughput and transport credits on the first hop ($\pm std.dev.$).

Figure 28: Simulation results for the network in Fig. 27a.

active, then only link $(1, 6)$ would be allowed to transmit concurrently. Nonetheless, node 1 is not aware of that and may probe link $(1, 2)$ or the anycast link. In this case, the probe fails since (one of) the receiver(s) is occupied and the link will be blocked (cf. section 3.1). On the other hand, the backoff of link $(1, 6)$ will eventually expire, so that both unicast links are concurrently transmitting, which takes this throughput away from the anycast link. The anycast link can only be activated in the less likely case that all other nodes are idle.

Interestingly, the performance of ExOR is only slightly better than DSR (cf. Fig. 28a). Remember that DSR is a single-path routing protocol, whereas ExOR uses multiple paths. But ExOR is unable to exploit spatial diversity in the considered scenario, since it uses anycast only on the first hop. On the other hand, the proposed protocol performs significantly better in comparison, because it limits its anycast usage (cf. Fig. 28d).

When increasing the target TA as shown in Fig. 28a, the system achieves a more efficient working point in terms of flow rate and goodput. We have determined an analytic flow rate of $f = 147.3$ pps in the very high efficiency regime, which is achieved without anycast. In Fig. 28d, we observe that the system gradually decreases the

throughput of the anycast link when increasing the target TA. This can be explained as follows. As in the previous section, the TAs are linearly scaled in the same way as the requested efficiency increases. However, the system persists longer in the states of maximal spatial reuse, which are the MISs of the LCG [24, 72]. Thus, the anycast link is not able to benefit from higher TAs in the same way as the unicast links, since the blocking gets more severe.

Starting from *identical initial conditions* and taking the *same actions*, the system implicitly identifies the type of diversity to use based on the differences in the underlying LCG. However, the system needs to actually probe and use *all* available (hyper)-links in order to identify its opportunities. In particular, no link can be excluded in advance, since the system identifies potential spatial or multi-user diversity from the answer of the network to its own input. However, the more (hyper)-links the system uses, the faster the per-node target TA is reached in the intra-flow adaptation (cf. section 6.2.1). In the considered scenario, the anycast link should not be used at best. But as shown in Fig. 27d, for example, it contributes the most to the cumulated per-hop TA. By excluding the anycast link, the TAs on the unicast links can be increased without violating the sum TA. Thus, the resulting efficiency factor K can be increased when not using anycast (cf. Fig. 28c), which also explains the higher analytic goodput in Fig. 28a when anycast is not used. The underlying cause is the increase of the optimality gap (4) of UO-CSMA within the number of involved links (cf. page 13). Nonetheless, an interesting question for our future work is whether heuristics can be developed to decide when to use anycast.

Furthermore, we estimated the influence of the target step size on the performance of the proposed protocol. Interestingly, the highest goodput is achieved with a medium step size of 0.005 (cf. Fig. 28b). We suppose that the limitations on the burstiness of the underlying network dynamics are responsible. When decreasing the target step size, the variability of the credit queues decreases, so that the short-term imbalances between different routing paths are reduced. However, virtual transmissions emerge when decreasing the target step size further, so that the overall performance suffers (cf. Fig. 28b). We suppose that the smaller step size might render it more difficult for the system to outweigh short-term imbalances, e.g. due to noise in the PSR measurements. Nevertheless, this point needs further investigation in our future work.

8. System Level Evaluation

After we have illustrated the operation of the system within small and synthesized topologies in the previous section, we now evaluate the system in larger scenarios. We are particularly interested in the following questions: Does it pay off to design a cross-layer MAC especially for WMNs? And if so, what potential for OR is left to a dedicated mesh MAC? To answer these questions, we have conducted network simulations using a detailed WMN simulator[10] to compare the performance of the proposed protocol and state-of-the-art protocols across a large set of randomized topolo-

[10] See `http://sarwiki.informatik.hu-berlin.de/Brn.Sim` for details on the Brn.Sim simulator.

Parameter	Value
Area	square $\{300\text{ m}, 600\text{ m}, 500\text{ m}, 1000\text{ m}\}$[1]
Topology	random with $\{30, 120, 50, 200\}$[1] nodes
Inter-node distance (min)	$\{30\text{ m}, 30\text{ m}, 50\text{ m}, 50\text{ m}\}$[1]
Path loss	Log-distance with exponent 3
Radio frequency	2.4 GHz
Channel	AWGN
Fading	i.i.d. Rician $K = \{0, 6, 12, 1000\}$
Reception model	SNR-PER via fading outages
Interference	Protocol model (no cuml. interference)
PHY	IEEE 802.11ag, $\{6, 12, 24, 48\}$ Mbps
SNR thresholds	$\{5.4, 7.0, 11.3, 18.6\}$ dBm
TX power	19 dBm
Noise floor	-92.965 dBm
CCA threshold	-88.5 dBm
Routing metric	ETT
Metric stretch	$\{1.05, 1.1, 1.3, 1.5, 1.8\}$[2]
Hop stretch	$\{1.5, 2.5, 3, 4, 5\}$[2]
Seeds	100

[1] The same indices go together, e.g. 300 m is used with 30 nodes and 30 m spacing, 500 m is used with 50 nodes and 50 m, and so on.
[2] The same indices go together.

Table 7: System Parameters for the Analytic Evaluation

gies within different configurations and environments. To start with, we evaluate the potential of spatial and multi-user diversity on random topologies. We have solved the underlying optimization problem numerically for a large set of mesh topologies in order to get a better understanding of the multi-path and anycast gains we can expect.

Table 7 and Table 8 summarize the parameters for the analytic and the simulative evaluation. Most importantly, the fading process is i.i.d. in time, so that the channel is memoryless. We applied fading to data frames only, since signaling frames are generally more robust against channel impairments because they are shorter and they are often transmitted using a lower bit-rate. Each configuration is repeated 100 times with different random seeds. Note that the node placement changes with each random seed. In order to avoid a clustering of nodes within certain areas, we constrain the distance between neighboring nodes to a specified minimal spacing.

8.1. Potential of Multi-User and Spatial Diversity on Random Topologies

In this section, we evaluate the potential of multi-user and spatial diversity within the considered topologies assuming UO-CSMA operates at the asymptotical optimum,

i.e. it achieves the optimal TDMA performance. We solve the opportunistic multi-commodity flow problem [100, 134] for a large set of topologies using a numerical solver and, in addition, we put the results in the context of our route pre-selection heuristics to illustrate its effectiveness (cf. section 6.5). In the following section, we will consider the extent to which the proposed protocol can realize the prospected potentials at typical operating points in the finite regime.

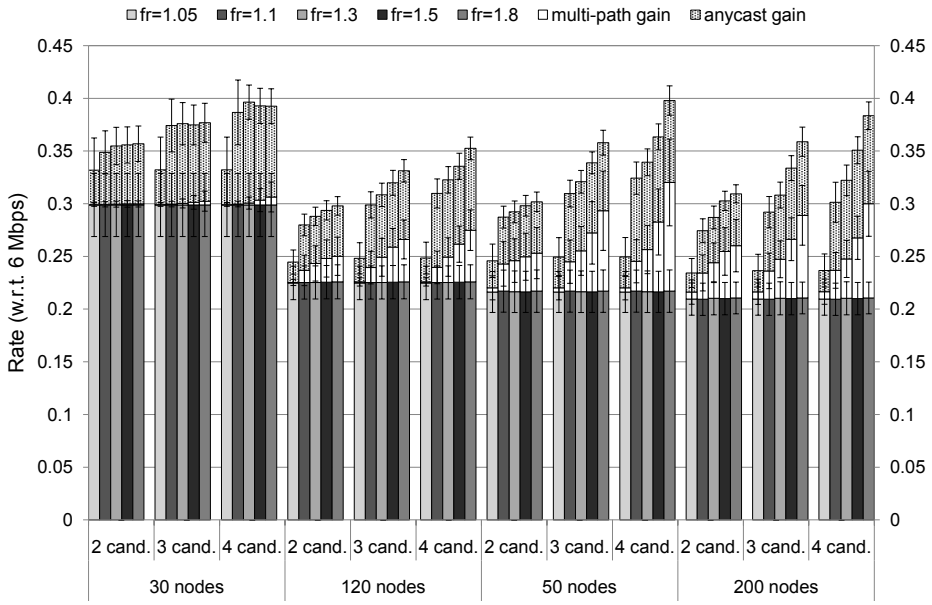
The considered topologies consist of 30-200 nodes randomly placed in a square area with varying node density (cf. Table 7). A single flow traverses the network from left to right. In the following, we use the terms adaptive and fixed bit-rate to refer to the results with all four bit-rates and only 6 Mbps, respectively. Note that the protocol model is used for interference modeling: Two links are in conflict and cannot be activated simultaneously if at least one node of every link (either sender or receiver) is within CS range of the other, i.e. the power of the received signal is above the CCA threshold. However, the model does not consider the impact of the cumulated interference of concurrent transmitters on the CCA and receiver operation. Thus, the obtained gains of spatial diversity should be understood as upper limits for a physical environment.

Every bar in Fig. 29 and Fig. 30 shows the average flow rate of the single-path variant in one particular system configuration with the multi-path and anycast gains on top. Note that the flow rate cannot decrease when going from single-path to multi-path and further to anycast, since the optimization gets more opportunities each time and chooses the best among them. The Rician fading with $K = 0$ equals to Rayleigh fading, and the higher the parameter K, the less severe the fading is and the closer the behavior gets to the AWGN channel. As shown in Fig. 30a and Fig. 30b, a Rician channel with $K = 12$ offers little MUD due to the low channel variability. The gains are below 0.04 in most cases, so that they become almost negligible with bit-rate adaption. On the other hand, Rayleigh fading has more potential for MUD. The gains are about 0.05 for 2 candidates and sufficient stretch parameters and even higher than 0.08 with adaptive bit-rates and 3 or more candidates (cf. Fig. 29a and Fig. 29b).

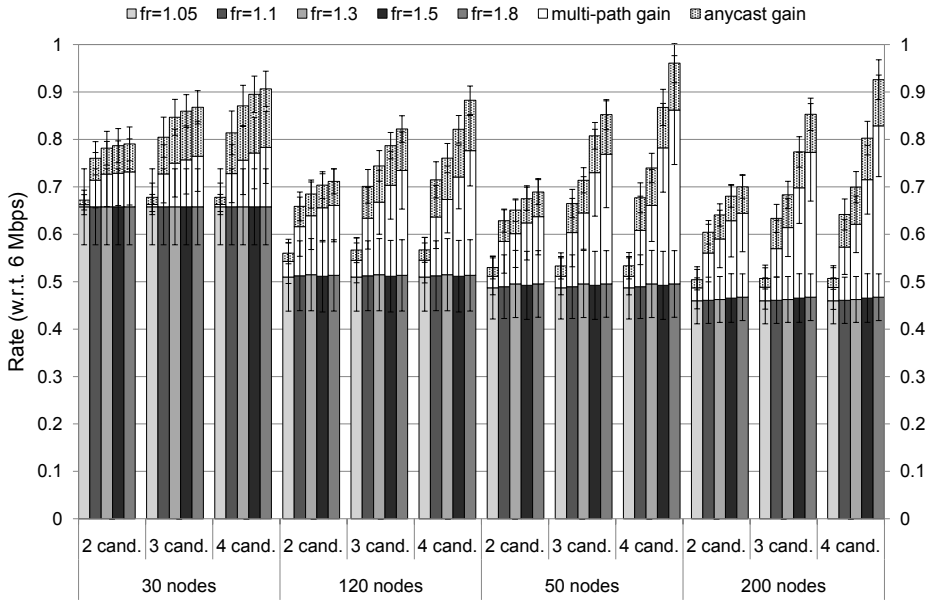
Interestingly, the relative MUD gains become smaller with bit-rate adaptation, most likely because the additional throughput of a higher bit-rate outweighs the MUD gains, which leaves fewer opportunities for anycast. By construction, the flow rate is non-decreasing in the number of candidates and in the stretch parameters of the route pre-selection. However, the returns of both larger stretches and more candidates are diminishing. For example, when increasing the stretch from 1.3 to 1.8, the additional MUD gains are below 0.015 in most cases.

We observed that two effects determine the achieved multi-path gains. With spatial diversity in a narrow sense, bottlenecks can be mitigated using alternative routes. However, in some cases the ETT heuristics[11] is not able to find the highest throughput route. With *route selection diversity*, we refer to the effect that additional paths can outweigh a sub-optimal routing decision. The absolute multi-path gains are almost equal across all Ks for fixed bit-rates. With adaptive bit-rates, on the other hand, the multi-path gains increase with K when using sufficient large routing stretches and

[11] ETT assumes that all nodes share the same collision domain [20].

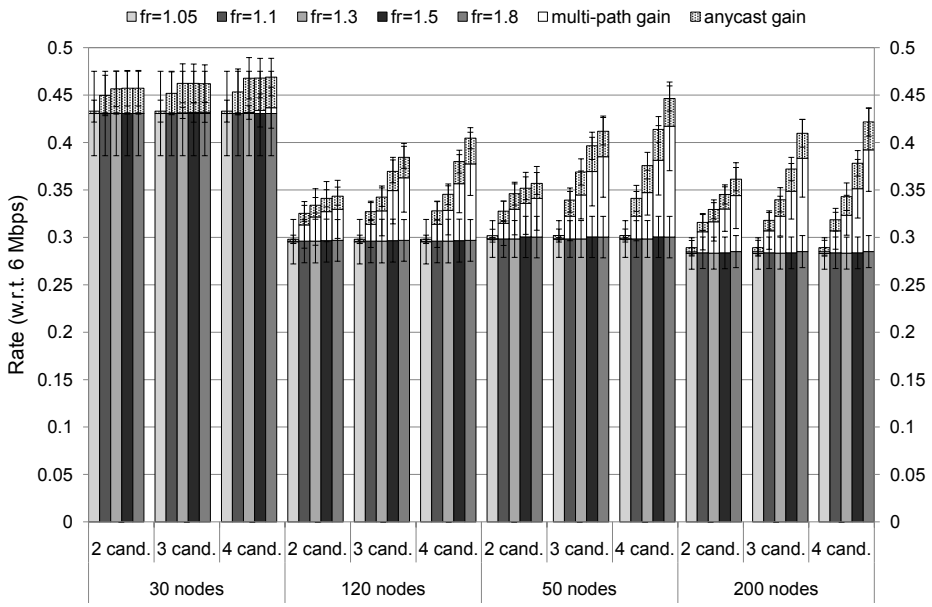


(a) Fixed bit-rate ($\pm std.dev.$).

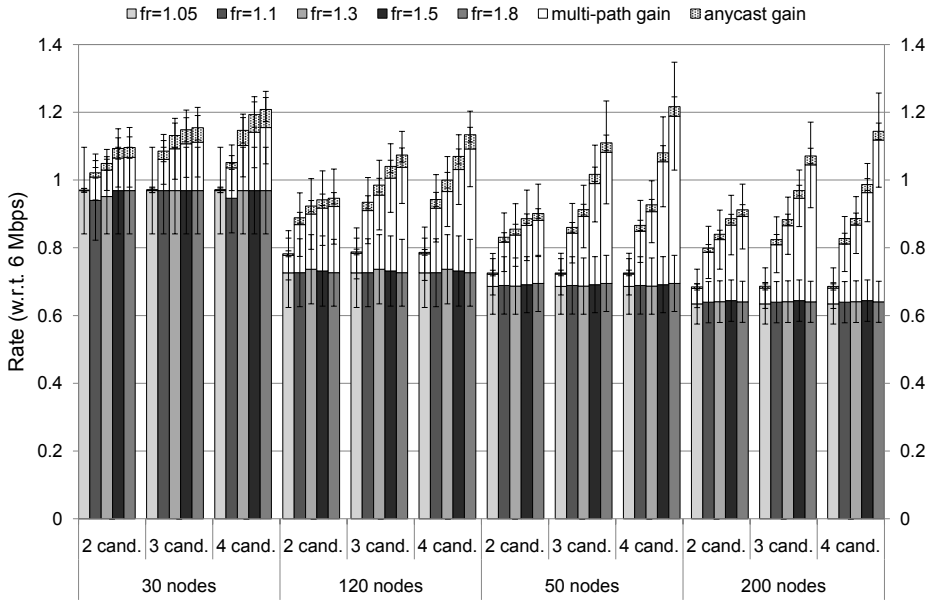


(b) Adaptive bit-rate ($\pm std.dev.$).

Figure 29: Flow rate obtained from numerically solving the multi-commodity flow problem ($K = 0$).



(a) Fixed bit-rate ($\pm std.dev.$).



(b) Adaptive bit-rate ($\pm std.dev.$).

Figure 30: Flow rate obtained from numerically solving the multi-commodity flow problem ($K = 12$).

network topologies. In either case, the absolute and relative multi-path gain of adaptive bit-rate selection significantly exceeds the gain when using a fixed bit-rate only. However, remember that our model does not account for cumulative interference and that the LCG is independent from the fading distribution.

The multi-path gain can be increased with larger routing stretches in most cases. Nevertheless, the returns for larger stretches are generally diminishing, but in a different way compared to the returns for MUD. With 2 candidates, the returns for stretches larger than 1.1 are small. In contrast, a third or fourth candidate may achieve significant returns beyond 1.1 given adaptive bit-rates and a sufficient large network. Thus, we suppose that the multi-path potential of larger networks is not used up, and both the number of candidates and the routing stretch are the main limiting factors.

We conclude that the multi-path potential is more pronounced in most of the considered topologies, whereas the potential for MUD can be significant depending on the fading distribution. With adaptive bit-rates, it is generally smaller in relation and it is used up quickly. With 2 candidates in particular, the anycast gains are about 10% and 20% for adaptive bit-rates and fixed bit-rate, respectively. Our route pre-selection heuristics succeeds in selecting paths with potential for MUD. The major gains can be realized with small routing stretches, and the topologies offer little additional returns when increasing the stretches further. On the other hand, the proposed route pre-selection is the limiting factor for realizing multi-path gains, which are within $10\% - 20\%$ and $20\% - 30\%$ for fixed and adaptive bit-rates assuming sufficient large networks and routing stretches. However, remember that the multi-path gains are optimistic since we do not account for cumulative interference and the LCG is independent from the fading distribution.

At typical working points of UO-CSMA in the finite regime, the system has to pay for every additional relay and every additional (anycast) link in terms of flow rate or delay. Thus, a main deficit of the route pre-selection heuristics is that it blindly includes additional nodes and anycast links in the routing decision when increasing the routing stretch and the number of anycast candidates, respectively, which may cause throughput losses for the real system. In our future work, better heuristics have to be developed that trade the incurring costs of additional relays and anycast links against the expected returns. For example, the flow source may solve the optimization problem locally using the link table information as an approximation of the LCG. This way, it can identify promising routes and exclude less promising relays using (possibly randomized) source routes within every packet.

8.2. A Single Flow on Random Topologies

In the following, we evaluate the performance of the proposed cross-layer protocol carrying a single flow through simulations. Our intention is twofold. On the one hand, we consider the extent to which the proposed protocol can realize the potentials that we have illustrated in the previous section at typical operating points of UO-CSMA in the finite regime. On the other hand, we compare the performance of the proposed protocol and state-of-the-art protocols across a large set of randomized topologies and within different configurations and environments to illustrate the ben-

Parameter	Value
Area	300 m $\times$ {300 m, 500 m}
Topology	random with 30 nodes
Fading	i.i.d. {Nakagami $\mu = .75$, synthetic $\frac{1}{2}$PSR}
Receiver	SNR-BER & cuml. interference
CCA	Mode 1, threshold -88.5 dBm
TA limits	4 . . . 8.8
Queue limits	16 . . . 100
Inter-flow hysteresis	1.1
Step size (initial/target)	$0.1 \cdot K^{-1}$ / $0.002 \cdot K^{-1} \cdot C^{-1}$
Step size update interval	200 ms
Target per-hop delay	0.8 s
Flow duration	20 s evaluated (excl. warm-up)
Packet size	1500 Byte

Table 8: Simulation Parameters (In addition to Table 7)

efits of a mesh-enabled MAC and a cross-layer routing protocol especially designed for WMNs.

We present simulation results obtained from a detailed WMN simulator. The system parameters are summarized in Table 8. In particular, we are using slightly different fading models with the simulations: A Nakagami fading [37] with parameter $\mu = 0.75$ (scenario B) and a synthetic model that halves the PSR of the AWGN channel regardless of the link's SNR (scenario A). Remember that a Nakagami fading with $\mu = 1$ corresponds to Rayleigh fading and Rician fading with $K = 0$. Thus, the severity of the considered fading is slightly higher compared to Rayleigh fading. To get an impression of the dimension of the simulation scenarios, Table 9 illustrates the average hop count across all 100 repetitions for DSR/UDP.

Scenario	Length	Bit-rate	Avg. Hop Count
A	300 m	single	2.0
		adaptive	3.2
	500 m	single	3.8
		adaptive	5.3
B	300 m	single	2.4
		adaptive	3.9
	500 m	single	4.2
		adaptive	6.3

Table 9: Average hop count for DSR/UDP in scenario A and B

In Fig. 31, we have plotted the goodput CDF across all seeds for scenario B (Nakagami fading) with a width of 500 m and adaptive bit-rates. In the figure, DSR and the cross-layer protocol with a single candidate (in short: 1/1), which are both single

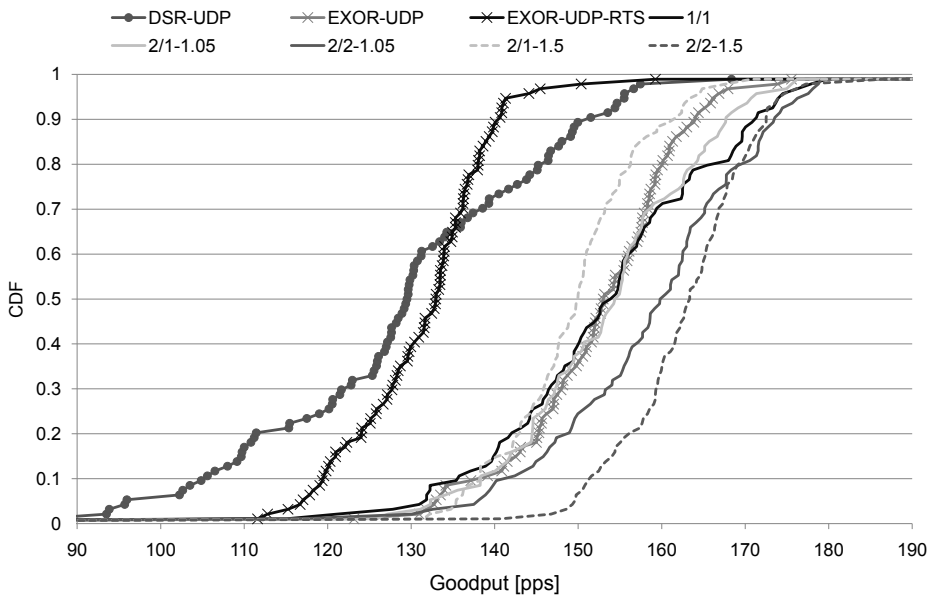


Figure 31: Goodput CDF for scenario B (Nakagami fading, adaptive bit-rates, 500 m, legend: no. candidates / no. anycast receivers - metric stretch).

path routing approaches, achieve a shallow CDF. The CDF gets steeper with multi-path routing. However, this is an indication of diminishing returns, since the diversity will get used up in the same way the links are improved. If multi-path is used without anycast (2/1), then goodput is lost especially in the high goodput regime (70% and above). Remember that additional links increase the optimality gap (4) of UO-CSMA (cf. page 13). Thus, the overall goodput suffers if the scenario does not offer enough spatial diversity to outweigh this loss. The same holds for anycast (2/2), although the scenarios offer sufficient MUD to increase the overall goodput. However, the benefits of anycast are not realized in the high goodput regime, which is most likely an indication that the MUD is used up there.

With increased routing stretches, the above mentioned multi-path and anycast effects become more pronounced. For example, with a metric stretch of 1.05, the anycast and single path CDFs are close to each other within the low goodput regime ($0\% - 20\%$), whereas a substantial improvement can be observed with a metric stretch of 1.5. Nevertheless, the losses also increase with higher routing stretches in the high goodput regime if multi-path is used without anycast.

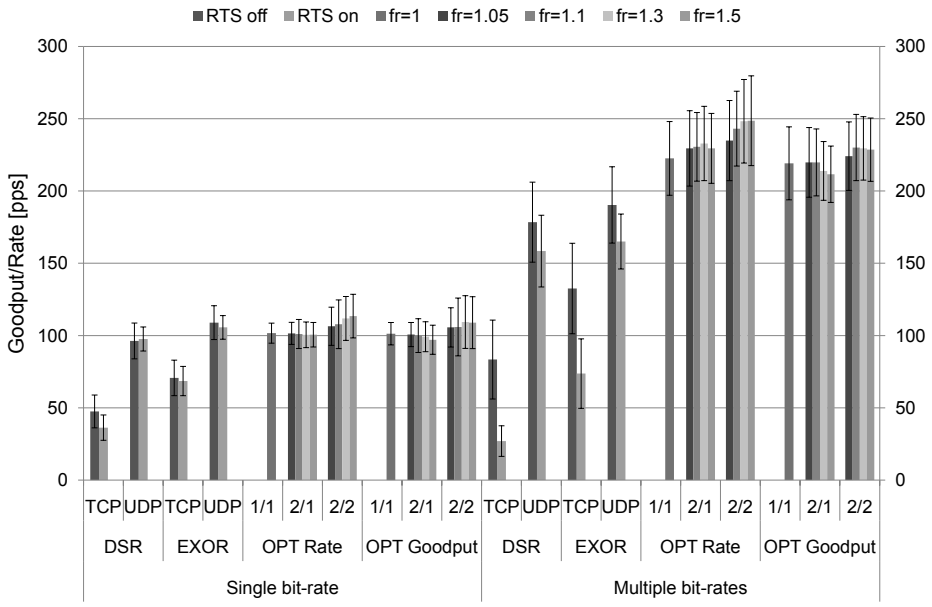
The figure furthermore illustrates that the cross-layer protocol significantly improves the single-path performance. For example, the median goodput increases to 154 pps from 133 pps with DSR/UDP. In relation, the benefits of anycast are smaller, e.g. the median goodput of the best anycast configuration is 163 pps. In the consid-

ered scenario, we observe that it pays off to consider the characteristics of WMNs in the MAC design. However, the returns of OR diminish at the same time since the opportunities for the routing layer to outweigh the impairments of the MAC vanish.

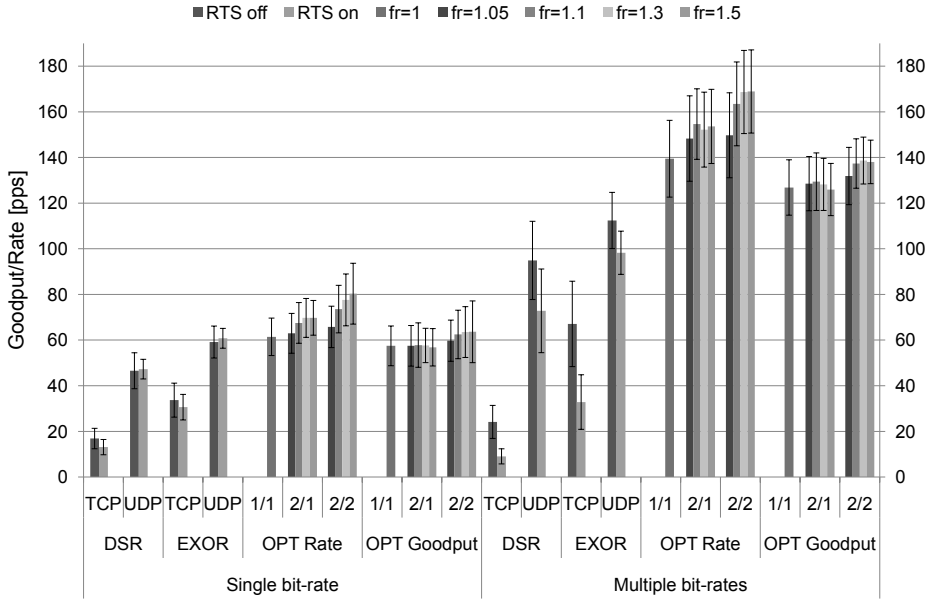
The average goodput and the estimated flow rates for scenario A and B are shown in Fig. 32 and Fig. 33, respectively. Our observations are consistent across all configurations, so that we refer to an individual diagram only if necessary. Using TCP with either DSR or ExOR results in a loss of goodput. The same holds for using RTS in most cases and especially with adaptive bit-rates. However, note that both protocols do not use TXOPs, so that the per-packet overhead increases with RTS at higher bit-rates. Furthermore, ExOR has a higher goodput performance than DSR in the considered scenarios.

The goodput performance of the proposed protocol with single-path routing (1/1) is better than DSR and close to ExOR with a single bit-rate and it even exceeds ExOR with adaptive bit-rates. However, remember that both DSR and ExOR are less conservative in terms of MAC exclusion due to the hidden node problem. The cross-layer protocol, on the other hand, reserves the medium around transmitter and receiver(s), which results in lower spatial reuse. To understand the impact of the larger exclusion region, we additionally conducted simulations with lower transmission powers P_{bt} for the data busy tone (BTD). This way, the exclusion region around transmitter and receiver(s) gets smaller and allows for higher spatial reuse, which comes at the expense of potentially lower SINR margins and thus an increased risk for receive failures. We have plotted the goodput results for the larger topology in Fig. 34. The diagram illustrates the tradeoff within the selection of P_{bt} especially for scenario B with the physical fading model: The highest goodput is achieved with a slightly reduced power setting. With the synthetic fading in scenario A, the tradeoff tends towards a more conservative power setting, which indicates that the SINR margins are smaller. Setting the optimal BTD power, the performance of the cross-layer protocol improves further in relation to ExOR.

When using multi-path routing without anycast (2/1), the goodput of the cross-layer protocol slightly decreases with higher routing stretches in most cases. On the other hand, in the previous section we have seen that the topologies offer spatial diversity especially with adaptive bit-rate selection. Nevertheless, remember that we use a physical interference model with the simulations, whereas the protocol model in the previous section leads to optimistic results for spatial diversity. At the considered working point, however, the multi-path gains are outweighed by the increased optimality gap of UO-CSMA due to the higher number of relays. We have also observed many cases that have their bottleneck at the first hop. Increasing the routing stretch is counterproductive in this case, since the bottlenecks cannot be bypassed. The end-to-end delay limiting further intensifies the effect: We observe that the estimated flow rate increases in most cases, whereas the goodput declines at the same time. The difference of flow rate and goodput is roughly the amount of virtual traffic (cf. section 6.3). With higher per-hop delay limits, both flow rate and goodput will converge and the virtual traffic will vanish (cf. section 7.4). The system can better utilize the multi-path gains in two different ways. It may decide to operate UO-CSMA at a working point that has a higher throughput efficiency, so that the optimality gap

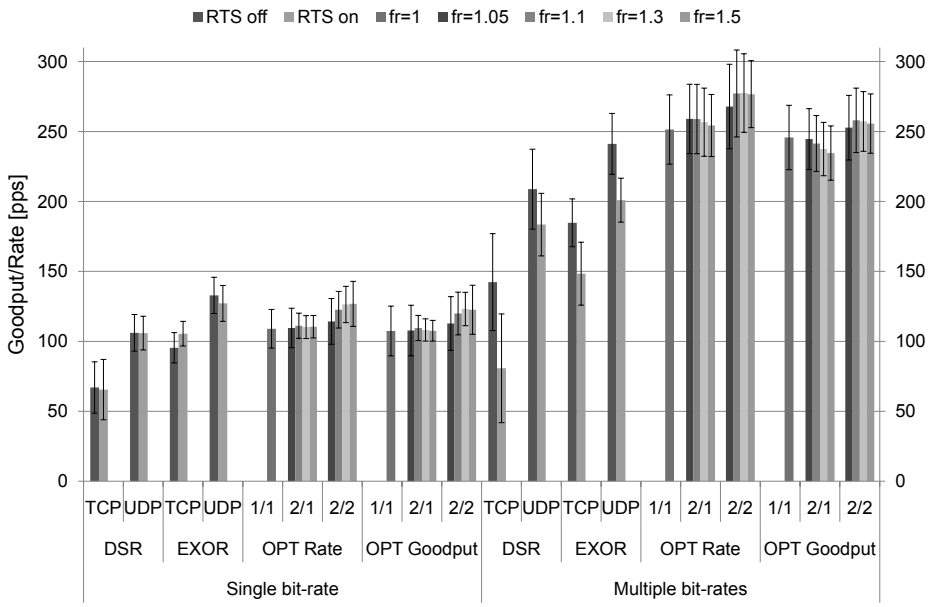


(a) Flow rate and goodput (300 m, $\pm std.dev.$).

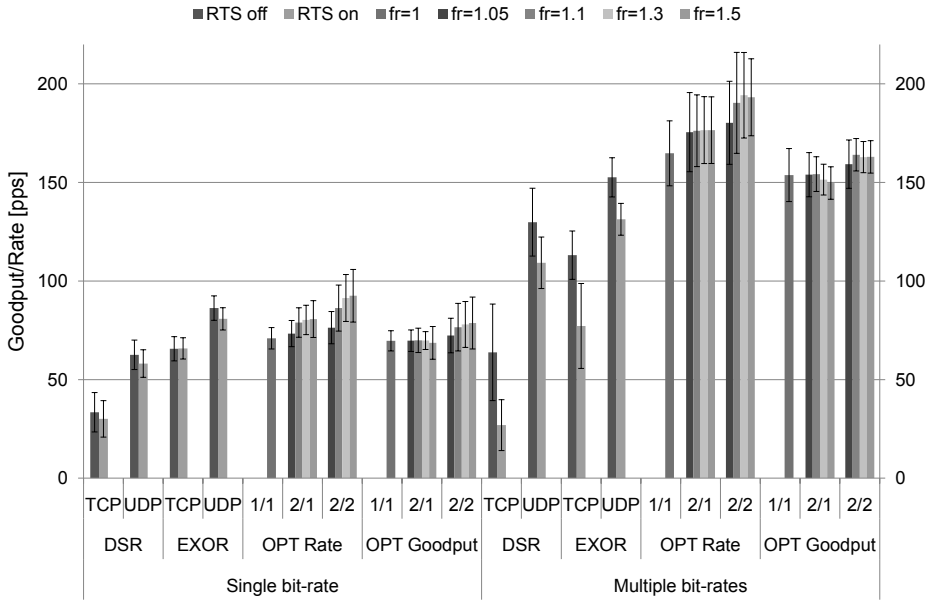


(b) Flow rate and goodput (500 m, $\pm std.dev.$).

Figure 32: Simulation results for Scenario A (Synthetic fading $\frac{1}{2}$PSR).



(a) Flow rate and goodput (300 m, $\pm std.dev.$).



(b) Flow rate and goodput (500 m, $\pm std.dev.$).

Figure 33: Simulation results for scenario B (Nakagami fading $\mu = .75$).

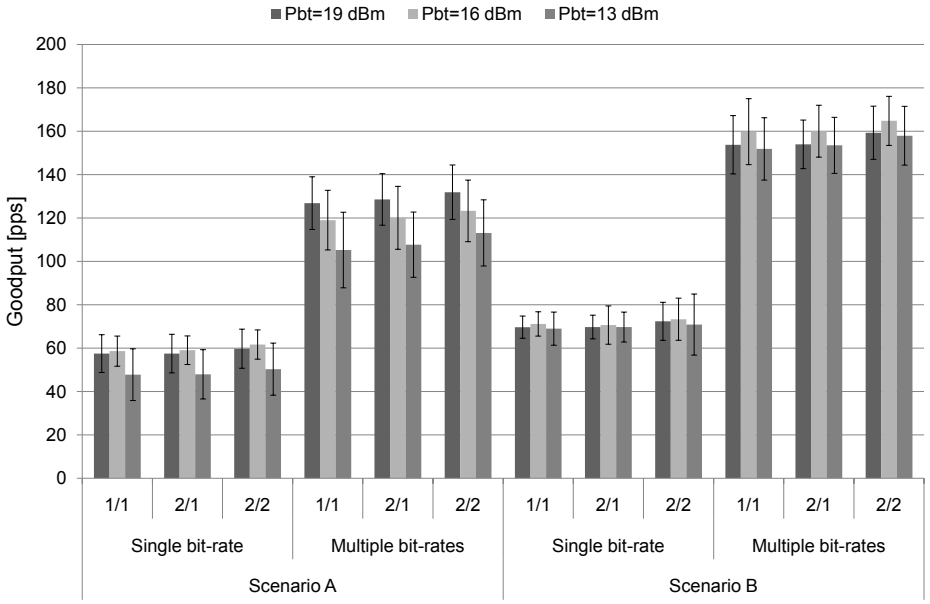


Figure 34: Goodput for both scenarios depending on the power of the data busy tone (500 m, stretch $f_r = 1.05$ and $f_h = 1.5$, $\pm std.dev.$).

is reduced. However, this requires the development of new technologies having the necessary capabilities. On the other hand, the system might still utilize additional multi-path gains if the route pre-selection trades the benefits of additional links and relays off against the incurring costs, as lined out in the previous section.

With anycast (2/2), the goodput performance can be increased even at small routing stretches. By increasing the stretch parameter further, the goodput increases up to a point where the benefits are outweighed by the costs of additional links. The anycast gains are slightly smaller than we would expect from section 8.1, but altogether they are close to the anticipated behavior. As mentioned above, the cost-benefit estimation per anycast link within the route pre-selection will most likely reduce the gap to the optimal results.

With only a single bit-rate, the performance of the cross-layer protocol with anycast is slightly lower compared to ExOR, which is especially pronounced for scenario B (Nakagami fading). The reasons are manifold. The SINR margins are larger with Nakagami fading (cf. Fig. 34). The gap reduces with a less conservative exclusion. Furthermore, remember that the cross-layer protocol achieves the optimum asymptotically. In the considered regime, it cannot be prevented that the system takes sub-optimal decisions, i.e. traffic is allocated to sub-optimal links to a particular extent (cf. section 7.6). For ExOR, on the other hand, the risk of wrong decisions is lower in small topologies since there are fewer possibilities. With adaptive bit-rates and

large topologies, the amount of possibilities explodes and a wrong decision becomes more likely for ExOR, so that the proposed protocol should have more advantages. In addition, the missing congestion control in ExOR contributes to the advantages of the cross-layer protocol especially in multi-rate environments. ExOR splits the traffic among different paths statically. However, if one path is not able to support the predetermined fraction of traffic, the queues will build up. The end-to-end congestion control that we used in the simulations assures that the number of in-flight packets does not exceed a fixed amount of 1000 packets in our case. If this limit is finally reached, the goodput on the other paths suffers from the static traffic split, as well, since the congestion controller will throttle its rate.

The generation of the link tables for the ETT routing metric [22] has been one particular problem in the simulations. There are conflicting objectives for the estimation of the link quality indicators (LQIs). On the one hand, they should be accurate, and on the other hand, the estimation should consume few medium resources and it should adapt to environmental changes in a timely manner. In the end, it boils down to the selection of the interval, packet sizes and bit-rates for link probing. For the simulations, we used a conservative setting with a very long settling time in order to get acceptable LQIs estimates, so that the results remain comparable. However, with typical parameter settings [20] the accuracy is very low, which gives rise to *route selection diversity*. In scenarios having neither multi-user nor spatial diversity, using multiple paths offers no advantages, however, it might mitigate the uncertainty in the LQIs. Thus, route selection diversity outweighs sub-optimal routing decisions by using multiple alternatives, i.e. it is a special form of MUD. We plan to explore this topic further in our future work.

In conclusion, this section has illustrated that it is necessary to consider the mesh nature in the design of MAC protocols. The state-of-the-art approach is to use MAC protocols with WMNs that have been designed for cellular environments. From our simulation results, we conclude that this approach leads to a waste of resources especially on the MAC layer. Advances techniques like OR might mitigate the design shortcomings to a certain extent. However, our results indicate that a proper MAC design within an integrated cross-layer protocol achieves the higher performance in most cases. Conversely, the benefits of MUD become smaller. OR is a technique to compensate the deficits of the lower layers. The less is wasted on the lower layers, the smaller the potential of OR.

Furthermore, our results clearly show that every additional link causes high costs in terms of the UO-CSMA optimality gap. In reducing the gap, two directions seem promising. New technologies for contention based medium access might shift the physical constraints on CSMA, which enables working points with higher efficiency. With Multi-Carrier Burst Contention, Roman *et al.* present a promising concept in this direction [104]. As an alternative direction, we can enhance the route pre-selection through individual cost-benefit estimations per anycast link, which remains for our future work. This way, the route pre-selection heuristics would be able to incorporate the expected spatial gains into its decisions. We have focused on MUD in the design and evaluation above. We used a low path loss exponent, so that a large separation is necessary to decouple two links spatially. Furthermore, the simulated propagation

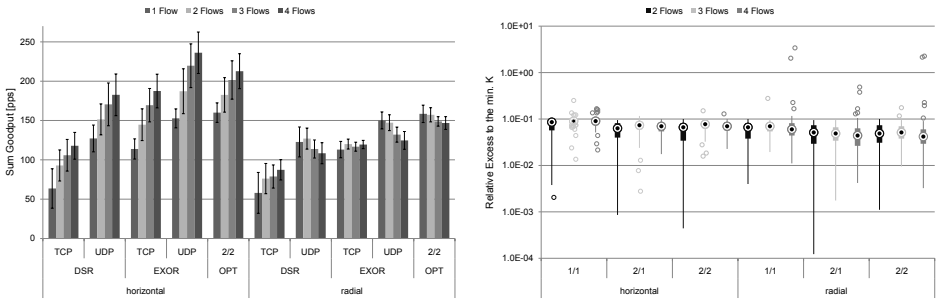
environment was homogeneous, whereas we would encounter obstacles in the propagation paths in reality. It remains open for our future work to evaluate the system with enhanced route pre-selection within several propagation environments. For example, the multi-slope path loss model [37, 120] may limit the spatial dependencies in the simulated WMN. In addition, spatially correlated shadow fading models have to be developed to account for inhomogeneity in the propagation environment.

8.3. Multiple Flows on Random Topologies

With multiple flows sharing resources of the WMN, the fairness between them becomes another objective for the network. The concept of utility is generally used to handle the tradeoff between throughput and fairness. In this section, we evaluate the proposed protocol in terms of achieved utility. Our intention is twofold once again. On the one hand, our objective is to characterize the behavior of the system with increasing number of flows in order to identify its limitations. On the other hand, we compare the performance of the system to state-of-the-art approaches within typical multi-flow scenarios. In particular, we consider two well-known scenarios: In the *flow-in-the-middle* scenario, the collision domains of the involved flows overlap to varying degrees. Traditional approaches are reported to handle this scenario unsatisfactory resulting in severe unfairness and starvation [30]. In the *congested-center* scenario, the central area of the network becomes overloaded due to traffic flows, which originate at the border and cross the network. Generally, a shortest path routing protocol chooses paths through the center of network, so that the center becomes congested with increasing number of flows [82]. Interestingly, Yi *et al.* report that multi-path routing can spatially smooth the traffic in this scenario, so that the congestion at the network center is mitigated [128].

The simulation setup is comparable to the previous section (cf. Table 8) with a few exceptions. The simulation topology consists of 50 nodes randomly placed in an area of 500 m $\times$ 500 m with a minimal neighbor distance of 50 m. We used Nakagami fading only with $\mu = 0.75$ and adaptive bit-rates. There are 1-4 flows that traverse the network either horizontally or radially. With radial traffic, we lay a circle over the network that has the same diameter. We place the flow sources equally spaced on the half circle. The destinations are placed in opposition. Based on these ideal positions, we use the closest network nodes as respective sources or destinations. For horizontal traffic, we place the ideal positions equally spaced on opposite edges of the network and determine the actual nodes according to their distance from the determined positions.

We have already reported results for one flow in the previous section. Nonetheless, we have plotted them in Fig. 36 and Fig. 37 as a point of reference. Note that the simulation topology is the same for either horizontal or radial traffic. However, the horizontal pattern smoothes the traffic spatially, whereas the center of the network will most likely become the bottleneck with radial traffic (*congested-center* scenario). And in fact, the cumulated goodput is bounded with radial traffic (cf. Fig. 35a). Hence, the absolute utility with horizontal traffic is generally higher compared to radial traffic except for one flow, of course. With 3 and more flows, the horizontal traffic pattern



(a) Cumulated goodput (RTS off, $\pm std.dev.$). (b) Excess of the max. K relative to the min. K.

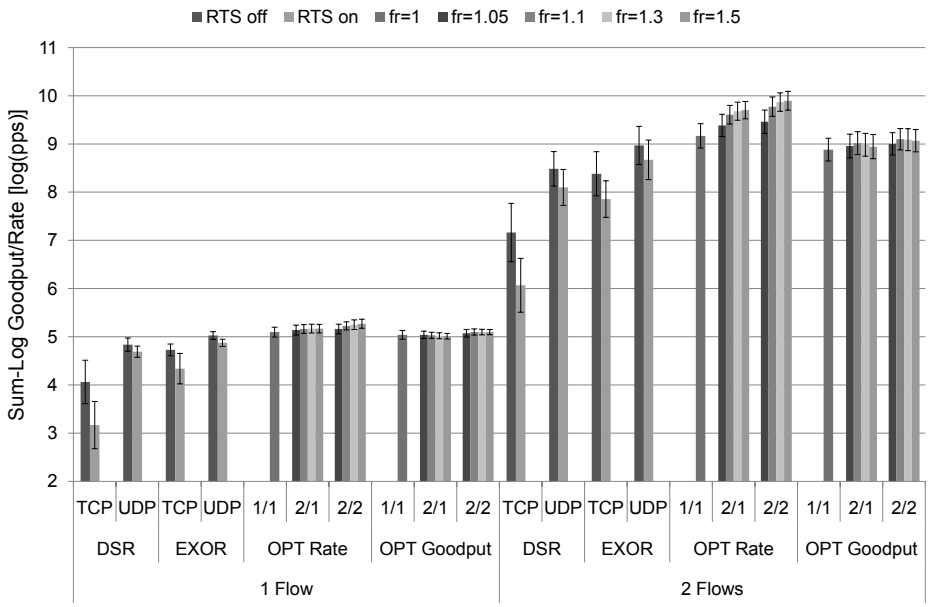
Figure 35: Cumulated goodput and excess of K between flows ($f_r = 1.05$).

results in the typical *flow-in-the-middle* scenario.

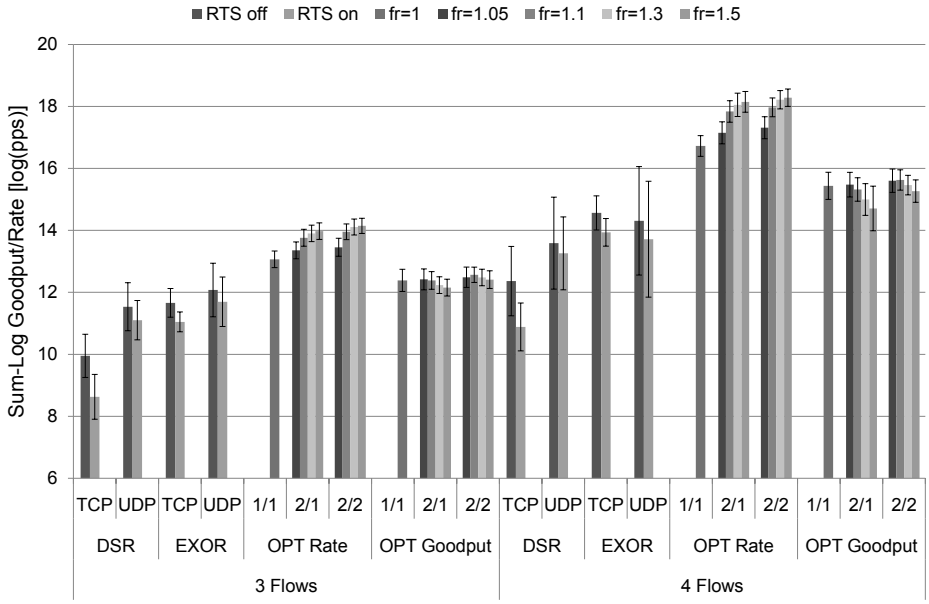
To start with, we take a closer look at the effectiveness of the inter-flow fairness adaptation. Remember that the inter-flow adaptation should ensure that neighboring flows within intersecting collision domains achieve proportional fair flow rates, which boils down to the problem of equalizing the K parameters of all flows (cf. section 6.2.2). Our objective for the inter-flow adaptation is that the involved Ks should not exceed the minimal K by more than a fixed ratio that we have set to $\delta = 1.1$ (cf. Table 8). In Fig. 35b, every bar shows the distribution of the relative excess in K across all 100 repetitions. The excess of the maximum K relative to the minimum K is defined as $\max_f K_f / \min_f K_f - 1$. The figure shows that the inter-flow adaptation is able to meet the given threshold of 10% with only few exceptions.

With increasing number of flows, the advantage of the proposed protocol in terms of aggregated utility increases with respect to DSR and ExOR, i.e. its working point is better in terms of proportional fairness (cf. Fig. 36 and Fig. 37). At the same time, the variability of the utilities across all random topologies is lower when using the cross-layer protocol. With DSR or ExOR, on the other hand, the variability significantly increases especially with horizontal traffic, which is most likely an indication of the flow-in-the-middle problem and the related starvation effects. Interestingly, ExOR/TCP exceeds its UDP counterpart in terms of aggregated utility with four flows in both Fig. 36b and Fig. 37b. Thus, ExOR achieves the higher overall goodput with UDP (cf. Fig. 35a) at the expense of fairness.

With increasing number of flows, the effect of larger routing stretches becomes mostly negative. We suppose that the multi-path and anycast benefits do not significantly increase if more flows have to share the same topology. However, the number of involved links increases, which increases the optimality gap of UO-CSMA further. Compared to the single flow scenarios, the limited multi-path and anycast gains with multiple flows are more often not able to outweigh the incurring costs of additional links. Thus, the cross-layer protocol achieves the best tradeoff for multiple flows with generally smaller routing stretches. As shown in Fig. 35a, the horizontal traffic pattern offers a higher multi-path and anycast potential. Hence, the system achieves

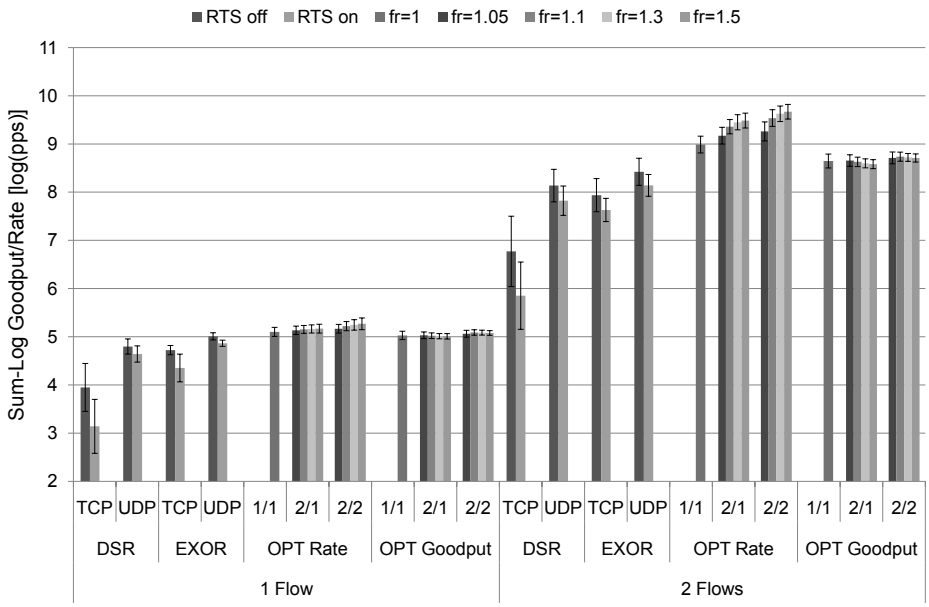


(a) Rate and goodput for 1 and 2 flows ($\pm std.dev.$).

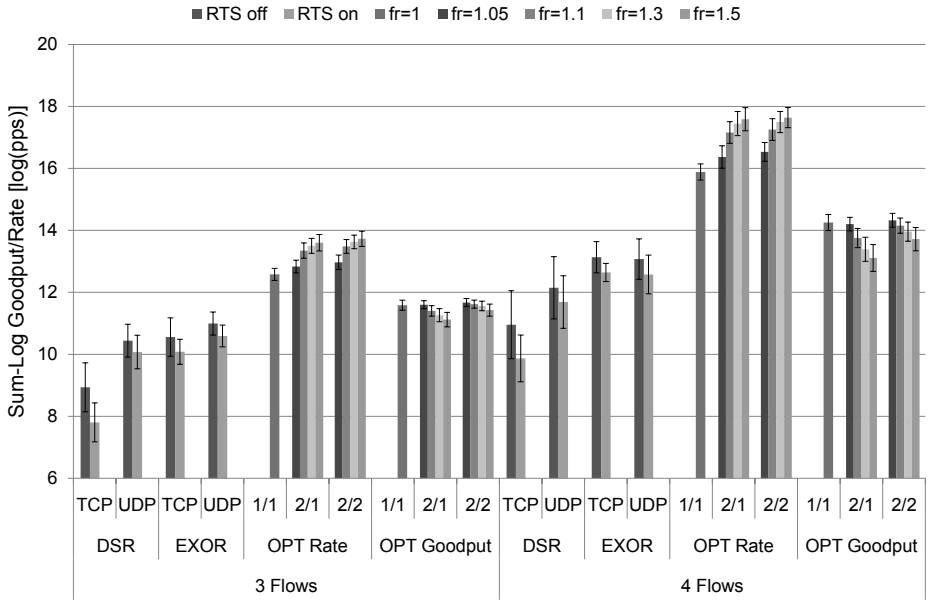


(b) Rate and goodput for 3 and 4 flows ($\pm std.dev.$).

Figure 36: Cumulated utility for horizontal flows.



(a) Rate and goodput for 1 and 2 flows ($\pm std.dev.$).



(b) Rate and goodput for 3 and 4 flows ($\pm std.dev.$).

Figure 37: Cumulated utility for radial flows.

the highest utility with 4 flows and horizontal traffic using a slightly larger routing stretch compared to radial traffic.

With every additional flow, the discrepancy between flow rate and achieved goodput increases, i.e. there is more virtual throughput. Remember that virtual throughput arises whenever the per-flow queue of a node underflows or overflows. In the former case, the saturation assumption is violated, whereas our delay limiting approach prevents the excessive accumulation of packets at a particular node in the latter case (cf. section 6.3.2). To a large extent, the throughput capacity of the network has to be shared by all flows (cf. Fig. 35a). Thus, the same per-hop delay targets translate to smaller arrival rates and smaller queue limits, which eventually increase the risk of queue overflows. Remember that excessive virtual traffic generally bears the risk that the physical flow rate deviates from proportional fairness since the congestion controller regulates the joint rate of physical and virtual traffic. However, the results indicate that this problem is not dominant in our case.

In conclusion, the results indicate that the cross-layer protocol can handle multiple flows more efficiently, i.e. its advantage in terms of utility tends to increase with additional flows. It can furthermore handle the flow-in-the-middle scenario, on which traditional approaches do not succeed in general [30]. On the other hand, we did not observe multi-path benefits in the congested-center scenario [128]. We suppose that the considered topology might not possess sufficient potential for spatial smoothing. Furthermore, the spatial spreading of the traffic involves the activation of additional links, so that the optimality gap of UO-CSMA becomes a further limiting factor.

The multi-path and anycast gains are generally smaller with multiple flows for two reasons. Within the same topology, the potential for both does not scale in the number of flows. Thus, the potential is used up faster, so that the effect of increased routing stretches becomes more often counterproductive. However, the results emphasize that the route pre-selection should be more selective. Instead of blindly including additional links, it should decide on an individual cost-benefit basis. With multiple flows, the problem of virtual traffic increases. As the results indicate, the consequences are still acceptable in our case. However, the problems will intensify in larger scenarios, so that we might not be able to ignore them anymore. One promising idea to mitigate the problem might be an architecture, in which the flows share per-neighbor queues as suggested by Bui *et al.* [12]. However, for CSMA the general tradeoff between throughput, delay and complexity remains. Due to varying step sizes, the credit and queue dynamics are only weakly coupled. A promising direction for our future work is the tighter reintegration of both processes in order to better understand and control the delay performance.

9. Conclusion and Outlook

In this paper, we have addressed the question of how opportunistic routing (OR) should be used in WMNs based on carrier sense multiple access (CSMA) in order to efficiently utilize multi-user and spatial diversity. We have presented a model for CSMA/CA with node-oriented CS that captures the characteristic tradeoff be-

tween multi-user gain and spatial reuse, which arises with OR. Even with node-based CS, CSMA can be operated in a simple and distributed way if the receiver blocking problem is handled properly. Based on the idea of dual busy tone multiple access (DBTMA) [42], we have designed a dedicated MAC protocol for WMNs called CSMA/HBT according to the model above. We have presented a distributed algorithm for congestion control, opportunistic routing and CSMA scheduling that targets WMNs with unreliable and memoryless links. The algorithm maximizes the cumulated throughput utility of elastic traffic flows that do not impose tight end-to-end delay constraints.

We have refined the above algorithm to a cross-layer protocol. We have defined a working point for CSMA in terms of contention aggressiveness that achieves a tradeoff between throughput efficiency and CSMA collisions. The working point can be approached via an intra- and an inter-flow feedback loop. Furthermore, we have presented an approach to control the throughput-delay tradeoff in CSMA via upper bounds on the physical queues and virtual transmissions. It reduces the convergence time and the end-to-end delay to a level of practical relevance. In addition, we have proposed an approach to pre-select the routing paths according to a WMN routing metric in order to bound the optimality gap (4) of UO-CSMA, which increases in the number of involved links.

We have illustrated the feasibility of the proposed protocol using a simulator prototype. Our solution fits well in the OSI reference model, whereas it does not comply with IEEE 802.11 in all particularities. We have evaluated the proposed protocol through analysis and simulation. Neither multi-user nor spatial diversity dominates in the analytic characterization of random topologies. Thus, both forms of diversity should be systematically and dynamically exploited. We have illustrated how the system handles the tradeoff between spatial reuse and multi-user gain. In particular, it is crucial that all anycast links properly contend for medium access at the transmitter, and the topology "decides" on the links to activate according to the CS relationship.

The simulation results suggest that the prototype is able to increase the throughput and fairness performance of WMNs in relation to state-of-the-art single-path and OR protocols. However, the optimality gap to the theoretical potentials increases with the utilized degree of multi-path and opportunism. On the one hand, a more robust MAC leads to fewer opportunities on the higher layers and thus smaller OR gains. On the other hand, a fundamental tradeoff arises between utility and delay on the one hand and the degree of multi-path and opportunism on the other. The more relays are used, the higher are the expected benefits of spatial and multi-user diversity, but they are generally diminishing and will eventually be eaten up by the costs of every additional relay in terms of the optimality gap of UO-CSMA. This effect intensifies in the number of competing flows, since the spatial and multi-user gains do not scale with the number of relay links in general. Altogether, the multi-flow OR performance of the cross-layer protocol exhibits a larger gap to its theoretical potential, despite its benefits in terms of utility seem to increase in relation to state-of-the-art approaches.

9.1. Theory-Practice Gaps, Limitations and Future Directions

Lee *et al.* have discussed the differences between theory on UO-CSMA and the practical results they have achieved [70]. In order to show current limitations of the protocol and future directions, we would like to continue this discussion in the following using our lessons learned.

Node-Oriented Carrier Sensing and Hidden Nodes. We have shown that CSMA can be controlled in a simple and distributed way even with node-oriented CS if the receiver blocking problem is properly considered and the exclusion regions are properly designed. However, both the extended CSMA model as well as the proposed protocol relies on assumptions. In particular, we have assumed that the synchronization rate is independent from all other TAs in section 3. Furthermore, we have assumed that no link becomes inactive as long as there is an ongoing probe within the contention neighborhood. In addition, we have assumed that the PSRs are stationary and i.i.d. in time. The MAC protocol generally allows a small amount of hidden interference, so that the PSRs become dependent on the protocol decisions. Nevertheless, CSMA/HBT is only one instance of a protocol that complies with the model. Furthermore, it remains open whether the high power busy tones are necessary, or if another protocol can be designed that has lower hardware requirements.

Carrier Sensing on Real Hardware. The IEEE 802.11 standard defines different types of CCA procedures [31]. In particular, IEEE 802.11g proposes a combination of energy and preamble detection with a mandatory energy threshold of -76 dBm and below. Thus, the receiver performance for a typical CSMA collision, for example, may be different from cases, in which the preamble can be used [69, 105]. Furthermore, proprietary features like the ambient noise immunity may further change the CS behavior [106]. However, it remains open whether the implementation decisions are made for technological or physical reasons that impose fundamental limits, or whether they are a compromise with respect to implementation complexity, costs or energy consumption constraints.

We have used the protocol model to derive the MAC protocol. Due to cumulative interference within the physical model, a perfect mutual exclusion requires a large amount of spatial resources. Thus, a relaxed exclusion is often more practical. In addition, the CS detectors are not perfect and their decisions are always subject to uncertainty. They experience propagation and processing delays resulting in CSMA collisions. Furthermore, the decision to block a link is not always correct. If the CTS is missing due to a blocked receiver, the transmitting link should be blocked. However, if the CTS is missing due to a CSMA collisions, the blocking is unnecessary. Nevertheless, the transmitter is unable to distinguish between both events.

Advanced Contention Procedures. We have explored the adaption of the backoff durations with UO-CSMA. Alternatively, the length of the TXOP can be adapted [47],

and combinations are also possible [78]. However, it may become necessary to repacketize the frames at intermediate hops to change the length of the TXOP, and the incurring costs have to be considered.

Furthermore, it is necessary to reconsider the contention mechanisms for advanced PHY technologies. For example, Jose *et al.* propose a technique called *measuring*, that tries to pack concurrent transmissions in a spatial more efficient way [57]. Qian *et al.* extend the CSMA model for the case of MIMO links, in which the state of link is no longer binary due to the capability of spatial multiplexing [95]. In addition, Coviello *et al.* propose a CS technique for MIMO networks that does not give a binary decision only, but tries to estimate the number of currently active data streams [21]. We strongly belief that alternate contention mechanisms are the key for WMNs to catch up with the speed of innovations at the PHY.

Lossy and Time-Varying Channels. The proposed cross-layer protocol is able to handle lossy links that are affected by a memoryless channel error process. Furthermore, the system is able to mitigate the channel impairments using the advanced technique of OR. The memoryless channel is the best case for OR since the error process cannot be predicted.

An important future direction is the investigation of channels with memory, for example the slow fading channel [103]. A viable alternative in this case is opportunistic scheduling, which tries to adapt the system to the instantaneous channel conditions [115]. Hence, the question arises whether opportunistic routing or opportunistic scheduling is better suited to extract MUD from slow fading channels. Furthermore, the impact of channel errors on signaling traffic and BTs should be considered in more detail for channels with and without memory. In particular, in LCG becomes time-varying in slow fading channels, and it is unclear whether this affects the control strategy of UO-CSMA.

Throughout this paper, we have assumed that channels are stationary. However, real links are expected to exhibit non-stationarity at macroscopic time scales from minutes to hours and days due to node mobility, shadowing and environmental mobility. In the long run, the system will adapt to the changed channel statistics. The question is how long it takes for the system to reach the new operating point after abrupt changes, which heavily depends on the current step size used in the parameter updates. It may become necessary to employ a change detection approach that adaptively increases the step sizes in order to reduce the re-convergence delay.

Working Points for the Collision-Efficiency Tradeoff. We have selected a working point for the collision-efficiency tradeoff that prevents the collision breakdown of the network as long as the node and flow density remains limited. A future direction is the consideration of alternative working points. Interesting work points are, for example, the ones that are independent from the node or flow densities. In particular, if the selected objective can be expressed in a similar model-predictive way, we expect that the proposed feedback loops can be used in the same way.

Relays and Their Costs in Multi-Path and Opportunistic Routing. UO-CSMA is asymptotic utility-optimal. The optimality gap (4) is determined by the contention aggressiveness V and the number of relay links $|\mathcal{L}|$. The intention of multi-path and opportunistic routing is to achieve spatial and multi-user diversity using additional relay. Thus, a tradeoff arises: Every additional relay achieves further diversity gains, but it causes costs in terms of the optimality gap. In particular, the returns of diversity are expected to diminish, whereas the relay costs will not. Hence, an optimal degree of multi-path and opportunism should exist, whereas the diversity benefits will be eaten up beyond that point. Our results suggest that the costs are non-negligible especially in the multi-flow case, so that it becomes difficult for OR to outperform single-path routing.

In detail, the problem is more involved and also affects the intra- and inter-flow adaption. For example, let us consider a diamond topology like Fig. 25a on page 79 without link errors. In the case the system uses single path routing, an efficiency parameter $V = 6.6$ achieves the target TA of 10.8. With multi-path and opportunistic routing, we have to lower V to 5.7 in order to achieve a cumulated TA of 10.8 at the flow source. Thus, the degree of multi-path and opportunism does not only affect the number of relays. Instead, it may alter the working point of the intra-flow adaptation, in addition. A smaller efficiency parameter V intensifies the increase of the optimality gap (4).

Since the degree of multi-path and opportunism may alter the working point of the intra-flow adaptation in terms of V, it also affects the inter-flow fairness adaptation and alters the proportional fair working point. Thus, a fairness-limited flow may become the technologically limited one that determines the V for all neighboring flows by simply adding relay nodes. Furthermore, the inter-flow adaptation introduces a form of global coupling between flows. The larger the degree of multi-path and opportunism, the higher becomes the risk that the flow has to compete with a technologically limited flow and thus, it has to sacrifice some of its efficiency in terms of V in the inter-flow adaptation.

In our future work, we need to understand the interaction between multi-path and opportunism and the optimality gap of UO-CSMA in a systematical way. In particular, we seek an answer to the questions: "What is the optimal degree of multi-path and opportunism?", and "How can it be estimated at runtime?" Furthermore, the fairness objective has to be reconsidered and stated more precisely, since the degree of multi-path and opportunism alters the network capacity region. We have observed that some cost factors are *preventable*, which otherwise contribute to the optimality gap (4). If we reconsider the diamond topology above, then it is obvious that the anycast link on the first hop has no benefits but incurs costs only. The same applies to dead ends within the topology. In the general case, however, such local decisions are not possible.

We see two directions for future improvements. On the one hand, a promising approach is the introduction of a bias into the back-pressure in a way that the optimality remains unaffected. The bias could be a heuristic metric that encourages the efficient usage of resources according to a second and less important objective. The approach has shown promising results using a routing metric as bias [12, 32, 83, 88, 132]. How-

ever, the approach caused severe routing oscillations within our preliminary investigation. Alternatively, better designs and heuristics might mitigate the tradeoff between the UO-CSMA optimality gap and the degree of multi-path and opportunism. For example, sources routes are used with Horizon as proposed by Radunovic *et al.* [99]. This way, the routing decision can be more selective in order to include the most promising paths and opportunities only. In addition, better heuristics have to be developed, which are able to identify the potential of spatial and multi-user diversity in a better way.

Breakdown of the Delay Limiting Approach. Our delay limiting approach is based on the idea to introduce virtual packets and virtual traffic flows. The upper queue limits are adapted according to the packet ingress rate in order to ensure an upper delay limit per hop. Whenever the physical queue limit would be violated in either direction, a virtual transmission takes place. An end-to-end flow of virtual traffic arises from the virtual transmissions. The congestion controller intends to find the fair rate for the joint flow consisting of virtual and physical traffic. Thus, virtual traffic increases the optimality gap in terms of physical throughput and may cause a lower fairness.

From our results, we have observed two particular problems. The more the traffic is spatially spread, the lower the traffic rate becomes on the involved links. Since the upper queue bound cannot be decreased beyond a lower limit, the delay control becomes ineffective. Furthermore, the risk of virtual transmission increases with the spatial spreading due to the generally smaller queue bounds, so that the optimality gap further increases. In the extreme case, the approach might break down in a way that the network does transport virtual packets only instead of physical traffic.

Thus, a tradeoff between delay and the degree of multi-path and opportunism arises. The lower the number of paths and opportunities, the lower the risk is that virtual traffic emerges and thus, the higher the probability that the delay limiting remains effective. Referring back to the previous section, the same two directions apply to handle the tradeoff. We might heuristically prune low-throughput paths in the case they do not significantly contribute to the end-to-end performance but increase the above mentioned risks only. In addition, there are also approaches that try to consider delay within the optimization problem [40, 84, 91, 112, 119].

Variability and Flow-Level Dynamics. In the theory on UO-CSMA, it is often assumed that the step sizes properly decrease until they diminish. From a practical point of view, we might not be interested in diminishing step sizes. For example, due to flow level dynamics or non-stationary changes in the environment, the system has to adapt to the new situation. Arbitrarily diminishing step sizes would slow down the process of convergence to the new operating point. Thus, the step sizes should not diminish arbitrarily. On the other hand, only a weak convergence is expectable for non-diminishing step sizes in a way that the system varies within a small neighborhood of the optimum. In combination with the intra-flow adaptation, the variability causes an additional problem. The intra-flow feedback loop estimates the minimum

efficiency parameter along a route. However, since all estimates are subject to random variations, a bias is introduced into the feedback loop that increases with the length of the route.

Appendix

A. Partial Balance in the Extended CSMA Markov Chain

According to [58, Corollary 9.7], a spatial process is a Markov random field (MRF) if it satisfies the partial balance equations

$$\pi(\boldsymbol{v}) \sum_{w} q(\boldsymbol{v}, \Theta_l^w \boldsymbol{v}) = \sum_{w} \pi(\Theta_l^w \boldsymbol{v}) q(\Theta_l^w \boldsymbol{v}, \boldsymbol{v})$$

with stationary distribution π and transition rates q for each link $l \in G$. The operator Θ_l^w changes the state of link l to w. Plugging (5) into the definition above, we get

$$\begin{aligned}
\frac{\prod_k P(v_k)}{C(\boldsymbol{r})} \sum_w R(v_l, w) &= \sum_w \frac{\prod_k P((\Theta_l^w \boldsymbol{v})_k)}{C(\boldsymbol{r})} R(w, v_l) \\
\prod_k P(v_k) \sum_w R(v_l, w) &= \sum_w \prod_k P((\Theta_l^w \boldsymbol{v})_k) R(w, v_l) \\
P(v_l) \prod_{k \neq l} P(v_k) \sum_w R(v_l, w) &= \sum_w P(w) \prod_{k \neq l} P(v_k) R(w, v_l) \\
P(v_l) \sum_w R(v_l, w) &= \sum_w P(w) R(w, v_l)
\end{aligned}$$

which is the partial balance equation for the truncated Markov chain, in which all links except l are frozen. Thus, the CSMA Markov chain (5) is a MRF if the truncated Markov process satisfies the partial balance equations.

In particular, the conflict relation of link l is determined by one of the conflict types in Fig. 3. Since only link l remains unfrozen, we can think of the truncation process in the depictured Markov chains as the removal of all other columns except for the column that corresponds to the frozen state of the remaining links. What remains is a Markov chain with either three, two or only one state, and it can be shown that the partial balance holds in all cases. Note that it is important that the truncated Markov chain does not contain a transition from state 2 to 0 if there is no probing state (1). Otherwise the truncated Markov chain would be non-reversible, which means that the unfrozen chain is not an MRF anymore.

B. Statistical Entropy of the Extended CSMA Markov Chain

Jiang *et al.* pointed out that directly using the maximum likelihood (ML) estimate leads to non-convexity [48]. Instead, they propose to use the statistical entropy. In the following, we show that this approach is also feasible for our formulation of the

extended CSMA model (cf. section 3). The statistical entropy is defined as

$$H(\tau, \boldsymbol{r}) = -\sum_l u_l(\tau) \log \frac{u_l(\tau)}{p_l(\boldsymbol{r})}.$$

By using the definition of p_l in (5) and $\mu_k^a(\tau) = \sum_l u_l(\tau)\delta_{l,k}^a \ \forall a$, we get

$$\begin{aligned}
H(\tau, \boldsymbol{r}) &= -\sum_l u_l(\tau)[\log(u_l(\tau)) - \log(p_l(\boldsymbol{r}))] \\
&= -\sum_l u_l(\tau) \log(u_l(\tau)) - \log(C(\boldsymbol{r})) \\
&+ \sum_l u_l(\tau) \sum_k \left[\delta_{l,k}^1(r_k - r^s) + \delta_{l,k}^2(r_k - r^l) + \delta_{l,k}^3(r^x - r^s)\right] \\
&= -\sum_l u_l(\tau) \log(u_l(\tau)) - \log(C(\boldsymbol{r})) \\
&+ \sum_k \left[\mu_k^1(\tau)(r_k - r^s) + \mu_k^2(\tau)(r_k - r^l) + \mu_k^3(\tau)(r^x - r^s)\right]
\end{aligned}$$

The adaptive CSMA algorithm solves the following optimization problem:

$$\begin{aligned}
&\max_u -\sum_l u_l \log(u_l) \\
&\text{s.t.} \sum_l \delta_{l,k}^a u_l \geq \lambda_k^a, \ \forall k, a \in \{1, 2, 3\} \\
&u_l \geq 0, \sum_l u_l = 1
\end{aligned}$$

If λ_k^2 is zero, there is no traffic on the link. Hence, it is not necessary to consider the link in the optimization. The partial Lagrangian is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{u}; \boldsymbol{y}) &= -\sum_l u_l \log(u_l) + \sum_{k,a} y_k^a (\sum_l \delta_{l,k}^a u_l - \lambda_k^a), \\
&= -\sum_l u_l \log(u_l) + \sum_{k,a} y_k^a \mu_k^a - \sum_{k,a} y_k^a \lambda_k^a.
\end{aligned}$$

The $\boldsymbol{y}$ are dual variables subject to the first constraint. Given $\boldsymbol{r}$, we know that the equilibrium distribution $u_l = p_l(\boldsymbol{r})$ maximizes $H(\tau, \boldsymbol{r})$. However, the equilibrium distribution also maximizes the above Lagrangian with $y_k^1 = r_k - r^s$, $y_k^2 = r_k - r^l$ and $y_k^3 = r^x - r^s$ because the coefficients $\sum_{k,a} y_k^a \lambda_k^a$ and $\log(C(\boldsymbol{r}))$ do not depend on $\boldsymbol{u}$. In particular, both y_k^1 and y_k^2 only depend on r_k and y_k^3 is fixed, which shows that the a priori knowledge of λ_k^3 is not necessary in order to update r_k. The partial derivate of the Lagrangian with respect to r_k is $\partial\mathcal{L}/\partial r_k = \mu_k^1 - \lambda_k^1 + \mu_k^2 - \lambda_k^2$. From the KKT conditions, we know that $\mu_k^2 = \lambda_k^2$ if $y_k^2 > 0$ or, in other words, $r_k > r^l$. In addition, it holds $\mu_k^1 > \lambda_k^1$ iff $\mu_k^2 > \lambda_k^2$ and vice versa for $r_k > r^l$ by construction. Hence, we

do not need a priori knowledge of λ_k^1 because $r_k \leftarrow r_k + \alpha[\lambda_k^2 - \mu_k^2]$ is a sub-gradient algorithm to update the dual variables. We conclude that the UO-CSMA algorithm updates the TAs via sub-gradients.

C. Anycast Goodput Region With Polynomial Number of Constraints

If we restrict the maximum number of (simultaneous usable) candidates, we can reduce the number of necessary constraints to describe the anycast goodput region (11) to a polynomial number. From the practical point of view, it is often favorable to restrict the number of candidates, because an additional candidate generally causes constant costs for diminishing returns. Let N be the number of neighbors. If we upper bound the number of candidates to a fixed constant C, the number of possible candidate sets is

$$\sum_{c=1}^{C} \binom{N}{c} \leq \sum_{c=1}^{C} N^c \leq CN^C.$$

Hence, the number of candidate sets is polynomial bounded in the number of neighbors up to a degree of C. This way, the number of necessary equations also reduces to a polynomial number, because all constraints in (11) having $|J| > C$ are linearly depending on the set of equations with $|J| \leq C$. To see this, consider the constraint for a fixed candidate set with $|J^\star| > C$. Because the throughput $q_{i,J}$ is zero for all J with $|J| > C$, (11) simplifies to

$$\sum_{j \in J^\star} x_{i,j} \leq \sum_{k=1}^{C} \sum_{K \in \mathcal{P}_k(J^\star)} p_{i,K} \cdot q_{i,K}, \tag{29}$$

where the k-subsets are given by $\mathcal{P}_k$. On the other hand, by summing the constraints having exactly C times an "x" on the left-hand side (LHS) over all combinations which contain the nodes $J^\star$, we get

$$\sum_{J \in \mathcal{P}_C(J^\star)} \sum_{j \in J} x_{i,j} \leq \sum_{J \in \mathcal{P}_C(J^\star)} \sum_{K \in \mathcal{P}(J)} p_{i,K} \cdot q_{i,K}$$

$$\sum_{j \in J^\star} \binom{|J^\star| - 1}{C - 1} x_{i,j} \leq \sum_{k=1}^{C} \sum_{J \in \mathcal{P}_k(J^\star)} \binom{|J^\star| - k}{C - k} p_{i,K} \cdot q_{i,K}$$

The introduced factor on the LHS corresponds to the number of times the term $x_{i,j}$ appears in the sum. We generate all C-subsets of a set of $|J^\star|$ elements and we are interested in how often a particular term appears. Hence, this corresponds to choosing $C - 1$ from $|J^\star| - 1$ elements, since we demand that the considered element is already contained. The introduced factor on the RHS may be derived in the same way.

$$\sum_{j\in J^\star} x_{i,j} \leq \sum_{k=1}^{C} \sum_{J\in\mathcal{P}_k(J^\star)} \frac{\binom{|J^\star|-k}{C-k}}{\binom{|J^\star|-1}{C-1}} p_{i,K}\cdot q_{i,K},$$

$$\sum_{j\in J^\star} x_{i,j} \leq \sum_{k=1}^{C} \sum_{J\in\mathcal{P}_k(J^\star)} \frac{(|J^\star|-k)!(C-1)!}{(|J^\star|-1)!(C-k)!} p_{i,K}\cdot q_{i,K},$$

$$\sum_{j\in J^\star} x_{i,j} \leq \sum_{k=1}^{C} \sum_{J\in\mathcal{P}_k(J^\star)} \prod_{l=1}^{k-1} \frac{C-l}{|J^\star|-l} p_{i,K}\cdot q_{i,K},$$

$$0 < \prod_{l=1}^{k-1} \frac{C-l}{|J^\star|-l} < 1$$

$$\sum_{j\in J^\star} x_{i,j} \leq \sum_{k=1}^{C} \sum_{J\in\mathcal{P}_k(J^\star)} 1\cdot p_{i,K}\cdot q_{i,K},$$

We conclude that (29) is a superset of the given linear combination of C-"x" constraints. Hence, it is sufficient to consider the constraints for candidates set sizes less or equal to C, which is polynomial in the number of neighbors.

D. Simultaneous Active Constraints of the Anycast Goodput Region

The anycast capacity region is given as

$$\sum_{j\in J} x_{i,j} \leq \sum_{L\in\mathcal{P}(J)\setminus\emptyset} p_{i,L} \sum_{K\in\mathcal{P}(N\setminus J)} q_{i,L\cup K}, \forall i\in N, \forall J\in\mathcal{P}(\mathrm{Nb}(i))$$

For the ease of notation, we define $p_\emptyset = 0$ and, furthermore, we drop the index i in the following. Let us associate a Lagrangian multiplier β_J with each constraint. Furthermore, let us assume without loss of generality (WLOG) $x_j > 0$ for all $j \leq j'$ ($j' \geq 1$) and $x_j = 0$ otherwise. The set J' covers all receivers $1,\ldots,j'$. In addition, we assume $p_L > 0$, which means that we only consider the neighbors of node i. The correlation of the PER across different receivers can be arbitrary. However, the return for another candidate has to be diminishing, but it must remain non-negative. In particular, given p_A and p_B for the non-empty receiver sets A and B, the joint packet success probability must satisfy $\max(p_A, p_B) \leq p_{A\cup B} < p_A + p_B$. Furthermore, given a shared receiver set C, the return of using the additional candidates B together with $A\cup C$ should not exceed the returns of adding B to the smaller set A: $p_{A\cup B\cup C} - p_{A\cup C} \leq p_{B\cup C} - p_C$.

In our CSMA model, the routing layer suggests a set of neighbors as potential forwarding candidates C, and the throughput of each used hyperlink $J\in\mathcal{P}(C)$ is positive $q_J > 0$, but possibly arbitrarily small. Note that there are technical constraints

that lower bound the minimal throughput. However, by adapting the system parameters this lower bound can be adjusted. For all unused hyperlinks having at least one receiver out of $N \backslash C$, the throughput is zero and the associated constraints are not considered. This decision is justified in the next paragraph.

We observe that for all $J \not\subseteq J'$, or equally $\exists j \in J : x_j = 0$, the TCs $\beta_J = 0$ are zero. To see this, let us consider the case $M = J \cup \{m\}$ with $J \subseteq J'$ and $m \notin J'$. We assume that constraint β_M is tight and equality holds. The difference between constraints β_M and β_J is

$$x_m \geq \sum_{L \in \{m\} \cup \mathcal{P}(J)} p_L q_L - \sum_{K \in \mathcal{P}(J)} p_K q_{K \cup \{m\}}$$
$$x_m \geq p_m q_m + \sum_{L \in \mathcal{P}(J)} (p_{J \cup \{m\}} - p_J) q_{J \cup \{m\}}$$

The RHS of the above expression is positive since we approach $q_m = 0$ only asymptotically within our CSMA model. Thus, a contradiction is shown and we conclude that the constraint β_M cannot be active.

On the other hand, at least one constraint must be active. Otherwise, we have strict inequality in all constraints. Thus, we can pick one arbitrary hyperlink having $q_J > 0$, and decrease its throughput until the first constraint becomes tight.

Furthermore, we observe that at each level $k = |J|$, there can be at most one constraint $\beta_J > 0$ be active. At level j', there is only one constraint, and we have shown above that it is active $\beta_{J'} > 0$. We proceed as follows: We consider level $m < j'$ and assume that there are at least two active constraints β_{M1} and β_{M2} at that level, what should lead to a contradiction. We partition $J' = A \cup B \cup C \cup D$ in a way that $A = M1 \backslash M2, B = M2 \backslash M1, C = M1 \cap M2, D = J' \backslash M1 \backslash M2$. The summation of both constraints β_{M1} and β_{M2} leads to

$$\sum_{j \in A \cup B \cup C} x_j + \sum_{j \in C} x_j = \sum_{a \in \mathcal{P}(A)} \sum_{b \in \mathcal{P}(B)} \sum_{c \in \mathcal{P}(C)} \sum_{d \in \mathcal{P}(D)} q_{a \cup b \cup c \cup d} (p_{a \cup c} + p_{b \cup c})$$

Now, we consider the sum of the constraints $\beta_{A \cup B \cup C} + \beta_C$. Note that we have to use inequalities because nothing is known about the activity of the involved constraints.

$$\sum_{j \in A \cup B \cup C} x_j + \sum_{j \in C} x_j \leq \sum_{a \in \mathcal{P}(A)} \sum_{b \in \mathcal{P}(B)} \sum_{c \in \mathcal{P}(C)} \sum_{d \in \mathcal{P}(D)} q_{a \cup b \cup c \cup d} (p_{a \cup b \cup c} + p_c)$$

Substitute the LHS with the above derived expression results in

$$0 \leq \sum_{a \in \mathcal{P}(A)} \sum_{b \in \mathcal{P}(B)} \sum_{c \in \mathcal{P}(C)} \sum_{d \in \mathcal{P}(D)} q_{a \cup b \cup c \cup d} \left[(p_{a \cup b \cup c} - p_{a \cup c}) - (p_{b \cup c} - p_c) \right]$$

The term in squared brackets does not depend on D, however, it is negative for all $c = \emptyset, a \neq \emptyset, b \neq \emptyset$ and non-positive otherwise. Since it exists at least one negative term by construction, the RHS of the expression is negative. Thus, the contradiction is shown and we conclude that there can be at most one active constraint per level.

On the other hand, this means that the number of active constraints is upper bounded by the number of candidates.

In addition, we observe that the constraint $\beta_K > 0$ at level k cannot be active if there exists an active constraint $\beta_L > 0$ at level $l < k$ with $L \not\subset K$. To see this, we assume both constraints are active and show the contradiction in the same way as above. The set J' is partitioned according to K and L as above. By summing the constraints $A \cup B \cup C$ and C and subtracting K and L, we get the following.

$$0 \leq \sum_{a \in \mathcal{P}(A)} \sum_{b \in \mathcal{P}(B)} \sum_{c \in \mathcal{P}(C)} \sum_{d \in \mathcal{P}(D)} q_{a \cup b \cup c \cup d} \left[(p_{a \cup b \cup c} - p_{a \cup c}) - (p_{b \cup c} - p_c) \right]$$

Due to our assumption about diminishing returns, the expression in squared brackets is non-positive in general, and negative in the cases $c = \emptyset, a \neq \emptyset, b \neq \emptyset$. Thus, the RHS becomes negative and we conclude that, if there exists an active constraint $\beta_L > 0$ at level l, then at higher levels each active constraint $\beta_K > 0$ contains $K \supset L$.

E. Opportunistic Back-Pressure Routing with Linear Queueing Complexity

In the following, we assume WLOG that the neighbors are sorted according to ascending costs, i.e. $C_j^f \leq C_{j+1}^f$. Using the described approach, the resulting flow rates $\boldsymbol{y}$ and $\boldsymbol{x}$ maximize the problem (12) if the system is in steady state with $\alpha_i^f = C_i^f$ and $\beta_{i,J}^f = TC_{i,j}, J = \{1, \ldots, j\}, C_j^f < C_i^f$ and $\beta_{i,J}^f = 0$ otherwise. Furthermore, $\boldsymbol{\gamma}$ denotes the dual variable for the non-negativity constraint (15). We show this point in a way similar to Radunovic *et al.* [99]. In particular, the KKT conditions for (12) are noted below, and we show that they are satisfied by the proposed approach. As pointed out in the given reference, however, this does not guarantee the existence of the steady state nor the convergence.

$$\sum_j x_{j,i}^f + y^f \mathbf{1}_{i=\sigma(f)} \leq \sum_j x_{i,j}^f \tag{30}$$

$$\sum_{j \in J} x_{i,j}^f \leq \sum_{L \in \mathcal{P}(J)} p_{i,L} \sum_{K \in \mathcal{P}(N \setminus J)} R_{i,L \cup K} q_{i,L \cup K}^f \tag{31}$$

$$0 \leq x_{i,j}^f, 0 \leq y^f, 0 \leq q_{i,J}^f, 0 \leq u_l, 0 \leq \alpha_i^f \tag{32}$$

$$0 \leq \beta_{i,J}^f \tag{33}$$

$$0 \leq \gamma_{i,j}^f \tag{34}$$

$$0 = \alpha_i^f \left(\sum_j x_{i,j}^f - \sum_j x_{j,i}^f + y^f \mathbf{1}_{i=\sigma(f)} \right) \tag{35}$$

$$0 = \beta_{i,J}^f \left(\sum_{L \in \mathcal{P}(J)} p_{i,L} \sum_{K \in \mathcal{P}(N \setminus J)} R_{i,L \cup K} q_{i,L \cup K}^f - \sum_{j \in J} x_{i,j}^f \right) \tag{36}$$

$$0 = \gamma_{i,j}^f x_{i,j}^f \tag{37}$$

$$\alpha^f = K \sum_f U'(y^f) \tag{38}$$

$$\alpha_i^f + \gamma_{i,j}^f = \alpha_j^f + \sum_{J \ni j} \beta_{i,J}^f \tag{39}$$

If the system is in steady state, the conditions (30) are satisfied with equality. This means, in turn, conditions (35) is satisfied, as well. Furthermore, the throughput, goodput, and flow rate are non-negative by construction. Using the proposed approach, the costs $\boldsymbol{C}$ are non-negative. Thus, the conditions (32) are satisfied. (38) follows from (22).

Traffic will be routed along link (i, j) only ($x_{j,i}^f > 0$), if node j is included in the routing decision ($C_j^f < C_i^f$). In this case, we set $\gamma_{i,j}^f = 0$ and conditions (34), (37) and (39) are satisfied. Furthermore, condition (33) holds since $\beta_{i,J}^f$ is either zero or equal to $TC_{i,j}$, which is non-negative by construction.

In the case $x_{i,j}^f = 0$, we do not assign packets to the link $TC_{i,j} = 0$. Thus, it holds $\beta_{i,J}^f = 0 \ \ \forall J \ni j$. We set $\gamma_{i,j}^f = \alpha_j^f + \sum_{J \ni j} \beta_{i,J}^f - \alpha_i^f$. Hence, conditions (33), (39) (37) and (34) are satisfied, since it holds $\gamma_{i,j}^f = C_j^f + 0 - C_i^f \geq 0$.

In addition, the condition (31) is satisfied due to our steady state assumption. If equality applies, the condition (36) is satisfied. If (31) is a strict inequality, then we consider the associated dual variable $\beta_{i,J}^f$. If J contains a node m carrying no traffic ($x_{i,m}^f = 0$), then node m does not have lower costs than node i ($C_i^f \leq C_m^f$). According to our construction above, we know that the TCs $\beta_{i,J}^f = 0 \ \ \forall J \ni m$ and (36) applies. Otherwise, all nodes $j \in J$ have lower costs than i ($C_j^f < C_i^f$) and carry traffic ($x_{i,j}^f > 0$). If J cannot be written as $J = \{1, 2, \ldots, j\}$ in the notation from above ($C_1^f \leq \ldots \leq C_j^f < C_i^f$), then $\beta_{i,J}^f = 0$ and condition (36) applies. In the remaining case, J can be written as $J = \{1, 2, \ldots, j\}$. Due to strict inequality in condition (31), the goodput supply for these receivers is higher than the demand. Thus, the costs C_j^f will increase. Let $K = J \cup \{j+1, \ldots, k\}$ be the smallest set of receivers containing J, for which $\beta_{i,K}^f > 0$. If no such $\beta_{i,K}^f$ exists, the costs C_j^f will eventually reach C_i^f and no traffic will be allocated to node j ($x_{i,j}^f = 0$). If $\beta_{i,K}^f$ exists, then equality applies to the associated constraint. Hence, C_j^f cannot increase above C_k^f and will settle there, which results in

$\beta^f_{i,J} = TC_{i,j} = C^f_k - C^f_j = 0$. We conclude that (36) applies in all cases.

References

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. *SIGCOMM Comput. Commun. Rev.*, 34 (4), 2004. ISSN 0146-4833.

[2] U. Akyol, M. Andrews, P. Gupta, J. Hobby, I. Saniee, and A. Stolyar. Joint Scheduling and Congestion Control in Mobile Ad-Hoc Networks. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008.

[3] A. Aziz, D. Starobinski, P. Thiran, and A. E. Fawal. EZ-Flow : Removing Turbulence in IEEE 802.11 Wireless Mesh Networks without Message Passing. In *ACM CoNEXT*, 2009.

[4] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler. Time-Space Opportunistic Routing in Wireless Ad hoc Networks: Algorithms and Performance Optimization by Stochastic Geometry. *The Computer Journal*, 2009. ISSN 0010-4620.

[5] R. Barr, Z. J. Haas, and R. van Renesse. JiST: an efficient approach to simulation using virtual machines. *Software: Practice and Experience*, 35(6), 2005.

[6] V. Bharghavan, A. J. Demers, S. Shenker, and L. Zhang. MACAW: A Media Access Protocol for Wireless LAN's. In *SIGCOMM '94: Proceedings of the conference on Communications architectures, protocols and applications*, 1994.

[7] U. N. Bhat. *An Introduction to Queueing Theory: Modeling and Analysis in Applications (Statistics for Industry and Technology)*. Birkhäuser Boston, 1 edition, 2008.

[8] S. Biswas and R. Morris. ExOR: opportunistic multi-hop routing for wireless networks. *ACM SIGCOMM Computer Communication Review*, 35(4), 2005. ISSN 0146-4833.

[9] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman. A simple Cooperative diversity method based on network path selection. *Selected Areas in Communications, IEEE Journal on*, 24(3), 2006.

[10] T. Bonald and M. Feuillet. On the stability of flow-aware CSMA. *arXiv cs.PF*, 2010.

[11] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[12] L. Bui, R. Srikant, and A. Stolyar. Novel Architectures and Algorithms for Delay Reduction in Back-Pressure Scheduling and Routing. In *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*. IEEE, 2009.

[13] F. Buschmann, R. Meunier, H. Rohnert, and P. Sommerlad. *A System of Patterns: Pattern-Oriented Software Architecture*. John Wiley & Sons, 2001.

[14] V. G. Cerf and E. Cain. The DoD internet architecture model. *Computer Networks*, 7(5), 1983.

[15] S. Chachulski, M. Jennings, S. Katti, and D. Katabi. Trading structure for randomness in wireless opportunistic routing. *SIGCOMM Comput. Commun. Rev.*, 37(4), 2007. ISSN 0146-4833.

[16] H. Chang, V. Misra, and D. Rubenstein. Fairness and Physical Layer Capture in Random Access Networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*, 2007.

[17] C.-K. Chau, M. Chen, and S. C. Liew. Capacity of Large-scale CSMA Wireless Networks. In *International Conference on Mobile Computing and Networking*, 2009.

[18] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle. Cross-Layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*. IEEE, 2006.

[19] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures. *Proceedings of the IEEE*, 95(1), 2007.

[20] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris. A high-throughput path metric for multi-hop wireless routing. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, San Diego, CA, USA, 2003. ACM.

[21] E. Coviello, A. Bhorkar, F. Rossetto, B. D. Rao, and M. Zorzi. A Robust approach to Carrier Sense for MIMO ad hoc networks. In *ICC*, 2009.

[22] R. Draves, J. Padhye, and B. Zill. Routing in multi-radio, multi-hop wireless mesh networks. In *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*. ACM Press, 2004.

[23] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Least-Cost Opportunistic Routing. In *Allerton Conference on Communication, Control, and Computing*, Monticello, Ill, USA, 2007.

[24] M. Durvy. *Modelling the IEEE 802.11 protocol in wireless multi-hop networks*. Ph.d. thesis, École polytechnique fédérale de Lausanne, 2007.

[25] A. Eryilmaz and R. Srikant. Joint congestion control, routing and mac for stability and fairness in wireless networks. *Selected Areas in Communications, IEEE Journal on*, 28(8), 2006.

[26] A. Eryilmaz, P. Marbach, and A. Ozdaglar. A Fluid-Flow Model for Backlog-Based CSMA Policies. In *International Conference on Wireless Internet*, 2008.

[27] Z. Fu, P. Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla. The impact of multihop wireless channel on TCP throughput and loss. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, volume 3, 2003.

[28] V. Gambiroza and E. W. Knightly. Congestion control in CSMA-based networks with inconsistent channel state. In *International workshop on Wireless internet (WICON)*. ACM, 2006.

[29] V. Gambiroza, B. Sadeghi, and E. W. Knightly. End-to-end performance and fairness in multihop wireless backhaul networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking - MobiCom '04*, New York, New York, USA, 2004. ACM Press.

[30] M. Garetto, T. Salonidis, and E. W. Knightly. Modeling Per-Flow Throughput and Capturing Starvation in CSMA Multi-Hop Wireless Networks. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, 2006.

[31] M. Gast. *802.11 Wireless Networks: The Definitive Guide, Second Edition (Definitive Guide)*. O'Reilly Media, Inc., 2005.

[32] L. Georgiadis, M. Neely, and L. Tassiulas. Resource Allocation and Cross-Layer Control in Wireless Networks. *Foundations and Trends in Networking*, 1(1), 2005. ISSN 1554-057X.

[33] J. Ghaderi and R. Srikant. On the Design of Efficient CSMA Algorithms for Wireless Networks. *arXiv*, 2010.

[34] B. Ghimire, G. Auer, and H. Haas. Busy Bursts for Trading off Throughput and Fairness in Cellular OFDMA-TDD. *EURASIP Journal on Wireless Communications and Networking*, 2009, 2009.

[35] C. Gkantsidis, W. Hu, P. Key, B. Radunovic, P. Rodriguez, and S. Gheorghiu. Multipath code casting for wireless mesh networks. In *Proceedings of the 2007 ACM CoNEXT conference on - CoNEXT '07*, New York, New York, USA, 2007. ACM Press.

[36] D. Gokhale, S. Sen, K. Chebrolu, and B. Raman. On the Feasibility of the Link Abstraction in (Rural) Mesh Networks. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008.

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

[38] G. R. Gupta and N. Shroff. Delay Analysis for Multi-Hop Wireless Networks. In *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*. IEEE, 2009.

[39] G. R. Gupta and N. B. Shroff. Delay Analysis for Wireless Networks with Single Hop Traffic and General Interference Constraints. *Networking, IEEE/ACM Transactions on*, 18(2), 2010.

[40] P. Gupta and T. Javidi. Towards Throughput and Delay Optimal Routing for Wireless Ad-Hoc Networks. In *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*. IEEE, 2007.

[41] P. Gupta and A. Stolyar. Optimal Throughput Allocation in General Random-Access Networks. In *2006 40th Annual Conference on Information Sciences and Systems*. IEEE, 2006.

[42] Z. Haas and J. Deng. Dual busy tone multiple access (DBTMA)-a multiple access control scheme for ad hoc networks. *IEEE Transactions on Communications*, 50(6), 2002. ISSN 0090-6778.

[43] IEEE 802 LAN/MAN Standards Committee. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Standard 802.11, 1999.

[44] IEEE 802 LAN/MAN Standards Committee. IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture, 2002.

[45] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot. Optimized link state routing protocol for ad hoc networks. In *Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century. Proceedings. IEEE International*, 2001.

[46] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia. Exploiting medium access diversity in rate adaptive wireless LANs. In *Proceedings of the 10th annual international conference on Mobile computing and networking - MobiCom '04*, New York, New York, USA, 2004. ACM Press.

[47] L. Jiang. *Optimization and Incentives in Communication Networks*. Ph.d. thesis, University of California, Berkeley, 2009.

[48] L. Jiang and J. Walrand. A Distributed Algorithm for Optimal Throughput and Fairness in Wireless Networks with a General Interference Model. In *Allerton Conference on Communication, Control, and Computing*, 2008.

[49] L. Jiang and J. Walrand. Approaching Throughput-optimality in a Distributed CSMA Algorithm with Contention Resolution. Technical Report UCB/EECS-2009-37, Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, 2009.

[50] L. Jiang and J. Walrand. Approaching throughput-optimality in a distributed CSMA algorithm. In *Proceedings of the 2009 MobiHoc S3 workshop on MobiHoc S3 - MobiHoc S3 '09*, New York, 2009. ACM Press.

[51] L. Jiang and J. Walrand. Convergence and Stability of a Distributed CSMA Algorithm for Maximal Network Throughput. Technical Report UCB/EECS-2009-37, EECS Department, University of California, Berkeley, Berkeley, CA, 2009.

[52] L. Jiang and J. Walrand. A Novel Approach to Model Control Throughput CSMA/CA Wireless Networks. Technical Report UCB/EECS-2009-8, EECS Department, University of California, Berkeley, 2009.

[53] L. Jiang, D. Shah, J. Shin, and J. Walrand. Distributed Random Access Algorithm: Scheduling and Congesion Control. *Information Theory*, 2009.

[54] L. Jiang, J. Ni, R. Srikant, and J. Walrand. Performance Bounds of Distributed CSMA Scheduling. In *Information Theory and Applications Workshop (ITA)*, 2010.

[55] L. B. Jiang and S. C. Liew. Improving Throughput and Fairness by Reducing Exposed and Hidden Nodes in 802.11 Networks. *Transactions on Mobile Computing*, 7(1), 2008.

[56] D. B. Johnson and D. A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. *Mobile Computing*, 353, 1996.

[57] J. Jose and S. Vishwanath. Distributed Rate Allocation for Wireless Networks. *arXiv Information Theory*, 2010.

[58] F. P. Kelly. *Reversibility and Stochastic Networks (Probability & Mathematical Statistics)*. John Wiley and Sons Ltd, 1979.

[59] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. *The Journal of the Operational Research Society*, 49(3), 1998. ISSN 01605682.

[60] T. H. Kim, J. Ni, and N. H. Vaidya. A Distributed Throughput-Optimal CSMA/CA. Technical report, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 2010.

[61] L. Kleinrock and F. Tobagi. Packet Switching in Radio Channels: Part I–Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics. *Communications, IEEE Transactions on [legacy, pre - 1988]*, 23(12), 1975.

[62] M. Kurth and J.-P. Redlich. Carrier sensing and receiver performance in indoor IEEE 802.11b mesh networks. In *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing Connecting the World Wirelessly - IWCMC '09*, New York, New York, USA, 2009. ACM Press.

[63] M. Kurth, A. Zubow, and J.-P. Redlich. Cooperative Opportunistic Routing Using Transmit Diversity in Wireless Mesh Networks. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[64] M. Lacage and T. R. Henderson. Yet another network simulator. In *Proceeding from the 2006 workshop on ns-2: the IP network simulator - WNS2 '06*, New York, New York, USA, 2006. ACM Press.

[65] P. Larsson. Selection diversity forwarding in a multihop packet radio network with fading channel and capture. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5 (4), 2001. ISSN 1559-1662.

[66] P. Larsson and N. Johansson. Multiuser diversity forwarding in multihop packet radio networks. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 4, 2005.

[67] R. Laufer and L. Kleinrock. Multirate Anypath Routing in Wireless Mesh Networks. In *INFOCOM 2009, IEEE*, 2009.

[68] M. Leconte, J. Ni, and R. Srikant. Mixing Time of Glauber Dynamics With Parallel Updates and Heterogeneous Fugacities. In *Workshop on Frontiers of Controls, Games, and Network Science with Civilian and Military Applications*, 2010.

[69] J. Lee, S.-J. Lee, W. Kim, D. Jo, T. Kwon, and Y. Choi. Understanding Interference and Carrier Sensing in Wireless Mesh Networks. *IEEE Communications Magazine*, 2008.

[70] J. Lee, J. Lee, Y. Yi, S. Chong, and A. Prouti. Implementing Utility-Optimal CSMA. In *Proceedings of Allerton conference*, 2009.

[71] J. W. Lee, M. Chiang, and R. A. Calderbank. Utility-Optimal Random-Access Control. *Wireless Communications, IEEE Transactions on*, 6(7), 2007.

[72] S. Liew, C. Kai, J. Leung, and B. Wong. Back-of-the-Envelope Computation of Throughput Distributions in CSMA Wireless Networks. *arXiv Networking and Internet Architecture*, 2007.

[73] S. C. Liew, C. Kai, J. Leung, and B. Wong. Back-of-the-Envelope Computation of Throughput Distributions in CSMA Wireless Networks. In *Communications, 2009. ICC '09. IEEE International Conference on*, 2009.

[74] X. Lin, N. B. Shroff, and S. Member. The Impact of Imperfect Scheduling on Cross-Layer Congestion Control in Wireless Networks. *IEEE/ACM Transactions on Networking*, 2006.

[75] J. Liu and A. L. Stolyar. Distributed Queue-Length based Algorithms for Optimal End-to-End Throughput Allocation and Stability in Multi-hop Random Access Networks. In *Forty-Fifth Annual Allerton Conference On Communication, Control, and Computing*, Monticello, IL, 2007.

[76] J. Liu, Y. Yi, A. Proutiere, M. Chiang, and H. V. Poor. Maximizing Utility via Random Access without Message Passing. Technical report, Microsoft Research, Cambridge, UK, 2008.

[77] J. Liu, Y. Yi, A. Proutiere, M. Chiang, and H. V. Poor. Convergence and Tradeoff of Utility-Optimal CSMA. *Applications of Mathematics*, 2009.

[78] J. Liu, Y. Yi, A. Proutiere, M. Chiang, and V. Poor. Towards utility-optimal random access without message passing. *Wireless Communications and Mobile Computing*, 2010.

[79] M. Lotfinezhad and P. Marbach. Delay Performance of CSMA policies in Multihop Wireless Networks : A New Perspective. In *Workshop on Frontiers of Controls, Games, and Network Science with Civilian and Military Applications*, 2010.

[80] H. Ma and S. Roy. Contention Window and Transmission Opportunity Adaptation for Dense IEEE 802.11 WLAN Based on Loss Differentiation. In *2008 IEEE International Conference on Communications*. Ieee, 2008.

[81] P. Marbach, A. Eryilmaz, and A. Ozdaglar. On the Throughput-Optimality of CSMA Policies in Multihop Wireless Networks. Technical Report CNRL-08, Computer Networks Research Lab, University of Toronto, Toronto, Canada, 2008.

[82] A. Mei and J. Stefa. Routing in Outer Space. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008.

[83] S. Moeller, A. Sridharan, B. Krishnamachari, and O. Gnawali. Routing Without Routes : The Backpressure Collection Protocol. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2010.

[84] M. Naghshvar and T. Javidi. Opportunistic Routing with Congestion Diversity in Wireless Multi-hop Networks. In *IEEE INFOCOM*, 2009.

[85] M. Neely. Optimal Backpressure Routing for Wireless Networks with Multi-Receiver Diversity. In *2006 40th Annual Conference on Information Sciences and Systems*. IEEE, 2006.

[86] M. Neely. Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(8), 2006. ISSN 0733-8716.

[87] M. Neely and R. Urgaonkar. Opportunism, backpressure, and stochastic optimization with the wireless broadcast advantage. In *Asilomar Conference on Signals, Systems and Computers*. IEEE, 2008.

[88] M. Neely, E. Modiano, and C. Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, volume vol. IEEE, 2003.

[89] M. Neely, E. Modiano, and C.-P. Li. Fairness and Optimal Stochastic Control for Heterogeneous Networks. *IEEE/ACM Transactions on Networking*, 16(2), 2008. ISSN 1063-6692.

[90] J. Ni and R. Srikant. Q-CSMA: Queue-Length Based CSMA/CA Algorithms for Achieving Maximum Throughput and Low Delay in Wireless Networks. In *INFOCOM*, 2010.

[91] J. Ni, B. Tan, and R. Srikant. Q-CSMA: Queue-Length Based CSMA/CA Algorithms for Achieving Maximum Throughput and Low Delay in Wireless Networks. *arXiv Networking and Internet Architecture*, 2009.

[92] A. Proutiere, Y. Yi, and M. Chiang. Throughput of Random Access without Message Passing. In *CISS*, 2008.

[93] A. Proutiere, Y. Yi, T. Lan, and M. Chiang. Resource Allocation over Network Dynamics without Timescale Separation. In *Conference on Computer Communications (INFOCOM)*, 2010.

[94] R. Punnoose, P. Nikitin, and D. Stancil. Efficient simulation of Ricean fading within a packet simulator. In *Vehicular Technology Conference, 2000. IEEE VTS-Fall VTC 2000. 52nd*, volume 2, 2000.

[95] D. Qian, D. Zheng, J. Zhang, and N. Shroff. CSMA-Based Distributed Scheduling in Multi-hop MIMO Networks under SINR Model. In *INFOCOM*, 2010.

[96] B. Radunovic and J. Le Boudec. Rate performance objectives of multihop wireless networks. *Mobile Computing, IEEE Transactions on*, 3(4), 2004. ISSN 1536-1233.

[97] B. Radunovic and J.-Y. Le Boudec. Rate performance objectives of multihop wireless networks. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 3. IEEE, 2004.

[98] B. Radunovic, C. Gkantsidis, P. Key, P. Rodriguez, and W. Hu. An optimization framework for practical multipath routing in wireless mesh networks. Technical Report MSR-TR-2007-83, Microsoft Research, 2007.

[99] B. Radunović, C. Gkantsidis, D. Gunawardena, and P. Key. Horizon: Balancing TCP over Multiple Paths in Wireless Mesh Network. In *Proceedings of the 14th ACM international conference on Mobile computing and networking - MobiCom '08*, New York, New York, USA, 2008. ACM Press.

[100] B. Radunovic, C. Gkantsidis, P. Key, and P. Rodriguez. Toward Practical Opportunistic Routing with Intra-session Network Coding for Mesh Networks. *ACM/IEEE Transactions on Networking*, 18(2), 2010.

[101] S. Rajagopalan and D. Shah. Distributed algorithm and reversible network. In *Annual Conference on Information Sciences and Systems (CISS)*, 2008.

[102] S. Rajagopalan, D. Shah, and J. Shin. Network Adiabatic Theorem : An Efficient Randomized Protocol for Contention Resolution. In *Joint International Conference on Measurement and Modeling of Computer Systems*, 2009.

[103] T. S. Rappaport. *Wireless Communications: Principles and Practice (2nd Edition)*. Prentice Hall PTR, 2001.

[104] B. Roman, F. Stajano, I. Wassell, and D. Cottingham. MAC 13-1 – Multi-Carrier Burst Contention (MCBC): Scalable Medium Access Control for Wireless Networks. In *2008 IEEE Wireless Communications and Networking Conference*. IEEE, 2008.

[105] J. Ryuy, J. Lee, S.-J. Lee, and T. Kwony. Revamping the IEEE 802.11a PHY simulation models. In *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems - MSWiM '08*, New York, New York, USA, 2008. ACM Press.

[106] L. Scalia, I. Tinnirello, and D. Giustiniano. Side Effects of Ambient Noise Immunity Techniques on Outdoor IEEE 802 . 11 Deployments. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 2008.

[107] D. Shah and J. Shin. Randomized Scheduling Algorithm for Queueing Networks. *arXiv*, 2009.

[108] D. Shah and J. Shin. Efficient Queue-based CSMA with Collisions. *arXiv*, 2010.

[109] S. Shakkottai and R. Srikant. Network optimization and control. *Found. Trends Netw.*, 2(3), 2007. ISSN 1554-057X.

[110] A. Sridharan, S. Moeller, and B. Krishnamachari. Investigating Backpressure-based Rate Control Protocols for Wireless Sensor Networks. In *ITA workshop*, 2009.

[111] Y. Su, P. Steenkiste, and T. Gross. Performance of TCP in Multi-Hop Access Networks. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, 2008.

[112] V. G. Subramanian and D. J. Leith. Draining time based scheduling algorithm. In *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007.

[113] 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(12), 1992. ISSN 00189286.

[114] F. Tobagi and L. Kleinrock. Packet Switching in Radio Channels: Part II–The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution. *Communications, IEEE Transactions on [legacy, pre - 1988]*, 23(12), 1975.

[115] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[116] M. Valenti and N. Correal. Exploiting macrodiversity in dense multihop networks and relay channels. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, New Orleans, LA, USA, 2003.

[117] P. V. D. Ven, A. Janssen, and J. V. Leeuwaarden. Optimal Tradeoff between Exposed and Hidden Nodes in Large Wireless Networks. *arXiv Networking and Internet Architecture*, 2010.

[118] P. V. D. Ven, J. V. Leeuwaarden, D. Denteneer, and A. Janssen. Spatial fairness in linear wireless multi-access networks. *Performance Evaluation (submitted)*, 2010.

[119] V. J. Venkataramanan and X. Lin. On scheduling for minimizing end-to-end buffer usage over multihop wireless networks. Technical report, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 2009.

[120] B. H. Walke, S. Mangold, and L. Berlemann. *IEEE 802 Wireless Systems: Protocols, Multi-Hop Mesh/Relaying, Performance and Spectrum Coexistence*. Wiley & Sons, 1. edition, 2007.

[121] J. Wang, H. Zhai, and Y. Fang. Opportunistic Packet Scheduling and Media Access Control for Wireless LANs and Multi-hop Ad Hoc Networks. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, 2004.

[122] J. Wang, H. Zhai, Y. Fang, J. M. Shea, and D. Wu. OMAR: Utilizing Multiuser Diversity in Wireless Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 5(12), 2006. ISSN 1536-1233.

[123] X. Wang and K. Kar. Throughput modelling and fairness issues in CSMA/CA based ad-hoc networks. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 1. IEEE, 2005.

[124] A. Warrier, S. Janakiraman, S. Ha, and I. Rhee. DiffQ: Practical differential backlog congestion control for wireless networks. In *INFOCOM*, 2009.

[125] F. Xue and P. R. Kumar. Scaling laws for ad hoc wireless networks: an information theoretic approach. *Found. Trends Netw.*, 1(2), 2006.

[126] Z. Yang, K. Zeng, and W. Lou. FSA: A Fast Coordination Scheme for Opportunistic Routing. In *Communications, 2009. ICC '09. IEEE International Conference on*, 2009.

[127] Y. Yi and M. Chiang. *Stochastic Network Utility Maximization and Wireless Scheduling*, chapter 1. Cambridge University Press, 2010.

[128] Y. Yi and S. Shakkottai. Hop-by-Hop Congestion Control Over a Wireless. *IEEE/ACM Transactions on Networking (TON)*, 15(1), 2007.

[129] Y. Yi, A. Proutiere, and M. Chiang. Complexity in wireless scheduling: Impact and tradeoffs. In *International Symposium on Mobile Ad Hoc Networking & Computing (Mobihoc)*, 2008.

[130] Y. Yi, J. Lee, M. Chiang, and A. Proutiere. Towards Optimal MAC without Message Passing in Wireless Networks. In *Fourth International Conference on Future Internet Technologies - CFI'09*, Seoul, Korea, 2009.

[131] Y. Yi, J. Zhang, and M. Chiang. Delay and effective throughput of wireless scheduling in heavy traffic regimes. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '09*, New York, New York, USA, 2009. ACM Press.

[132] L. Ying, S. Shakkottai, and A. Reddy. On Combining Shortest-Path and Back-Pressure Routing Over Multihop Wireless Networks. In *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, Rio de Janeiro, Brazil, 2009. IEEE.

[133] Y. Yuan, H. Yang, S. H. Y. Wong, S. Lu, and W. Arbaug. ROMER: Resilient Opportunistic Mesh Routing for Wireless Mesh Networks. In *IEEE Workshop on Wireless Mesh Networks (WiMesh)*, Santa Clara, CA, USA, 2005.

[134] K. Zeng, W. Lou, and H. Zhai. On End-to-End Throughput of Opportunistic Routing in Multirate and Multihop Wireless Networks. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008.

[135] X. Zhang and B. Li. Dice: a game theoretic framework for wireless multipath network coding. In *International Symposium on Mobile Ad Hoc Networking & Computing*, Hong Kong, Hong Kong, China, 2008. ACM.

[136] Z. Zhong and S. Nelakuditi. On the Efficacy of Opportunistic Routing. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*, 2007.

[137] H. Zimmermann. OSI Reference Model–The ISO Model of Architecture for Open Systems Interconnection. *Communications, IEEE Transactions on*, 28(4), 1980.

[138] M. Zorzi and R. R. Rao. Geographic Random Forwarding (GeRaF) for ad hoc and sensor networks : multihop performance. *Mobile Computing, IEEE Transactions on*, 2(4), 2003.

[139] A. Zubow. *Kooperatives Forwarding in drahtlosen Maschennetzen*. Ph.d. thesis, Humboldt-Universität zu Berlin, 2009.

[140] A. Zubow, M. Kurth, and J.-P. Redlich. An Opportunistic Cross-Layer Protocol for Multi-Channel Wireless Networks. In *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2007.

[141] A. Zubow, M. Kurth, and J.-P. Redlich. Considerations on Forwarder Selection for Opportunistic Protocols in Wireless Networks. In *2008 14th European Wireless Conference*. IEEE, 2008.

[142] A. Zubow, M. Kurth, and J.-P. Redlich. Opportunistic Protocols in Multi-rate Environments. In *Sensor Technologies and Applications, 2008. SENSORCOMM '08. Second International Conference on*. IEEE, 2008.

Acronyms

ACK acknowledgement. 15, 16, 24, 43, 44, 55, 67, 81

AWGN additive white Gaussian noise. 6, 27, 29–31, 66, 85, 86, 90

BER bit error rate. 27, 30, 66, 90

BPSK binary phase shift keying. 30

BRN Berlin RoofNet. 66

BT busy tone. 23–25, 31, 32, 104

BTC contention busy tone. 23, 24, 26

BTD data busy tone. 24, 92, 95

BTS synchronization busy tone. 24

C node credit. 37, 39–41, 44, 47

CA collision avoidance. 1, 7, 12, 18, 20, 101

CCA clear channel assessment. 27, 28, 31–34, 45, 85, 86, 90, 103

CCDF complementary cumulative distribution function. 28, 29

CDF cumulative distribution function. 90, 91

CS carrier sensing. 1, 2, 7, 8, 10, 11, 15, 17–19, 23, 30, 86, 101–104

CSI channel state information. 78

CSMA carrier sense multiple access. 1, 2, 6–8, 10–14, 17–22, 25, 27, 29, 32, 34, 36, 37, 42–46, 48, 56, 57, 59, 60, 68, 79, 80, 96, 101–104, 108–112

CSMA/HBT CSMA/CA with hierarchical busy tones. 22–25, 27, 32, 67, 102, 103

CTS clear to send. 12, 18, 21, 24, 28, 57, 66, 67, 81, 103

DBTMA dual busy tone multiple access. 7, 23, 102

DCF distributed coordination function. 22, 23

DSR dynamic source routing. 66–68, 74, 80, 83, 90–92, 98

ETT expected transmission time. 15, 64, 85, 86, 96

ETX expected transmission count. 15, 16, 63, 64

EWMA exponentially weighted moving average. 46, 48, 57

ExOR extremely opportunistic routing. 80, 83, 92, 95, 96, 98

FTP file transfer protocol. 6, 51

HFD hidden node free design. 17, 18, 21

i.i.d. independent and identically distributed. 6, 30, 32, 34, 85, 90, 103

IEEE Institute of Electrical and Electronics Engineers. 5–7, 11, 15, 17, 18, 20, 22, 26, 27, 30, 48, 66, 67, 78, 80, 85, 102, 103

IQR interquartile range. 52, 69, 70, 72

ISO International Organization for Standardization. 5, 6

KKT Karush-Kuhn-Tucker. 40, 109, 113

LAD layering as optimization decomposition. 8–10

LAN local area network. 5

LCG link conflict graph. 11, 19, 20, 36, 66, 78, 84, 89, 104

LHS left-hand side. 110, 112

LIFO last in, first out. 14

LQI link quality indicator. 96

MAC medium access control. 5, 6, 8, 9, 13, 15–18, 22, 24, 26, 41, 43, 46, 48, 49, 59, 60, 67, 79, 84, 90, 92, 96, 102, 103

MCS modulation and coding scheme. 6, 10, 30, 34, 49, 64

MIMO multiple-input multiple-output. 5, 10, 104

MIS maximal independent set. 11, 13, 84

MISO multiple-input single-output. 16

ML maximum likelihood. 21, 22

MRF Markov random field. 10, 11, 20, 21, 108

MUD multi-user diversity. 2, 6, 8, 15, 17, 64, 78–81, 84–86, 89, 91, 96, 102, 104–106

MWS maximal-weight scheduling. 9, 10, 12, 13, 15, 17, 24, 55

NP nondeterministic polynomial time. 5, 9, 12, 17

NUM network utility maximization. 5, 8–10, 56

OFDMA orthogonal frequency-division multiple access. 5, 24

OLSR optimized link state routing. 64

OR opportunistic routing. 1, 6–8, 15–18, 34, 36, 40, 84, 92, 96, 101, 102, 104, 105

OSI Open System Interconnection. 5, 6, 102

PDF probability density function. 32

PER packet error rate. 27, 85

PF proportional fairness. 49, 50, 59

PHY physical layer. 5, 8, 16, 26, 27, 30, 48, 66, 85, 104

PID proportional-integral-derivative. 46

PSR packet success probability. 28–36, 43, 57, 66, 76–79, 81, 82, 84, 90, 93, 103

RF radio frequency. 10

RHS right-hand side. 13, 110, 112, 113

RPC randomized pick-and-compare. 13, 14

RTS request to send. 12, 18, 23, 24, 28, 57, 66, 67, 74, 81, 92, 98

SDF selection diversity forwarding. 15

SINR signal to interference and noise ratio. 27, 29–31, 92, 95

SNR signal to noise ratio. 18, 49, 85, 90

STC space-time coding. 5, 16

TA transmission aggressives. 20, 22, 23, 28, 29, 42–48, 50, 52, 53, 57, 60–63, 66–74, 76–84, 90, 103, 105, 110

TC transmission credit. 39–41, 44, 45, 47, 53, 57, 60–63, 77, 78, 81, 112, 114

TCP transmission control protocol. 5, 9, 59, 66–68, 74, 80, 92, 98

TDMA time division multiple access. 9, 17, 50, 86

TX transmission. 30, 85

TXOP transmission opportunity. 28, 57, 81, 92, 103, 104

UDP user datagram protocol. 66–68, 74, 80, 90, 91, 98

UO-CSMA utility-optimal CSMA. 10–15, 17, 43, 45, 47, 54, 55, 59, 74, 84, 85, 89, 91, 92, 96, 98, 101–106

WLOG without loss of generality. 40, 111, 113

WMN wireless mesh network. 1, 2, 5–10, 15–18, 20, 34, 36, 43, 48, 49, 57, 59, 64, 67, 84, 90, 92, 96, 97, 101, 102, 104

WSN wireless sensor network. 15