Technische Universität München
Lehrstuhl für Kommunikationsnetze

Efficient Data Transport in Cellular Multi-Hop Networks

Hans-Martin Zimmermann

Vollständiger Abdruck der von der Fakultät Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. techn. Josef A. Nossek
Prüfer der Dissertation:	1.	Univ.-Prof. Dr.-Ing. Jörg Eberspächer
	2.	Univ.-Prof. Dr.-Ing. Thomas Bauschert, Technische Universität Chemnitz

Die Dissertation wurde am 28.08.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 25.06.2009 angenommen.

Abstract

Future mobile communication is characterized by targeted data rates in the gigabit range. One possible technical realization of these high data rates includes the reduction of the physical distance between transmitter and receiver by the introduction of relay stations. This thesis investigates a new network architecture, where a given cellular network is enhanced by fixed relay stations. The thesis provides a framework to combine the clustering process of the new network into cells and the balancing of load among different cells. It examines routing strategies to enable the handover process between stations of the same cell with respect to their resource efficiency and feasibility. The thesis finally analyzes, whether the application of network coding in cellular multi-hop networks is reasonable. The compatibility to existing standards and the preservation of service quality are important constraints to the new overall concept.

Zusammenfassung

Die Erhöhung der Datenrate ist eine Herausforderung für zukünftige Mobilfunknetze. Die Verringerung der physikalischen Distanz zwischen Sender und Empfänger durch den Einsatz von drahtlosen Relaisstationen ist dazu ein möglicher Ansatz. Die vorliegende Arbeit präsentiert Lösungsvorschläge, wie eine mit Relaisstationen erweiterte Netzarchitektur dynamisch in Zellen untergliedert und dabei ein Lastenausgleich zwischen den Zellen erzielt werden kann. Darüber hinaus werden verschiedene Routing-Ansätze, welche den Handover von Nutzern zwischen den Sendestationen einer Zelle ermöglichen, auf Ressoureneffizienz und Realisierungspotenzial hin untersucht. Eine Analyse, inwiefern Network Coding als Transporttechnik in der gegebenen Architektur sinnvoll angewendet werden kann, rundet die Arbeit ab. Die Kompatibilität zum existierenden Standard und der Erhalt der bestehenden Dienstgüte stellen wichtige Randbedingungen für das neue Gesamtkonzept dar.

Preface

> *The outcome of any serious research can only be to make two questions grow, where only one grew before.*
> *— Thorstein Veblen, economist and sociologist*

This doctoral thesis is the result of several years of research at the Institute of Communication Networks at the Technische Universität München. I want to thank primarily Prof. Dr.-Ing. Jörg Eberspächer for his guidance, his support, his encouragement, and also his critical analysis over the last years. Furthermore, I would like to thank Prof. Dr.-Ing. Thomas Bauschert for his efforts as second examiner and Prof. Dr. techn. Josef A. Nossek as chairman of the examination board. Similarly, I thank Prof. Dr.-Ing. Eckehard Steinbach for the successful cooperation in the organization of the student exchange programs. I also appreciated the helpful hands of Dr.-Ing. Martin Maier and Sabine Strauß in administrative issues, which saved me a lot of time and discouragement in the daily work routine. I really enjoyed working at the institute, which is a result of the collegial spirit. So, I want to say thank to all my current and former colleagues for the interesting talks, useful hints, and gentle gestures in daily work. Particularly, I want to mention my fellows of the Mobile Communications Group, Stephan Eichler, Jan Ellenbeck, Ulrike Korger, Silke Meister, Robert Nagel, and Robert Vilzmann. I want to thank all of them for the fruitful discussions and helpful comments concerning this work. Special appreciation is due to my longstanding office colleague Dr.-Ing. Christian Hartmann for his patience, his gentle humor, and his scientific expertise. Concerning my former colleagues, I'm especially thankful to Dr.-Ing. Ingo Gruber for his initial guidance in my project and to Dr.-Ing. Claus Gruber for the development and friendly provisioning of GRAPH library. Apart from that, I thank Rafael Blanco, Florian Franz, Simon Schubert, and Alexander Seitz in place of all my students for their commitment and great support.

The excellent cooperation with my former research project partners Rüdiger Halfmann and Dr.-Ing. Matthias Lott from Nokia Siemens Networks GmbH and Michael Einhaus from RWTH Aachen within the research project Wireless Gigabit With Advanced Multimedia Support (WIGWAM) funded by the German Ministry of Education and Research has brought up many ideas, which are included in this thesis.

I thank my temporary supervisor Dr. Ying-Chang Liang and my gentle offices mates Lijuan Geng and Yiyang Pei, representatively for all other colleagues at the Institute for Infocomm Research in Singapore. I really appreciated their hospitality, innovative ideas, and helpful discussions.

Last, but not least, I thank my family and friends, which accompanied my work, for their unconditional support, their kind comprehension, and their friendly encouragement. Thank you very much.

München, August 2008 Hans-Martin Zimmermann

Contents

List of Figures ix

List of Tables xiii

1. Introduction 1
- 1.1. Data Transport . . . 1
- 1.2. Contribution . . . 4
- 1.3. Overview of the Thesis . . . 5

2. Cellular Multi-Hop Networks — An Overview 9
- 2.1. Cellular Multi-Hop Networks within the Network Family . . . 9
- 2.2. Classification and Definition of Cellular Multi-Hop Networks . . . 11
- 2.3. The Design of Multi-Hop Networks — Some General Observations . . . 14
  - 2.3.1. Coverage . . . 14
  - 2.3.2. Spectral Efficiency . . . 15
  - 2.3.3. Capital and Operational Expenditures . . . 17
  - 2.3.4. Seamless Handover in Networks with Small Cell Size . . . 18
  - 2.3.5. Capacity of Cellular Multi-Hop Networks . . . 20
- 2.4. Architecture of a Cellular Multi-Hop Network . . . 26
  - 2.4.1. Cellular Multi-Hop as an Extension of LTE System Architecture . 26
  - 2.4.2. System Assumptions . . . 28
- 2.5. Methodology . . . 30
  - 2.5.1. Graph Theory . . . 30
  - 2.5.2. Numerical Optimization . . . 31
  - 2.5.3. Simulation . . . 33

3. Dynamic Cell Clustering in Cellular Multi-Hop Networks 37
- 3.1. Introduction . . . 37
- 3.2. Motivation . . . 37
- 3.3. Related Work . . . 42
  - 3.3.1. Load Balancing in Cellular Networks . . . 42
  - 3.3.2. Clustering in Wireless Multi-Hop Networks . . . 45
  - 3.3.3. Summary . . . 45
- 3.4. A Conceptional Study of Feasibility and Impacts of Dynamic Cell Clustering . . . 46
  - 3.4.1. System Model . . . 46
  - 3.4.2. Numerical Optimization based on Global Knowledge . . . 52

3.4.3. Greedy Algorithm based on Local Knowledge 55
3.4.4. Simulative Study 59
3.5. Towards a Reasonable Protocol Stack 68
3.5.1. Preliminary Thoughts on the Implementation 68
3.5.2. Centralized Approach with Numerical Optimization 69
3.5.3. Decentralized and Hybrid Approach with Greedy Algorithm . . 70
3.5.4. Simulative Study 72
3.6. Conclusion 74

4. Intra-Cell Routing in Cellular Multi-Hop Networks 77
4.1. Introduction 77
4.2. Motivation 77
4.3. Related Work 79
4.3.1. Routing Protocols for Wired Networks 80
4.3.2. Routing Protocols for Wireless Networks 82
4.3.3. Summary 85
4.4. Influence of Routing Metrics on Resource Demand 86
4.4.1. System Model 86
4.4.2. Optimum Solution to the Routing Problem 88
4.4.3. Simulative Study 93
4.5. Efficient Topology Exploration within a Multi-Hop Cell 101
4.5.1. Topology Exploration Protocol WOSPF 101
4.5.2. Inter-Layer Routing Support by MAC Layer 106
4.5.3. Simulative Study 110
4.6. Data Switching in a Multi-Hop Cell 112
4.6.1. Label Switching in Cellular Multi-Hop Networks 112
4.6.2. Comparison between Cellular Multi-Hop Label Switching and MPLS 117
4.6.3. Intra-Cell Handover 118
4.7. Conclusion 121

5. Network Coding in Cellular Multi-Hop Networks 123
5.1. Introduction 123
5.2. Motivation 123
5.3. Related Work 125
5.3.1. Wireless Multicast Graph Model 125
5.3.2. Linear Network Coding Approach 126
5.3.3. Routing versus Network Coding 129
5.4. Minimum Energy Broadcast Transmission 131
5.5. Network Coding in Bidirectional Relaying Scenarios 134
5.6. Conclusion 136

6. Summary and Outlook 139
6.1. Results and Contribution 139
6.2. Outlook 141

A. Simulation Model **143**
A.1. Environment Model 143
A.2. Device Model 145

B. Parameter Setup — Overview **149**

C. Packet Definitions **151**

D. Abbreviations **153**

List of Figures

2.1. Classification of different wireless standards with respect to service and cost efficiency. 11
2.2. Relaying can help to increase coverage and spectral efficiency in a cellular network. 16
2.3. Illustration of the handover problem due to lack of infrastructure density. 20
2.4. Deployment scenarios with different expansion states and reuse factors. 22
2.5. Division of a cell into grid elements and illustration of the natural link capacity within a cell. 23
2.6. The figures illustrate a single-hop and a multi-hop cellular network. While both of them are equal in coverage and apply full frequency reuse in each cell, only the single-hop network suffers from severe interference at the cell border. 24
2.7. Illustration of the link capacity in a Manhattan grid topology. 25
2.8. The LTE system architecture with a relay-based multi-hop extension. . . 27
2.9. Data processing in Access Gateway (AGW) and Enhanced NodeB (ENB). 27
2.10. The picture shows a possible OFDMA frame structure as used in cellular multi-hop networks. Relay stations are assigned subframes in the uplink frame of the access point (UL AP). Relay stations may use these resources to schedule an independent subframe structure with downlink phase (DL RS) and uplink phase (UL RS). 29
2.11. The overall simulation flow around the network simulator ns-2. 35
2.12. Layered node implementation of the wireless devices and extrinsic models, which influence the nodes' interaction. 35

3.1. Example network configuration before and after the reclustering process: load is transfered from cell A to cell C. 39
3.2. Visualization of the Dynamic Cell Clustering strategy as a combination of load balancing and clustering in a cellular multi-hop network. 43
3.3. Randomly meshed network topology with 16 nodes, mesh degree $\gamma = 3$, and $\rho = \frac{3}{16}$. 47
3.4. Torus-shaped example network topology with 360 nodes ($X_{max} = 10$, $Y_{max} = 36$), mesh degree $\gamma = 4$, and access point density $\rho = 0.2$. 49
3.5. Sketch of the influencing parameters for load balancing and reconfiguration in a network with dynamic traffic load. 51
3.6. Plain field example network topology with 28 nodes and mesh degree $\gamma = 3$. 56
3.7. Local environment $\mathcal{C}_l(V_A)$ of access point V_A. 57

3.8. Self-organization of cell clustering process in plain field example network with 28 nodes. 59
3.9. Mesh degree as topologically defined performance boundary for load balancing in torus-shaped network topology with 360 nodes. 63
3.10. Influence of access point density and network size on load balancing in torus-shaped random network topologies with $\gamma = 4$. 65
3.11. Influence of load dynamics on balancing performance in torus-shaped random network topology with 360 nodes. 67
3.12. Message sequence charts for centralized and decentralized reclustering approaches. 71
3.13. Effects of reclustering process in a Manhattan scenario. 74
3.14. Evaluation of the balance index β in a Manhattan example scenario. . . 75

4.1. Each node choses distinct multi-point relay nodes to decrease the overhead for flooding the network. 83
4.2. Multi-hop scenario with single access point, 8 relay stations, and — for illustration purpose — only 10 mobile terminals. 93
4.3. The figures indicate that the choice of the routing strategies has serious impact on the overall resource consumption and the maximum resource consumption within the multi-hop cell. 96
4.4. Resource consumption and rerouting events. 98
4.5. The mutual overlapping coverage area between neighboring routers is an important extrinsic impact factor on the efficiency of the routing strategy. 99
4.6. The multi-hop cell can be divided hierarchically into a static infrastructure network and a dynamic user subnetwork. 102
4.7. The WOSPF routing protocol uses periodical HELLO and corresponding HELLO reply packets to explore local link conditions. The link state information is globally distributed within the cell by LSU messages. . . 104
4.8. Exemplary TDD-OFDMA frame structure for centralized medium access control. 108
4.9. The introduction of a new MAC service interface enables the inter-layer information exchange between MAC and network layer. 108
4.10. User velocity versus signaling overhead. 113
4.11. Beacon interval versus signaling overhead. 113
4.12. Node number versus signaling overhead. 113
4.13. Separation of the multi-hop cell in a static label-switched subnetwork and dynamic user subnetwork. 114
4.14. Intra-cell handover between two relay stations of a multi-hop cell. 120

5.1. Basic principle of network coding. 125
5.2. Graph-theoretical model for unicast and multicast transmission. 126
5.3. Illustration of solution approach with example. 129
5.4. Spheric example scenario and evaluation results. 134
5.5. Connectivity models for simple relaying scenarios. 135

A.1. Illustration of a potential movement defined by the random direction mobility model. 144
A.2. Stepwise approximation of the envelope of the HiperLAN/2 link adaptation curve. 146
A.3. Transmission queues are attached to access points and relay stations. Each packet passes exactly one transmission queue, which models the finite transmission capacity of the wireless medium and its optimum assignment. 146
A.4. The radio resource management unit assigns transmission resources exclusively in a periodic superframe structure. 147

C.1. MPLS header and cellular multi-hop label switching header. 151
C.2. Topology table and routing table entry. 151
C.3. AODV packet format. 151
C.4. WOSPF and OSPF HELLO packet format. 152
C.5. WOSPF and OSPF link state update packet format. 152

List of Tables

2.1. Frequently used graph-related notations. 31

4.1. Classification and comparison of different routing protocols. 85
4.2. The label switching header is included in the data packet in between the layer 3 and layer 2 header. 115
4.3. Comparison between MPLS and cellular multi-hop label switching. . . . 118

B.1. Chapter 3.4. Dynamic Cell Clustering (graph model). 149
B.2. Chapter 3.5. Dynamic Cell Clustering (ns-2 simulation). 150
B.3. Chapter 4.4. routing metrics (general parameter setup). 150
B.4. Chapter 4.4. routing metrics (individual parameter setup). 150
B.5. Chapter 4.5. topology exploration (general parameter setup). 150
B.6. Chapter 4.5. topology exploration (individual parameter setup). 150
B.7. Chapter 5.4. network coding. 150

1. Introduction

1.1. Data Transport

Transport optimization is a considerably old problem. Imagine ancient trading caravans, which have transported valuable goods like spices or gold over thousands of kilometers from a producing area to a consumer area. Intrinsic impact factors as the limited long-distance transportation capacity at that time and external impact factors as the risk of losses due to raids or weather influence are very apparent reasons, which have made efficient transport necessary. Even today's logistic companies with different warehouses are subject to transportation problems, as they are supposed to deliver a given set of goods to distinct customer places. Typically, the source and destination locations are connected by a network of transport paths. The use of a transportation path is related to costs mainly depending on the physical conditions and the length of the path. Intermediate crossings along a path can be used for a rearrangement of the corresponding load.

With the raise of information and computer technology, a new class of transportation problems has been defined: data transport problems. Data source and destination can be considered thereby as information warehouses, while fiber or wireless connections act as roads with given transport capacities to enable the data exchange between source and destination. Continuous data is discretely encapsulated in packets for that purpose. Although the coherence to the classical transport problem is obvious, there is one important difference: Digital data can be broken down from large packets into a single logical bit. The resulting scalability is therefore by far higher than for conventional goods. In addition to that, the physical integrity of a data packet is not a condition for the integrity of the overall information at the final destination, as digital data flows can be logically combined as well as separated again. In other words, quoting Ralf Kötter:"Bits are not cars"[EKM07].

The classical capacity-restricted Vehicle Routing Problem (VRP) as known from operations research [TV02] therefore only partly matches to the requirements of data transport. Data transport in mobile communication networks involves all kinds of tasks, which are covered by the OSI[1] reference model from the physical layer up to the application layer. Let us exemplarily name the signal processing to assure the correct transmission and reception of data over a wireless channel. Alternatively, refer to the higher layer protocol suites to offer sockets for software applications, to preprocess data, and to monitor the end-to-end transmission. In this thesis, we consider the transport problem from the network layer perspective. We consequently concentrate on the determination of optimum transmission paths between a backbone node and associated mobile terminals in wireless multi-hop networks. In doing so, it is possible

---

[1] Open Systems Interconnection (OSI)

to distinguish three principle approaches to solving this transport problem: single-path routing, multi-path routing, and network coding. We will take a closer look on the difference in between these approaches in the following.

Single-path Routing Single-path routing is a transport approach, in which a unique path between source and destination is defined upon a network topology. Data is subsequently forwarded on that path. Single-path routing is an adequate strategy, whenever the transportation costs on distinct links play the dominant role in the optimization process and the capacity on these links is large compared to the demand of individual data flows. Fixed costs can be assigned to distinct path segments in this case. Consequently, the routing problem can be solved as a minimum cost problem, whose optimum solution is expressed by a single routing path. Choosing an alternate path or a set of alternate paths will not improve the overall cost situation, even if several independent paths achieve equal costs. A single path is thus sufficient to provide the optimum solution. Routing is generally subject to the assumption that incoming data packets are decoded at any intermediate node and subsequently forwarded to the next-hop destination. The logical integrity of the packet content remains untouched during the whole transfer.

Multi-path Routing Multi-path routing is an extension of the single-path routing approach, as multiple edge-disjoint paths between source and destination are determined for the transport. Multi-path routing can be applied for resilience purpose, in case of one or several path segments failing unexpectedly. In doing so, redundant information is transmitted in parallel to compensate potential failures. We will not consider this type of error prevention mechanism. Alternatively, let us turn our attention to the ability of multi-path routing to distribute the overall data onto several paths. Multi-path routing can thus achieve better results than single-path routing, for instance in terms of end-to-end throughput, when the capacity limitation of certain links is the dominant property of the network. Imagine the capacity of a link to be small compared to the demand of a single traffic flow. The distribution of the distinct traffic flow onto several independent routing paths can then increase the throughput significantly. In contrast to wired networks, wireless multi-hop networks are characterized by a fixed amount of available transmission resources, which are assigned to arbitrary links within the network. The individual link capacity is flexible and adjustable to the demand of a data flow. We are therefore looking for a suitable path assignment, which minimizes the overall resource requirement for data transmissions within the network. Suppose that spatial frequency reuse within the multi-hop network is impossible, as the corresponding area is considerable small and mutual interference would be too strong. We also imply that the considered transmission resources are fully orthogonal and physical effects like inter-carrier interference are negligible. Fixed costs can be assigned to each link in this case. These costs depend on the transmission distance and behave inversely to the physical link quality indicators. Consequently, the intended optimum resource assignment is determined by the solution of a minimization problem, more precisely of a minimum cost problem. As already explained, the solution of such a problem is given by a single routing path. We will therefore not consider multi-path routing in the following and concentrate our analysis to single-path routing.

Network Coding Network coding can be considered as a generalization of the routing concept. Network coding allows data flows to be logically combined and separated at any intermediate network node. In contrast to routing, the logical integrity of the data flow is not guaranteed anymore during the transfer. Incoming data is decoded and processed by intermediate nodes. The output of this process is forwarded to other nodes then. In the special case of a point-to-point connection between a single source and a single destination node, the network coding solution is equivalent to the multi-path routing solution. Although network coding is not restricted to multicast applications, network coding can achieve significant performance gains, for example regarding the number of required transmission resources, in point-to-multi-point scenarios, where a single source node is connected to an arbitrary number of destination nodes. If fixed costs are assigned to individual links again, the network coding problem can be written as a minimum cost optimization problem as well. The network coding solution potentially allows an arbitrary number of independent paths from the source node to each of the multicast nodes. In contrast, the single-path routing solution is restricted to one of the feasible paths. The possible set of routing solutions is therefore described by a subset of potential network coding solutions, as will be explained later on.

This initial classification has shown that there are different solution strategies to the data transport problem. Although discovered only a few years ago, network coding provides a general solution to the data transport problem. Historically, single-path or multi-path routing have been studied extensively as solution to the data transport problem. For this reason, the term "routing problem" is still often used as synonym to "data transport problem". This habit generates the paradox that routing is referred to as a generalization of the network coding problem on the one hand, but as its special case on the other hand. We will follow a nested view in this thesis. Accordingly, the overall problem category is referred to as data transport problem, whose general solution is described by network coding. The single-path routing solution is in turn a special case of the network coding solution.

It is obvious that any data transport problem can be divided into two different subproblems, namely the route planning as initial task and the subsequent data delivery itself. The route planning is associated with protocol design and signaling issues to explore the network topology and to determine the optimum route. The data delivery part mainly deals with addressing issues to practically realize the data transfer. This thesis is entitled as an efficient data transport scheme, whereby the term "efficient" refers to the trade-off between signaling overhead for the route planning process and resource consumption for the data delivery. The less signaling overhead will be invested in the optimization of the route planning, the higher the necessary resource consumption for the data delivery will typically be. The challenge is therefore to keep both factors balanced on a small overall level. One key restriction during this process is the warranty of quality of service in a network. Economically, quality of service can be interpreted as a trust relationship between a provider and a subscriber on previously agreed resources. Technically, quality of service is seen as the provisioning of these resources within the physically feasible framework. This thesis investigates data transport in cellular multi-hop networks from the operator's perspective: On the one hand, the operator wants to have a satisfied costumer base. This strongly correlates with the realization of the depicted quality of service issue. On the other hand, the operator wants to maximize the ratio between physical resource input and resulting capacity output. The presented proposals within this thesis are technical approaches to fulfill these ambitious requirements.

1.2. Contribution

This thesis investigates a relay-extended cellular network architecture. The concept enriches paradigms from today's cellular networks with the idea of self-organization and other aspects known from ad hoc networks. The beauty of the concept is based on the fact that it is an extension of the Long-Term Evolution (LTE) standard and thus fully compliant to the current draft. However, the concept also enables a new philosophy, a different deployment strategy, and therefore an innovative operation of a cellular network.

Most sources from literature consider fixed relay stations as a mechanism to extend coverage of a base station or to achieve higher spectral efficiency. This thesis presents relay stations as the missing link between current and future cellular network architectures: Today's network providers are afraid of the increasing competition by self-administrating low-cost Wireless Local Area Network (WLAN) technology. We show in this work that relay stations can be used to introduce self-organization in cellular networks and to cut down costs without losses in service quality. In detail, the contribution includes three major issues:

- *Dynamic Cell Clustering: a concept how to establish a cellular multi-hop network*

  Existing work about relay-based cellular networks typically considers a permanent assignment of a fixed relay station to a distinct access point. This thesis shows that the dynamic assignment of relay stations to access points can increase the network capacity. The assignment itself is motivated by the idea to equalize traffic load among different cells. We investigate the dependency between network topology and achievable load balancing gains. The result of this analysis gives insight into the network planning process of relay-based wireless networks as well as into the application of self-organization in cellular networks.

- *Intra-cell routing: a proposal how to operate a single multi-hop cell*

  Starting with an analysis of existing routing protocols, we develop a proprietary routing protocol for the operation of a single multi-hop cell. We show that a proactive routing protocol is not necessarily related to high signaling overhead, if link state information of the underlying Medium Access Control (MAC) protocol is used efficiently. Additionally, we investigate, how the topological characteristics of the network relate to the choice of a routing metric in order to obtain a spectrally efficient routing path. A very important observation is the property that the overlapping coverage area between two wireless routers and the link conditions of the relay links are key drivers to influence the efficiency of a routing path.

- *Network coding: a comparison of routing and network coding in cellular multi-hop networks*

  We investigate the application of network coding to cellular multi-hop networks. Firstly, we consider broadcast applications. We show, how network coding can help to save transmission resources in a relay-extended cellular multi-hop network and indicate potential performance gains in dependence on the access point density of the network. Secondly, we take a closer look on the case of bidirectional relaying and address the feasibility of bidirectional relaying in cellular multi-hop networks.

Parts of this thesis have been previously published in [ZL04], [ZG05], [ZLE05], [ZLA05], [Reu06] [ZHW06], [ZFH06], [ZSH06], [Zim06], [EEH^{+}07], and [ZL08].

1.3. Overview of the Thesis

Future mobile communication will be characterized by a significantly higher demand in available data rates. This statement is without controversy among experts from academia as well as from industry. We are talking about target data rates somewhere in the gigabit domain. Although these data rates will not be reached everywhere, these ambitious goals seem to be realistic for city center areas and users with low mobility. First prototypes already indicate that suitable commercial systems will be realizable in about five to ten years from now on.

Even in times of advanced antenna technologies and sophisticated signal processing algorithms, the physical path loss between transmitter and receiver is still a dominant limitation for the achievable link capacity according to Shannon's Theorem. Therefore, one probable technical realization of these high data rates includes the reduction of the physical distance between transmitter and receiver. The natural consequence is a tremendous increase in the number of required wireless infrastructure stations, typically known as access points or base stations. It turns out that the necessary density of access points is economically not realizable with only access points. Therefore, the introduction of access points without wired backbone access, so-called relay stations, is a direct consequence of economic constraints.

In this thesis, we investigate a cellular network architecture, where a given cellular network is enhanced with a large amount of fixed relay stations, i.e. about four to eight or even more relay stations per access point on average. This architecture is referred to as cellular multi-hop network architecture. The challenge of this thesis is to establish a proposal, how to operate such a wireless multi-hop network. Thereby, one very basic requirement is the compliance of the new concept to existing cellular standards. Compliance implies that the new proposal shall be an extension of the existing standard. Compliance also implies that features and service quality known from today's cellular networks shall not be restricted in any regard by the relay-based extension.

This thesis applies three major methods for the scientific investigation: graph theory for the modeling of wireless networks, numerical optimization for the calculation of potential performance gains, and simulation for the verification of the proposed protocol stacks. The thesis includes four major chapters, the first of which presents a general motivation for the relay-based network concept. We explain key drivers for this type of network architecture and identify its critical aspects. We also give a brief explanation, how the presented architectural relay concept is embedded in today's network standards. Additionally, we comment on the question, how the proposals of this thesis relate to the future LTE standard from the technical point of view. We characterize possible network types, address the economic benefits of the relay concept, and present key parameters for the design of a cellular multi-hop network.

Each of the following three main chapters is arranged similarly: At first, we explain the necessity and motivation of the distinct technical innovation, which we want to introduce hereafter. Subsequently, a survey of existing related approaches is given. This is intended to illustrate the advantages and disadvantages of these approaches. It is followed by an explanation, about how much these approaches incorporate in our method and differ from our own proposal. A graph-theoretical analysis identifies potential sources of performance gains in the main part of each chapter. Afterwards, this thesis comes up with a practical implementation of the presented ideas in a real protocol stack. Each chapter ends with a short summary. The composition of these chapters can be considered as a circle from the illustration of the problem, over the identification of potential improvements, to the realization of a proposal, concluding with a final verification of the initial problem being solved.

Chapter 3 is focused on the question, how to establish a cellular network based on the unstructured set of access points and relay stations. For that purpose, we propose a Dynamic Cell Clustering strategy to attach relay stations to distinct access points. This strategy combines the access point discovery mechanism with load balancing among different cells. We develop a centralized and a distributed strategy and show, how potential performance gains depend on the network topology itself. The result is that Dynamic Cell Clustering can help to reduce the maximum cell load significantly and thus extend the network's overall capacity. Additionally, it is shown that a decentralized strategy and therefore self-organizing network architecture is feasible for networks with moderate load dynamics. The interested reader should take especially notice of the fact that Dynamic Cell Clustering adds elements of self-organization to cellular networks and thus is one potential approach to reduce the gap between Wireless Local Area Network architectures and typical cellular networks.

Assuming that the network has been already divided into different cells as a direct result of the clustering procedure, we try to find an operational concept for a single multi-hop cell in Chapter 4. The question is, how to efficiently guide data from a backbone node to a mobile user terminal. We take a look on today's routing protocols and the ambivalence between reactive and proactive approaches. Subsequently, we focus on the question, how the choice of a suitable routing metric will influence the overall resource consumption. One significant result is that the overlapping area between neighboring routers heavily influences the choice of the routing metric. Moreover it turns out that a routing strategy, where a mobile terminal is attached to the station with best received signal strength, can lead to comparably high resource consumption. In the second part of this chapter, we can show that proactive approaches are not necessarily subject to higher signaling overhead than reactive approaches. This observation holds for networks, where the available link state information of a centralized MAC can be used. Finally, we illustrate that label switching is a suitable addressing mechanism for the operation of a multi-hop cell.

In Chapter 5, we focus on the application of network coding to cellular multi-hop networks. We investigate the question, whether network coding outperforms conventional routing in terms of resource consumption for broadcast applications. The results indicate that potential performance gains will primarily depend on the access point density. Subsequently, we comment on the feasibility of bidirectional relaying in cellular multi-hop networks. The thesis ultimately concludes with a summary of all preceding results and a corresponding outlook about further research activities.

2. Cellular Multi-Hop Networks — An Overview

2.1. Cellular Multi-Hop Networks within the Network Family

Innovation in wireless communication networks is characterized by a trend to provide services with advanced quality at low costs per bit. Although the starting point and evolution path of various wireless network technologies are different, all technologies aim at this common trend. Figure 2.1 shows a two-dimensional chart, where the abscissa roughly outlines costs for deployment and operation of a technology and the ordinate describes the degree of service quality provided by the distinct technology. According to this chart, the technology with highest service quality at lowest costs is located in the upper left corner of the second quadrant. Obviously, this point represents the optimum technological development, which research activities try to approach step-by-step. A closer look to ongoing developments reveals that today's standards indeed move on this two-dimensional chart toward the outlined goal. In detail, we observe

- *a trend to cut down costs while keeping or even extending the service in cellular networks:* Initially starting with the first analog generation of wireless communication networks, the big economic success of cellular networks has been accomplished by the digital second generation (2G) of mobile communication networks based on the Global System for Mobile Communications (GSM) [EVB01] standard. Currently, providers are dealing with the controversial third generation (3G) of cellular networks represented by the Universal Mobile Telecommunications System (UMTS) [ETS98] and its enhancements High Speed Downlink Packet Access (HSDPA) [3GP08b] and High Speed Uplink Packet Access (HSUPA) [3GP08c]. The major motivation of cellular network providers is the offer of area-wide voice, data, and even multimedia services. The future of cellular networks is defined by the LTE standard, which is supposed to use Orthogonal Frequency Division Multiple Access (OFDMA) and to enable multi-antenna transmission technology [Poo07]. LTE is mainly intended to result in higher spectral efficiency and smaller costs per bit [Sch08]. This shall be achieved by technological changes in the air interface and by architectural changes in the radio access network. Although the LTE standard is not completely fixed yet, scientists already think about an extension of the LTE standard. The vision is to extend the cellular network topology by additional fixed relay stations and thus to support cellular multi-hop communication. Relaying is thereby considered as a technical approach to provide cheaper network deployment and maintenance combined with a comparable or even higher capacity than corresponding cellular single-hop networks. The classification of Figure 2.1 shows this trend of cellular network technology as a horizontal approach from the upper right corner to the optimum in the upper left corner.

- *the desire to provide quality of service in wireless low-cost networks:* IEEE[1] 802.11 WLAN standard [Com99a] is a popular wireless network technology. Its minor spectral efficiency is well-known, but compensated by its simple handling and cheap usage. IEEE 802.11 is mainly used as wireless access technology to provide Internet access in a home or office environment and in highly populated hot-spot areas. In addition to that, WLAN technology can also be applied to mutually connect wireless devices into an autonomous and self-organizing ad hoc network. The original IEEE 802.11 standard has been extended by continuously improved air interfaces published as IEEE 802.11a [Com99c], IEEE 802.11b [Com99b], IEEE 802.11g [Com03], and future IEEE 802.11n [Xia05] standard. The available maximum physical raw data rates have been increased from 11 MBit/s up to about 100 MBit/s. Furthermore, IEEE 802.11i [Com04] has put effort to assure security, while IEEE 802.11e [IEE05b] is known for the introduction of traffic classes to provide better service guarantees. The future IEEE 802.11s standard [IEE06a] will create meshed wireless infrastructure networks to enable wireless communication at very low costs. Figure 2.1 illustrates this development in a portfolio placement from the lower left corner toward the upper left corner of the chart.

- *the idea to replace wired by wireless services:* Probably the most prominent service of this category is IEEE 802.16 [IEE04], better known as Worldwide Interoperability for Microwave Access (WIMAX) [LQLG07], with the original motivation to replace wired Digital Subscriber Line (DSL) services. IEEE 802.16 can be seen as the legal successor of the technologically interesting, but commercially disastrous ETSI[2] High Performance Radio Local Area Network (HiperLAN) Type 1 (HiperLAN/1) [ETS96] and Type 2 (HiperLAN/2) [ETS00] projects. HiperLAN can be interpreted as a compromise between cost and service efficiency. Thus, HiperLAN is positioned in the point of origin of the chart. The extension IEEE 802.16e [IEE05a] provides mobility to the IEEE 802.16 standard. IEEE 802.16j[3] enables the use of fixed relay stations such that analogies to cellular multi-hop networks are clearly visible. Figure 2.1 classifies this trend as an arc-shaped movement toward the second quadrant.

The different approaches obviously move on to a common goal: to combine a high service quality with respect to data rate, security, delay, and mobility support with low-cost infrastructure components. It is possibly unrealistic to assume the arrival

---

[1] Institute of Electrical and Electronics Engineers (IEEE)

[2] European Telecommunications Standards Institute (ETSI)

[3] IEEE 802.16j Relay Task Group is currently developing a standard draft "Air Interface for Fixed and Mobile Broadband Wireless Access Systems — Multihop Relay Specification".

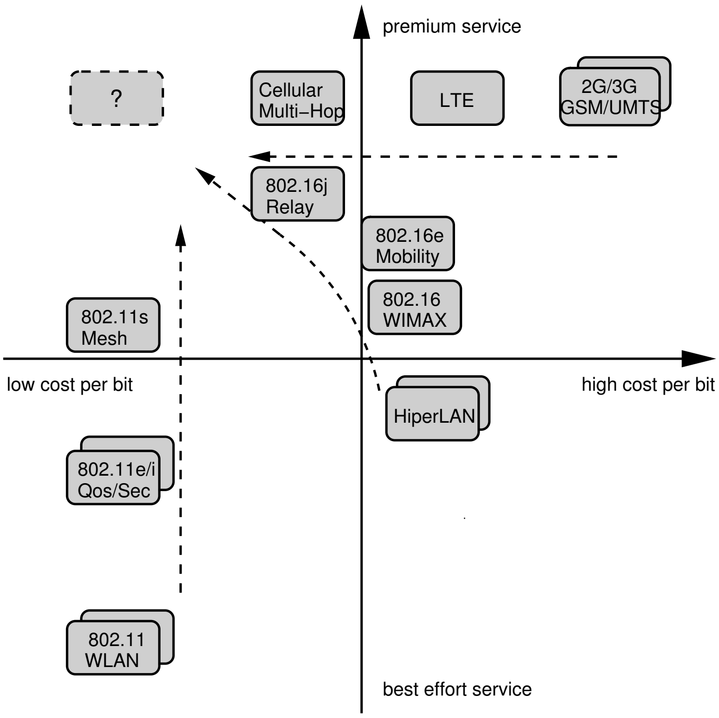


Figure 2.1.: Classification of different wireless standards with respect to service and cost efficiency.

of a future superior technology, which will occupy the question mark's position in Figure 2.1. More probably, the different technologies will converge into a common framework. Figure 2.1 thus can also be interpreted as an illustration of a phenomenon, which is commonly referred to as fourth generation (4G) convergence of mobile communication networks. We can summarize that service quality and cost efficiency are the major technological drivers in the evolution of today's wireless communication networks. We additionally record that the relay-extended cellular network architecture will play an important role in the future of mobile communication. We will therefore concentrate on this special network architecture in the following and characterize different forms of cellular multi-hop networks first of all.

2.2. Classification and Definition of Cellular Multi-Hop Networks

Present wireless communication networks are typically single-hop networks, where base stations or access points provide wireless access for mobile users to some kind of wired infrastructure network. Nevertheless, there can be more than a single wireless hop between the backbone-connected gateway device and the mobile terminal, when relay stations act as forwarding nodes for other devices. This type of network architecture is typically referred to as multi-hop access network. We identify three typical occurrences of multi-hop access networks, which are distinguished by the specific properties of the relaying device:

Categories of Relaying Technologies

Ad Hoc Relaying Probably the most apparent form of a multi-hop access network uses ad hoc relaying nodes. In this case, any available ad hoc device in the network may act as relay node for other mobile terminals [GBL03]. Christian Bettstetter characterizes ad hoc networks by direct link connections between unreliable devices, by the abandonment of fixed infrastructure, by the establishment of multi-hop connections, and by the lack of centralized databases [Bet04b]. Ad hoc relaying helps to extend the coverage of base stations [GBL03] and to increase the duration of a physically feasible connection, i.e. to lengthen the path life time [GL04b]. [GL04a] and also [ACM07] promote the idea of a coverage extension of a cellular base station by ad hoc relaying based on IEEE 802.11 technology. The nature of ad hoc systems typically implies that the relaying devices are owned and operated by private users. Therefore, a very crucial point is the motivation of private users to provide relaying services to other users. Even though game theoretical approaches exist to achieve a stable operating point, it is very hard to realize permanent quality of service. This is due to the unpredictability of ad hoc scenarios. For these reasons, it is unlikely that conventional network operators will enforce the establishment of ad hoc relaying services.

Fixed Relaying Fixed relaying is characterized by relaying stations, which are permanently deployed at fixed positions. Fixed relay stations act like access points from a mobile terminal's perspective. However, relay stations do not have any wired connection to the backbone infrastructure network, but establish wireless single-hop or multi-hop connections to access points. For this reason, relay stations are much cheaper in deployment than conventional access points, as power supply is the only constraint and the fiber connection becomes superfluous. This also implies that relay stations are operator-owned or at least operator-steered devices. Traffic lights or street lighting could be potential locations for fixed relay stations. In doing so, an easy deployment is possible, unlimited power supply is granted, and additionally good signal propagation characteristics could be expected because of the exposed position. According to current literature [Pab04], fixed relaying in cellular networks is primarily intended to gain coverage or to achieve higher spectral efficiency in a cost-efficient way. Further research in [FAMZ03] and [HCC05] moves into a different direction assuming multi-hop communication with fixed relay stations based on IEEE 802.11 technology.

Mobile Relaying Mobile relaying is an architectural concept somewhere in between fixed and ad hoc relaying. Again, the intention is to provide coverage in areas without coverage. Mobile relaying has been extensively studied in the WINNER[4] project. Mobile relay stations are wireless devices, whose positions are not restricted to fixed

---

[4] Wireless World Initiative New Radio (WINNER) is an international research project under framework program 6 of the European Commission.

locations. However — in contrast to ad hoc relaying — these relay stations are owned and operated by the network provider. We distinguish two major categories:

- *temporal network access:* While fixed relays provide permanent service, mobile relaying is used for temporal network access. Possible scenarios include major events with unusually high user demand, for example a soccer match, concerts, maybe also a conference, or generally all kinds of public events. The technical challenge in this case is the optimum placement and relocation of these wireless routers [WXCA07] to achieve appealing performance and to absorb peak loads.
- *mobile routers:* Additionally, mobile relaying also includes fixed devices mounted on mobile locations, for example buses or trains of a public transportation system [BL05]. In this case, the relaying device is permanently mobile and therefore only temporarily attached to a distinct access point. Given the bus example, a group of users can even travel around with a relay station. Thus, the users stay connected to the relay device, while the relay device is changing its serving access point [WYT$^+$05]. The major challenge of this scenario is the guarantee of a permanent wireless access of the relay device to a given infrastructure network. The difficulty is especially the seamless handover of the mobile relay stations with its attached users between different access points.

Cooperative Relaying Contemporary research literature [KGG05] introduces the notion of "cooperative relaying". The majority of publications follows the definition of cooperative relaying as a concept, which exploits the diversity of the wireless channels. Corresponding examinations rely on a model referred to as "relay channel". The relay channel is a link constellation, in which a wireless device can communicate on a direct link and an alternative two-hop link with another device. Transmissions on both of the two alternate paths result in redundant information at the destination device, which can be exploited in the decoding process to achieve performance gains. Cooperative relaying is thus not a supplement to the three presented categories. Instead it is a concept, which can be applied to any of those categories. The major obstacles in the practical realization of cooperative networks are the mutual dependency between devices, the high coordination effort, and reliability constraints [DMS08]. The complement to the cooperative relay channel is known as two-hop channel and considers the case, in which the direct transmission is infeasible. Communication is fully relying on the intermediate relay station then. Coverage extension is the typical application scenario of this two-hop channel. In the following, we will concentrate on routing issues and therefore primarily work with the non-cooperative two-hop channel model.

Characteristics of a Cellular Multi-Hop Network

Considering the cellular principle from a more general perspective, a set of exclusive transmission resources is assigned to a predefined local service area in order to enable wireless communication between a distribution system, i.e. typically the base station, and mobile terminals. If a single access point and a variable number of fixed relay stations become part of the distribution system, this special type of multi-hop access network is defined as cellular multi-hop network. Assuming that the functionality of access points and base stations will become indistinct in future networks, these expressions will be used synonymously in the following. According to the cellular principle, a multi-hop cell is a set of relay stations and a single access point, which share the same wired access to the backbone network and a common set of wireless transmission resources. The relay stations are assigned to a distinct access point as gateway node for that purpose. Access points as well as fixed relay stations act as wireless routers, which provide service to mobile user terminals. We will use the term "wireless routers" in the following to address the combined set of access points and fixed relay stations within the network. The cellular multi-hop network itself is established by a set of multi-hop cells. This type of network architecture primarily intends to provide coverage and services for a local area with dense traffic load. We will rely on this definition subsequently and consequently restrict the further analysis to the case of fixed relaying. Let us start with some very general design issues about relay-based cellular networks for that purpose.

2.3. The Design of Multi-Hop Networks — Some General Observations

2.3.1. Coverage

Cellular networks are characterized by the network operator’s efforts to provide sufficient network coverage to mobile subscribers in a certain area of interest. It has been shown that infrastructure-based, i.e. fixed, relay stations or mobile relay stations can help to extend the coverage area of a given base station [Pab04]. Future cellular networks will require data rates in the gigabit range in order to provide advanced multimedia services. These data rate requirements are likely to be realized by the use of transmission frequencies in the 2.5 GHz or even in the 5 GHz spectrum. The idea of coverage extension by relaying becomes especially important, when frequencies in the 5 GHz domain are used. In this case, cell size will shrink dramatically due to propagation path losses. The amount of required infrastructure stations to achieve the same coverage thus increases dramatically. Relay stations extend an existing access point’s coverage and therefore are an alternative to the allocation of new access points instead.

In contrast to ad hoc relaying, fixed relaying provides permanent and reliable coverage. [PMB08] examines the capacity of IEEE 802.16 base stations, whose coverage is extended by fixed relay stations for the two-hop case. The analysis shows that there is a trade-off between coverage extension and cell capacity. As the coverage area becomes larger, the average cell capacity will decrease for uniformly distributed users as a result of higher propagation losses. However, it is also shown that the decrease in cell capacity can be partly compensated by the efficient use of radio resource management skills: The use of relay stations may enable a higher spatial frequency reuse, if the same resource can be used simultaneously at multiple relay stations. The results of the analysis in [PMB08] explicitly recommend the replacement of relay stations by new base stations, once the capacity limit is reached. Following this idea, the concept of coverage extension by fixed relay stations is especially interesting during the initial deployment phase of a new network or for networks, which are permanently operated far beneath capacity, for instance in rural areas.

Example To illustrate the potential of coverage extension, we consider an access point, whose coverage is extended by four fixed relay stations[5] as shown in Figure 2.2(a). Assuming the same maximum transmission radius R at the access point and the relay stations, the transmission area of the initial access point equals $A_{ap} = R^2 \cdot \pi$, if neither fading effects nor shadowing are taken into account. The additional use of relay stations extends the overall multi-hop coverage area to $A_{mh} = 4 \cdot R^2 + 4 \cdot \frac{R^2 \cdot \pi}{2}$. In doing so, the percentage of coverage extension $\frac{A_{mh}}{A_{ap}} = \frac{4 \cdot R^2 + 4 \cdot \frac{R^2 \cdot \pi}{2}}{R^2 \cdot \pi} = 327\%$, i.e. the area of the new multi-hop cell is more than three times as large as the initial transmission area of the access point.

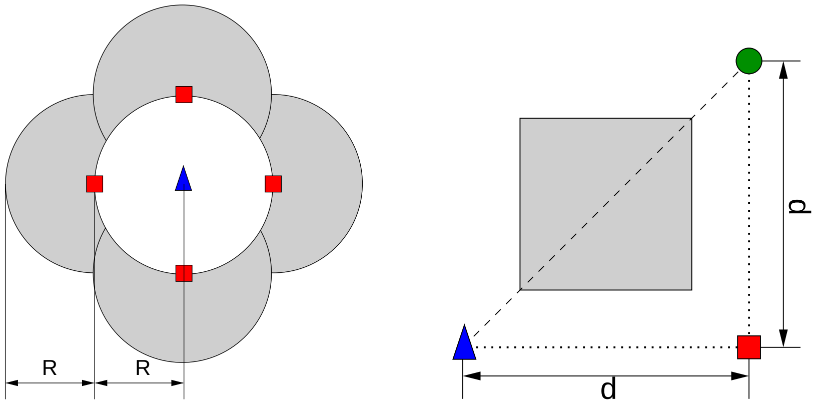
Summary To conclude, relay stations can extend the coverage range of an access point. However, the extended coverage area will typically result in a lower average capacity, as the cell size and thus the average transmission distance between access point and mobile terminal is increased.

2.3.2. Spectral Efficiency

Spectral efficiency is defined by the information in terms of bit, which is transmitted in a communication network using a predefined frequency spectrum as resource. A higher spectral efficiency means a better utilization of the frequency spectrum such that a higher overall data rate is achieved with a constant set of resources. A typical value for the spectral efficiency of user links is about 0.5 bit/s/Hz in a GSM network [FNO99]. The desired spectral efficiency of the future LTE standard is located in the range between 10 bit/s/Hz and 20 bit/s/Hz [Sch08]. It can be shown that multi-hop links are not only a possibility to extend coverage, but splitting a single-hop link into two or more hops can lead to a spectrally more efficient usage of the transmission resources from the network's perspective. This is especially the case for non line of sight (nLOS) connections as shown in [Pab04]. If we consider again a network operating in the 5 GHz frequency band, propagation conditions and thus link conditions become especially dependent on physical propagation effects. The electro-magnetic wave is even more viable to effects like fading or penetration losses due to physical

---

[5] Three symbols will be used in the following figures: Any access point (AP) or base station (BS) is symbolized by a blue triangle. Any relay station (RS) is represented by a red square and each mobile terminal (MT) is depicted by a green circle.



(a) The initial coverage area of the access point is increased by 327% through the use of four relay stations.

(b) Splitting a nLOS connection into two LOS connections can raise the spectral efficiency of a transmission.

Figure 2.2.: Relaying can help to increase coverage and spectral efficiency in a cellular network.

obstacles. The performance gains in terms of spectral efficiency achieved by relaying concepts are therefore higher then.

Example Let us illustrate this statement by a scenario as shown in Figure 2.2(b), where a mobile terminal can choose between a two-hop line of sight (LOS) connection and a single-hop nLOS connection to a given access point. We want to investigate this scenario from the spectral efficiency point of view. For that purpose, let us consider a simple free space propagation model to determine the received signal strength $P_r(d)$ as a function of the distance d between sender and transmitter for a given transmission power P_s

$$P_r(d) := P_s \left(\frac{c}{4\pi f} \right)^2 \cdot \frac{1}{d^\delta}, \tag{2.1}$$

whereby the transmission frequency $f = 5.3$ GHz, the speed of light $c = 3 \cdot 10^8$ m/s, a fixed overall transmission power $P_s = 200$ mW, and the path loss coefficient $\delta = 2$ for the LOS and $\delta = 4$ for the nLOS case are taken into account. A measurement campaign[6] has shown that this simple model is reasonably accurate for transmissions in the 5 GHz domain. We define the distance between the terminals to be $d = 30$ m. In order to allow a fair comparison, we split the transmission power for the nLOS connection into two equal parts of $P_s' = 100$ mW for the first hop and the second hop in the two-hop case, while the single-hop connection operates with $P_s = 200$ mW. A corresponding calculation reveals that $P_{r,mh}(30\text{ m}) = 2.25 \cdot 10^{-9}$ W for both hops of

---

[6] A corresponding measurement campaign has been performed by Nokia Siemens Networks GmbH, but results are not publicly available.

the multi-hop connection and $P_{r,sh}(42\text{ m}) = 1.25 \cdot 10^{-12}$ W for the single-hop case. According to Shannon's famous theorem, the resulting capacity of a link C is determined by

$$C := B \cdot \log_2 \left(1 + \frac{P_r}{P_n}\right), \tag{2.2}$$

where we assume a bandwidth $B = 10$ MHz and a noise power of $P_n = 10^{-12}$ W. In the two-hop case, the capacity of the overall link is determined by the minimum of all links. The link capacity is equal on both links in the given scenario. The application of Shannon's capacity theorem shows that the capacity boundary for the single-hop case is $C_{sh} = 11.7$ MBit/s. However, the multi-hop capacity is given by $C_{mh} = 111.3$ MBit/s. The huge performance difference is mainly due to the different propagation conditions. Nevertheless, this example clearly illustrates, how relaying may help to raise the spectral efficiency of a wireless network.

Summary It is important to record at this point that relaying can indeed increase the spectral efficiency of a link connection and thus increase the capacity of a cell. Performance gains are mostly achieved for spectrally inefficient links, which are divided into two or even more links with good link conditions. This effect becomes even more dominant with raising carrier frequency, as the risk for bad link constellations increases due to heavier propagation path losses.

2.3.3. Capital and Operational Expenditures

Any financial investment of a company into fixed assets with the aim of generating future profits is called Capital Expenditure (CAPEX). Concerning a mobile communication network, the technical infrastructure equipment itself, the deployment of this equipment, and the planning process of the network are related to CAPEX. When introducing a new technology, the necessary CAPEX for the start phase of the network is of enormous importance for the feasibility of a technological investment.
According to the analyst report in [Bro04], a GSM base station has been sold for about 70000 € around 2002. It can be purchased for around 10000 € or even less money today. Another source in [GLDS07] considers a WIMAX deployment scenario and assumes around 110000 € for purchase and allocation of a WIMAX base station. Similarly, [Sch08] estimates 90000 € for a 4x4 MIMO base station with 20 MHz bandwidth and additional 30000 € for the opening of a new site. It can be concluded from these data that for any kind of cellular infrastructure network, a new site with latest-edge technology equipment will roughly cost about 100000 €. Let us consider an investing provider, who wants to establish a new wireless communication network with sufficient coverage in the 5 GHz frequency band. Obviously, the deployment of such a network is practically infeasible concerning CAPEX from the economic point of view, if coverage entirely with base stations shall be achieved. As the coverage area is dependent on the square of the transmission range, the deployment costs to achieve coverage increase disproportionately with decreasing transmission range. As relay

stations are much cheaper than access points in terms of purchase costs as well as site development, relay stations help to cut down CAPEX significantly [Pab04]. In contrast to access points or base stations, relay stations do not require any wired fiber connection. A simple electricity supply to allow permanent power consumption is sufficient. The usage of relay stations thus accelerates the network roll out phase and makes it also cheaper.

Operational Expenditures (OPEX) include the ongoing costs related to the network's daily operation. Typical OPEX comprise the costs for maintenance and service of the infrastructure, the rent of the wired backbone connection, expenses for power, and last but not least rental costs for the site itself. According to [Sch08], a 200 MBit/s fiber line to connect a base station causes monthly leasing costs of 1625 €. [WIM04] assumes yearly OPEX of base station equipment to amount to about 5% of its CAPEX. It is clearly visible from these numbers that OPEX also play an important role in the economic business model of an operator. The deployment of relay stations helps to reduce operational costs, as — again — relay stations need not be wired by fiber to the operator's backbone. Imagine a scenario, where the network provider signs an agreement with a city government. The agreement allows the potential deployment of relay stations into public buildings and on street or traffic lights. An in-return allowance of the operator guarantees additional funding for the city and therefore makes it attractive. The two-party contract between a government and a provider will also alleviate the deployment in a city center in comparison to site-by-site rental deals with different contractors. The rental costs per site are presumably small for this model compared to today's site costs of a base station. Simultaneously, the service staff of the network provider could comparably quickly allocate a large number of relay stations almost everywhere. Some access points will placed additionally at locations, where fiber connections are already available. A first trial with IEEE 802.11 equipment has been accomplished in the Wireless Silicon Valley project, where 45 wireless transmission stations have been deployed on street lights and buildings in a test project [Fea07].

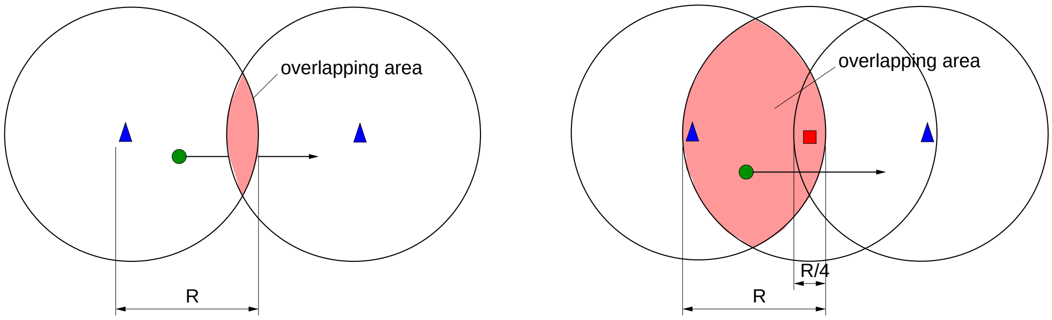
Summary To conclude, high CAPEX and OPEX result in high costs per bit for the network operator. These costs have to be transfered to the subscribers in any reliable business model. The attractiveness is therefore fairly low for both, the provider and the customer. Cellular multi-hop networks are a technical concept to overcome these economic hurdles by reducing CAPEX and OPEX. Costs are therefore kept at a considerably low level. This makes the multi-hop network's operation a beneficial business for the provider, especially during the initial grow-up phase of the network.

2.3.4. Seamless Handover in Networks with Small Cell Size

Mobile users in a cellular network typically handover from one cell to another cell. The frequency of these handovers depends on two issues, namely the cell diameter and the user velocity. The handover frequency is thereby defined as the expected average number of handovers between two cells per second. On the one hand, the faster the user moves around, the higher the resulting handover frequency will be. On the other hand, a small cell diameter also results in a high handover frequency. An analytical analysis in [Emm05a] has established a relationship between cell size and required overlapping coverage area in between neighboring base stations to perform a seamless handover. It is shown that the ratio between required overlapping coverage area and overall cell area increases with raising handover frequency. An extended study of [Emm05a] is published as technical report in [Emm05b]. While intuitively high velocities are related to a large number of handovers, the importance of the cell diameter is not so obvious on the first view. Especially indoor, office, and urban scenarios turn out to be related to high handover frequencies, surprisingly by far higher than high-speed highway scenarios. Although the absolute amount of the required overlapping area remains constant for a fixed user velocity, its relative fraction of the overall cell area increases with decreasing cell size. The relative area, which has to be covered by multiple access points just to enable seamless handover, related to the overall coverage area is therefore much larger for a network operating in the 5 GHz domain than for the typical GSM frequency range.

Example Let us consider a car passenger driving at a low speed of 10 m/s through a hot-spot cell with a radius of $R = 50$ m. Assuming a LOS channel and following the assumptions of [Emm05b], the solution of the corresponding analytical model results in a required minimum overlapping range of 46% of the transmission radius R. This means neighboring stations have to be placed in a mutual distance of $d = 77$ m, if a seamless handover shall be performed. Figure 2.3(a) illustrates the scenario for $R = 50$ m. Two base stations are located in a distance of 87.5 m. Although the transmission ranges are overlapping and thus coverage is achieved in a static scenario, the emerging density is not sufficient to enable a seamless handover for a user with a velocity of 10 m/s moving from one cell to the other cell. This is simply due to the fact that 87.5 m > 77 m and therefore the minimum overlapping range is not achieved. An additional relay station is placed in the middle between both base stations in Figure 2.3(b). In doing so, the overlapping coverage range is increased to more than 100% of the transmission radius. According to [Emm05b], the required minimum overlapping is exceeded by far such that a seamless handover between all stations is possible in this case.

Summary Summing up the consequences of a high carrier frequency, we can record two major phenomena: A high carrier frequency will result in a small cell size and therefore an increased number of transmission stations to achieve coverage in a way that every location within the transmission area has sufficient signal strength. In addition to that, the small cell size increases the handover frequency in comparison to conventional 2G or 3G networks. The analytical study in [Emm05b] has shown that the handover area between two cells takes a larger percentage of the overall cell area, if a seamless handover shall be performed. Consequently, a higher infrastructure density is required. The use of relay stations can increase the infrastructure density in wireless networks and thus increases the overlapping area between neighboring stations. For this reason, relay stations help to fulfill the coverage requirements for a seamless



(a) Network topology with a mutual overlapping range of 25% of the cell radius achieves coverage, but it is not sufficient to allow a seamless handover.

(b) The extended network topology with relay station results in a overlapping coverage range of more than 100% of the transmission radius such that a seamless handover is feasible according to [Emm05b].

Figure 2.3.: Illustration of the handover problem due to lack of infrastructure density.

handover as presented in [Emm05b].

2.3.5. Capacity of Cellular Multi-Hop Networks

Expansion State of a Wireless Network

The capacity of a cellular network is mainly dependent on two extrinsic parameters: the number of wireless infrastructure stations per area and the available bandwidth. An increasing number of access points or base stations will typically enhance the capacity of a network. This due to the fact that the propagation path loss is the determining factor for the achievable channel capacity. Provided a fixed amount of signal power, a higher infrastructure density reduces the average distance between transmitter and receiver. Consequently, the expected path loss is smaller such that the achievable link is increased. According to Shannon's law, the capacity C of a channel is limited by

$$C := \frac{B}{r} \log_2(1 + \gamma(r)). \tag{2.3}$$

In contrast to Equation (2.2), we assume this time that the bandwidth $B = 200$ MHz is divided by the reuse factor r into subbands, which are assigned as transmission resource to the base stations. Additionally, γ is not only subject to noise, but in a cellular network primarily related to interference from other base stations.

Bandwidth is generally an extrinsic input factor, which is provided in terms of licenses by government agencies. The overall amount of available bandwidth is thus an economic factor depending on the business model and beyond technical scope. Any increase of r will raise the signal to interference and noise ratio (SINR) of the transmission signals. However, it will also decrease the available bandwidth per station, as only a fixed overall bandwidth is available. The operator's technical challenge is to maximize the capacity, implicitly given by the fraction $\frac{\log_2(1+\gamma(r))}{r}$. Obviously, the optimum choice is determined by a compromise between frequency reuse r and resulting SINR $\gamma(r)$.

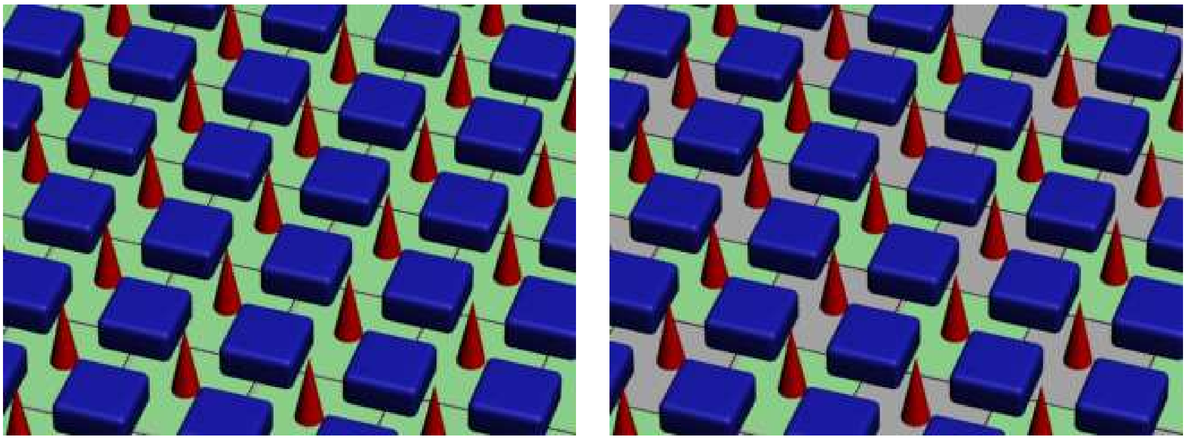
Example Let us consider a network deployed on a Manhattan grid topology with a building size of 75 m and a street width of 15 m. The expansion state e of a distinct network is introduced as the percentage of crossings, which are covered by an access point. Following this definition, the minimum expansion state e_{min} equals to $e_{min} = 0\%$ in the meaningless case, in which there is no access point deployed. In contrast, the maximum expansion state e_{max} amounts to $e_{max} = 100\%$, if exactly one access point is located at each cross-way. This implies that access points are only deployed at street crossings as proposed in [SWPI03]. Exemplary deployment scenarios for $e = 100\%$ and $e = 20\%$ are illustrated in Figure 2.4(a) and Figure 2.4(d). Again, signal propagation is determined by the free space formula according to Equation (2.1) with the same parameter set as in Section 2.3.2. A corresponding SINR is calculated in dependence on the interference power of surrounding stations P_i^k by

$$\gamma = \frac{P_r}{P_n + \sum_k P_i^k} \approx \frac{P_r}{\sum_k P_i^k} \tag{2.4}$$

under the assumption that $P_n \ll \sum_k P_{i,k}$. Four interfering stations in LOS and the eight nearest stations in nLOS are taken into account for the interference calculation. We want to determine the expected average capacity $\tilde{C}$ of a user link in such a network in dependence on the expansion state and the reuse factor. For that purpose, the coverage area A of a cell is divided into N grid elements Δa_n as exemplarily illustrated in Figure 2.5(a) for one cell in a network with expansion state $e = 0.2$. The coverage area is limited to the road level, users within buildings are not considered. Cross-road areas, which are covered from stations in both vertical LOS and horizontal LOS, are always supplied by the station in horizontal LOS by definition. The maximum cell range d_{max} is chosen to be half of the distance between neighboring stations in LOS and depends on the expansion state. The expected link capacity $C(\Delta a_n)$ is calculated according to Equation (2.3) for each grid element Δa_n. Subsequently, the averaged link capacity is divided by the cell area A to obtain an area-specific capacity indicator:

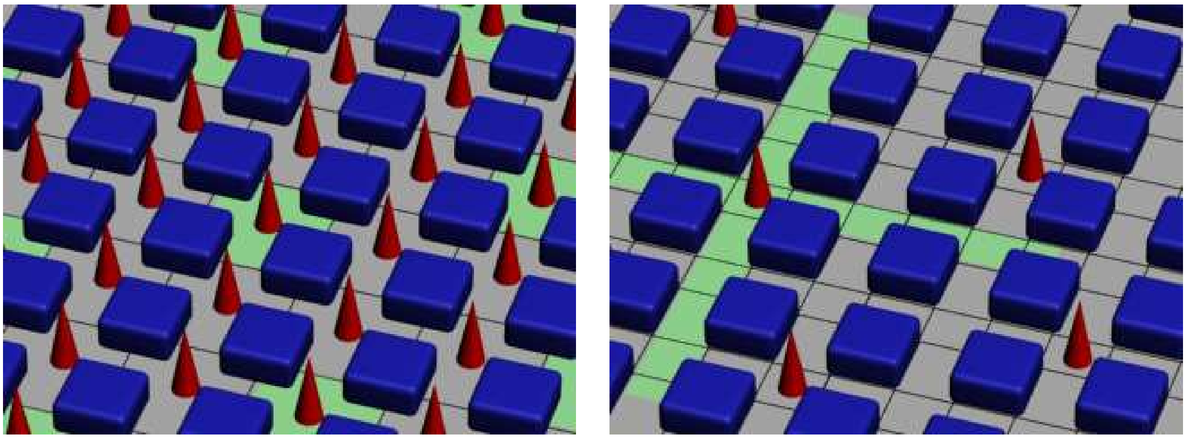
$$\tilde{C} = \frac{1}{A} \cdot \frac{1}{N} \cdot \sum_{\Delta a_n \in A} C(\Delta a_n). \tag{2.5}$$

The capacity has been evaluated for a set of reuse factors $r \in \{1; 2; 5\}$ and expansion states $e \in \{0.05; 0.1; 0.2; 0.5; 1.0\}$. The applied reuse patterns are shown for $e = 1.0$ in Figure 2.4(a) - 2.4(c). Figure 2.7(a) indicates that the overall capacity increases almost linearly in dependence on the expansion state. This obviously confirms that a higher infrastructure density has positive effects on the network capacity. Figure 2.7(a) also suggests that a small reuse factor $r = 1$ is favorable from the overall capacity perspective, even if users far away from the base station at the cell border are subject to very high interference then.



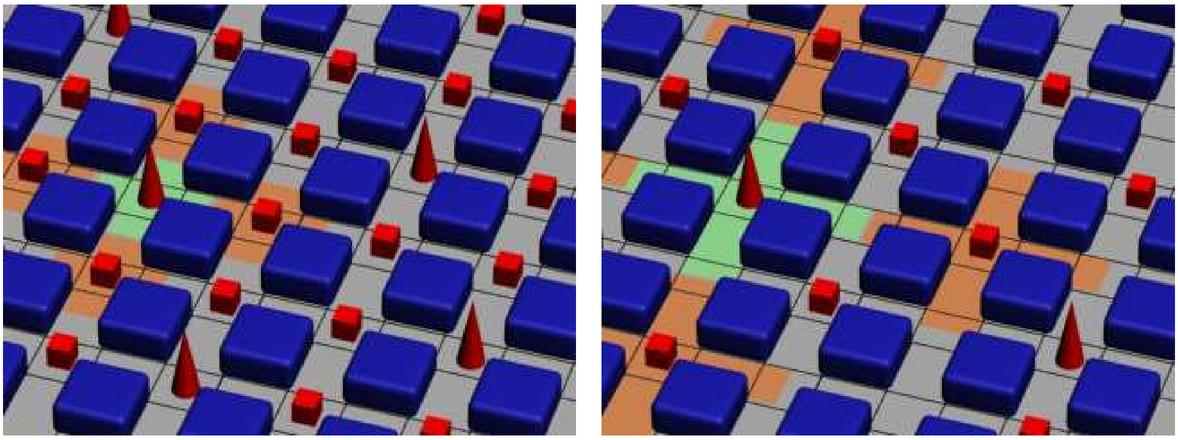
(a) Network with expansion state $e = 1$ and reuse factor $r = 1$.

(b) Network with expansion state $e = 1$ and reuse factor $r = 2$.



(c) Network with expansion state $e = 1$ and reuse factor $r = 5$.

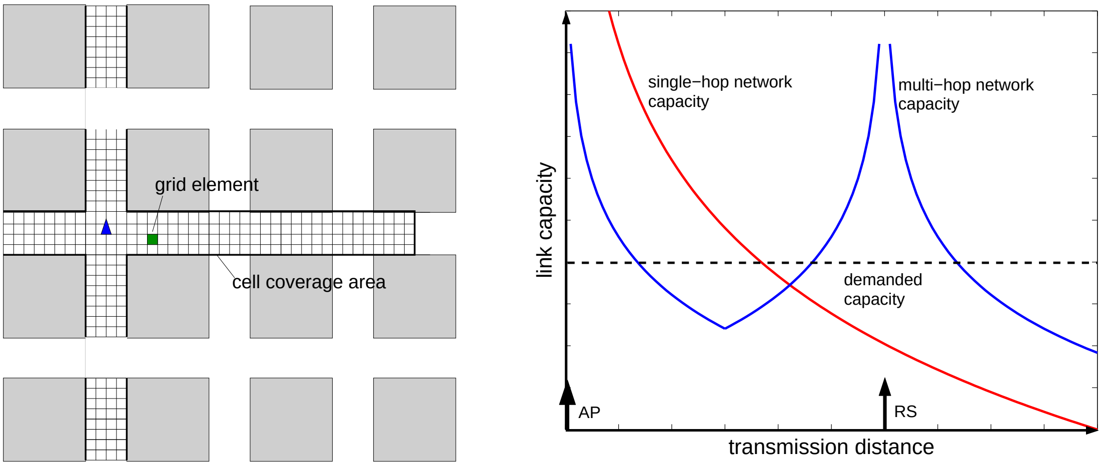
(d) Coverage area of an access point in a network with expansion state $e = 0.2$.



(e) Coverage area of an access point extended by four relay stations in a network with expansion state $e = 0.2$

(f) Coverage area of an access point extended by four relay stations in a network with expansion state $e = 0.1$.

Figure 2.4.: Deployment scenarios with different expansion states and reuse factors.



(a) The cell area is divided into multiple grid elements to determine the average capacity.

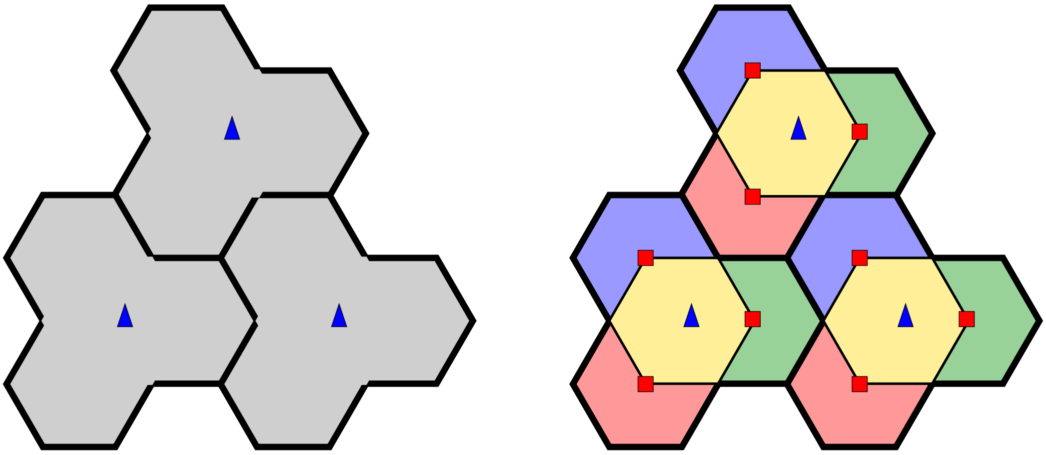
(b) The use of relay stations equalizes the natural link capacity throughout the cell.

Figure 2.5.: Division of a cell into grid elements and illustration of the natural link capacity within a cell.

Summary It is very important to emphasize at this point that the capacity of a wireless network can be increased — in theory — arbitrarily by raising the number of access points or base stations correspondingly. It makes only sense to compare networks with a predefined and fixed access point density for this reason. Whenever a relay station within the network is replaced by an access point, the capacity will increase. Therefore, any comparison between networks with unequal access point density is not meaningful. Additionally, one should keep in mind that small reuse factors tend to maximize the average capacity of the network.

Reuse Partitioning with Relays

Let us continue the frequency reuse discussion and especially address the role of relay stations within a network with small reuse factor. We have already seen that a reuse factor $r = 1$ is beneficial from the capacity perspective such that it becomes understandable that full reuse $r = 1$ is intended in LTE networks for reasons of spectral efficiency. Figure 2.6(a) shows an exemplary cellular network with reuse factor $r = 1$, the users of which will most likely face severe interference at the cell boundary, at least for full traffic load. Let us alternatively consider a cellular multi-hop network as illustrated in Figure 2.6(b). Suppose the available transmission resources to be separated into four subsets. Each of these subsets is assigned exclusively to the access point and the three relay stations. Interestingly, the network in Figure 2.6(b) uses the same average amount of resources per cell area as the network in Figure 2.6(a), but does avoid the interference problem at the cell boundary. The use of relays within the cell enables a radio resource management strategy, which is commonly referred to as reuse partitioning. Given this example, it becomes understandable that the use of relay stations facilitates reuse partitioning strategies, as the relay deployment naturally



(a) The figure shows a single-hop cellular network with full frequency reuse in all cells. Users at the cell borders will be subject to heavy interference from surrounding stations.

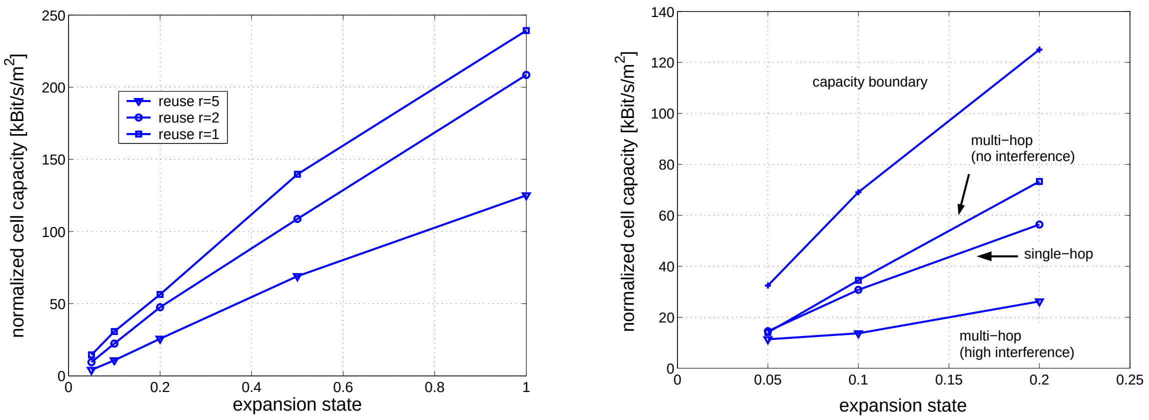
(b) The radio resource management divides the overall resources into four parts and assigns the resource subsets to different cell areas. Relay stations help to apply this strategy, which is commonly referred to as reuse partitioning.

Figure 2.6.: The figures illustrate a single-hop and a multi-hop cellular network. While both of them are equal in coverage and apply full frequency reuse in each cell, only the single-hop network suffers from severe interference at the cell border.

separates the cell into different areas.

The allocation of relay stations also helps to equalize the offered capacity over the cell area. The user's capacity demand is typically distributed over the whole cell area and therefore independent on the distance to the access point. However, due to propagation path losses, the natural link capacity of users far away from the access point is way smaller than the capacity of a user next to the access point as sketched in Figure 2.5(b). We obviously observe a mismatch between demand and offer. This mismatch can only be partly compensated by corresponding MAC or PHY mechanisms to guarantee a minimum data rate support. It turns out that the concept of relaying also leads to a more equal distribution of the natural link capacity as illustrated in Figure 2.5(b).

Summary We keep in mind that networks are generally intended to be operated at small reuse factors. This is also the case for the LTE networks such that interference problems and a general mismatch between demand and offer at the cell border are a major challenge. The use of relay stations facilitates reuse partitioning strategies, equalizes the capacity distribution throughout the cell area, and therefore helps to offer steady service quality to users independently on their position.



(a) The average capacity of a cellular infrastructure network increases with raising expansion state e and decreasing reuse factor.

(b) The average multi-hop capacity is increased in comparison to the single-hop case for good relay link conditions, but is decreased in case of strong interference.

Figure 2.7.: Illustration of the link capacity in a Manhattan grid topology.

Influence of Relay Link on the Network Capacity

We have used the preceding sections to explain, why cellular multi-hop networks are a technical way to combine infrastructure density and costs in a reasonable fashion. However, this cost advantage is associated with the disadvantage that additional resources are required for the transmission between the access point and relay station. This set of links is referred to as relay links in the following. The amount of necessary resources for the relay link determines, whether the relay-enhanced cell provides higher capacity than a corresponding single-hop cell with equal size. Interference is thereby a major issue. If relay links suffer from strong interference, the achievable physical data rate on the relay link is very small. Provided a fixed overall amount of resources, the amount of remaining resources for user links is decreased to the extend, to which the resources for the relay links are employed. The interesting question is now, whether the relay-enhanced cell will indeed dispose of higher capacity. Alternatively, the additional bandwidth requirement could even degrade the overall capacity compared to the single-hop case. It now depends on the efficiency of the relay link itself, whether the average overall capacity of the multi-hop cell is increased or even decreased by the use of relay stations. Let us explain this issue again with an example.

Example We consider a single-hop infrastructure network with reuse factor $r = 1$ in a Manhattan topology as reference for that purpose. Starting from this given infrastructure network, four relay stations are assigned to each access point. Relay stations are placed northwards, southwards, eastwards, and westwards of the corresponding access point on the grid. This infrastructure upgrade is performed exemplary for three basic network topologies with $e = 20\%$, $e = 10\%$, and $e = 5\%$. Two exemplary deployments for basic topologies with $e = 20\%$ and $e = 10\%$ are shown in Figure 2.4(e) and Figure 2.4(f). We want to compare the average cell capacity for the single-hop and the multi-hop case and therefore evaluate the capacity according to Equation (2.5). The evaluation results in Figure 2.7(b) indicate that the overall capacity of the multi-hop network even decreases in comparison to the single-hop network. It is thereby assumed that the multi-hop cell is distributed in five equal areas and the resources are divided correspondingly. The analysis reveals that heavy mutual interference is a major problem on the Manhattan topology, as the relay stations at the cross-roads are subject to strong interference from surrounding stations. In a second step, we imagine the relay links to be free of any interference and to only suffer from natural path loss according to the free space propagation. Although this assumption seems to be rather unrealistic on the first view, a combination of beamforming and smart scheduling can create constellations, in which this assumption is reasonably justified. The user links are not concerned by the new assumption and treated unalteredly. By neglecting interference on the relay links, the capacity is increased significantly and outperforms the single-hop benchmark in Figure 2.7(b). Replacing all relay stations by access points, a scenario is created, which defines a capacity boundary for the multi-hop network. The higher infrastructure density theoretically allows the reduction of the reuse distance. Nevertheless, the upper boundary has been calculated for an unchanged reuse distance, as otherwise a fair comparison between the relay-extended cellular network and the single-hop network is infeasible. The capacity margin between the single-hop capacity and the upper boundary visualizes the potential of relay-based networks. The unfavorable capacity result for the high interference case illustrates possible dangers of the relay approach and emphasizes the tremendous need for very efficient relay link constellations.

Summary Summarizing the above section, the importance of the relay link within the cellular multi-hop network has been pointed out. The relay link is clearly the linchpin of the entire cellular multi-hop network. If the relay link conditions are good, we observe capacity gains throughout the network. Otherwise, the capacity can be even reduced in comparison to the single-hop case.

2.4. Architecture of a Cellular Multi-Hop Network

2.4.1. Cellular Multi-Hop as an Extension of LTE System Architecture

The intended cellular multi-hop network architecture is an extension of the current Long-Term Evolution (LTE) system architecture according to the specification of the 3GPP[7] standardization working group [3GP08a]. While initially discussed within LTE standardization groups, the use of relays has been currently removed from the agenda. The only active working group dealing with the standardization of relay devices is IEEE 802.16j relay task group. However, it is still foreseen to include relay devices in future developments of the LTE standard. The road map is about 4-8 years

---

[7] Third Generation Partnership Project (3GPP)

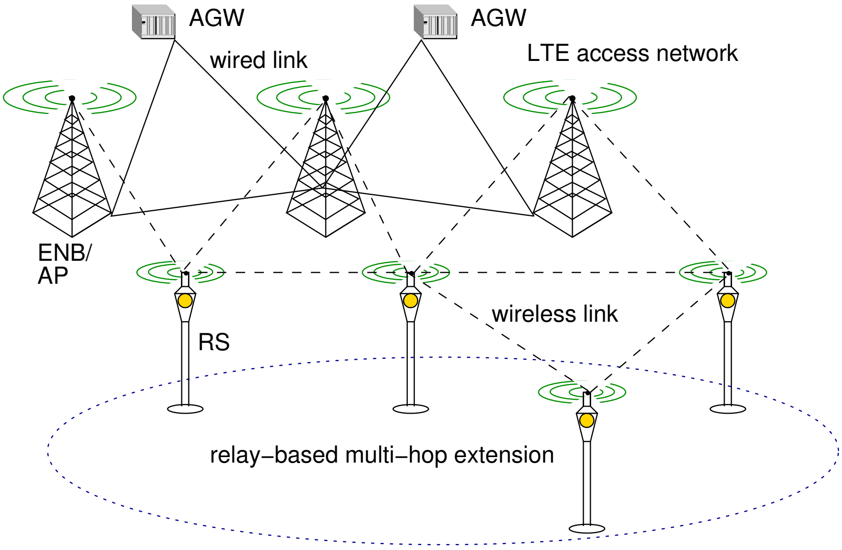


Figure 2.8.: The LTE system architecture with a relay-based multi-hop extension.

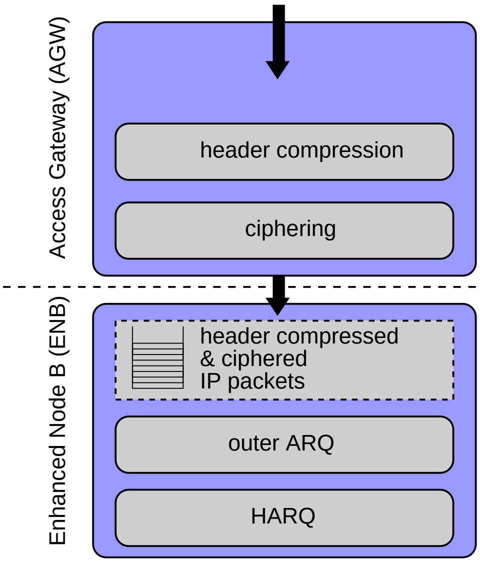


Figure 2.9.: Data processing in AGW and ENB.

from now on, i.e. 2012-2016. The system architecture of LTE as shown in Figure 2.8 includes two major components, namely access gateways (AGW) and enhanced nodeB (ENB) devices. According to [EFK$^+$06], the AGW node summarizes functionalities of the Gateway General Packet Radio Service (GPRS) Support Node (GGSN), the Serving GPRS Support Node (SGSN), and the Radio Network Controller (RNC) as known from the UMTS system architecture. The redesign of the UMTS protocol stack with an aggregation of these nodes into a single node leads to reduced call setup times and additionally saves development and purchase costs [EFK$^+$06]. The ENB is in charge of the connection mobility control and the radio resource management [Bar06]. Therefore, the handover of a mobile terminal is accomplished by the mutual interaction between two ENBs (see [BHKV07] and [RTR07]), which cause the AGW to redirect the current data traffic from the source ENB to the destination ENB. As shown in Figure 2.9, the AGW sends out header compressed and ciphered user data. This data is already prioritized according to predefined service classes. While processing the queued packets, the ENB accounts for the service level. Lossless delivery is achieved by an outer n-process stop-and-wait Automatic Repeat Request (ARQ) loop and a Hybrid ARQ (HARQ) with subsequent PHY/ARQ interaction [Bar06].

The relay station in a cellular multi-hop network is intended to form a wireless extension of the current LTE architecture as sketched in Figure 2.8. The relay station is supposed to act as an extension to the ENB and to forward the data in a wireless multi-hop connection to the user terminal. The multi-hop connection requires an enhancement of the outer ARQ loop into a two-level scheme with an outer and several inner ARQ loops as proposed in [WMLC05] or similarly in [Lot05]. The outer ARQ loop is established between user terminal and ENB, while the inner ARQ loops monitor the transmission hop-by-hop. In doing so, service quality is provided even for multi-hop connections.

2.4.2. System Assumptions

As it is planned to enhance the LTE architecture in future by relay stations, one important constraint of this thesis is the compliance to the LTE standard. Otherwise, there is no realistic chance to apply the proposed ideas. The cellular multi-hop concept may suggest some extensions and minor changes to LTE components, but shall not concern the basic system architecture. In compliance with LTE, this work is subject to the following system assumptions:

TDD-OFDMA The wireless devices are assumed to operate in half-duplex mode. This means that devices cannot receive and transmit at the same time. Considering relay stations as an extension of the cellular infrastructure, it is clear that the radio resources of relay stations are primarily scheduled by the access point. The access point in turn provides subsets of the frequency band to the relay stations. The same set of transmission resources is thereby intended for reuse in all cells ($r = 1$). The overall frequency band is typically divided in a periodic frame structure for that purpose. The relay station may run its own scheduler within the distinct resource subset as

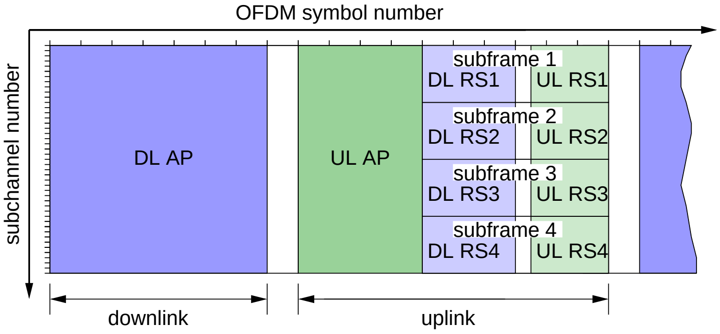


Figure 2.10.: The picture shows a possible OFDMA frame structure as used in cellular multi-hop networks. Relay stations are assigned subframes in the uplink frame of the access point (UL AP). Relay stations may use these resources to schedule an independent subframe structure with downlink phase (DL RS) and uplink phase (UL RS).

sketched in Figure 2.10. This requires the idea of a hierarchical radio resource management scheme. An exemplary realization of a hierarchical radio resource management for OFDMA systems in the single-hop case can be found in [RRE06]. The LTE specification defines OFDMA for the downlink transmission from the access point to the mobile terminal and foresees Single Carrier Frequency Division Multiple Access (SC-FDMA) for the uplink. As the physical layer is not focused in this work, we will rely on a simplified FD-TDMA realization to model the OFDMA access scheme. OFDMA systems at 5 GHz have been investigated by Michael Einhaus[8] in [ESW07]. His concepts concerning PHY and MAC complement the presented approaches in this thesis.

Single Interface Technology In contrast to some existing work, we assume that all wireless devices must operate on a single interface. According to some manufacturers, this is a major constraint for the economic saleability of a technology. It is very important to differentiate the cellular multi-hop approach from existing approaches based on microwave technology at this point. Some operators use microwave radio to mutually connect base stations by wireless links. Microwave connections require very reliable LOS links such that all base stations are mounted on exposed sites. The exposed site and the very rigid channel requirements characterize the microwave concept, in which the microwave technology is a substitute for a wired fiber connection. Instead, the relay-based approach considers relay stations to be subordinated to one distinct access point. Relay stations are placed on the street level. The resulting links are supposed to be good, but not reliable enough to operate a microwave connection. The laxer channel requirements and the subsequent higher flexibility in deployment are major advantages of the relay concept. The single interface is also reflected in significantly lower manufacturing costs. Therefore, both approaches are not directly comparable, as the deployment requirements are different. However, please keep in

---

[8] Michael Einhaus and the author have collaborated within the Wireless Gigabit With Advanced Multimedia Support (WIGWAM) project funded by the German Federal Ministry of Education and Research.

mind that a relay station can never be an adequate replacement of a base station. The "in-band" relay link transmission will always occupy parts of the overall set of transmission resources. Provided the expose site, the "out-band" microwave approach automatically outperforms the relay concept from the capacity point of view. From this perspective, the relay concept is always economically motivated and differs in the deployment requirements from the microwave approach.

Relay Device Beyond the LTE constraints, there are some assumptions to the relay device itself. We require the relay station to dispose of layer 3 functionality. This means that any relay device is able to receive, to decode, and finally to digitally process an incoming data packet. The processing includes for instance the identification of header structures and addresses. Devices with this type of capability are typically referred to as decode-and-forward devices [BL07]. This is a major difference to amplify-and-forward devices, also known as repeaters. Repeaters can be used to extend a base station's coverage range and only accomplish non-intelligent retransmission [RSW06]. This implies a restriction to simple layer 2 functionality. Advanced multi-antenna transmission technologies are supposed to be available for the communication among wireless routers. Consequently, relay stations are equipped with more than a single transmission antenna. The resulting antenna gain is a key factor to guarantee the efficiency of the relay link. In contrast, there are no antenna requirements to mobile terminals such that there is no antenna gain assumed for the communication toward the mobile terminals. Any interaction within the network has to be initiated by a wireless router, as mobile terminals are not assigned transmission resources. This typically corresponds to the rules of cellular networks. We specifically focus the 5 GHz carrier frequency range. This choice has a major impact on the size of cells. The principles of our work are not dependent on the frequency. However, it will be shown that the cellular multi-hop concept becomes even more valuable and relevant for high carrier frequencies.

2.5. Methodology

Three major methods have been used in this thesis to investigate cellular multi-hop networks: graph theory, numerical optimization, and simulation. Let us give a short overview about these methods, before we start in detail with the analysis.

2.5.1. Graph Theory

A Graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is defined by a set of nodes $V_i \in \mathcal{V}$ and a set of edges $\vec{e}_i \in \mathcal{E}$ to mutually interconnect the nodes. Graph theory is a mathematical method to explain the properties of such graphs. In doing so, graph theory is a helpful tool to abstractly describe, to analyze, and finally to solve problems in communication theory. Wireless devices can be modeled by nodes, while edges describe link relationships between different devices. Thereby, any edge $\vec{e}_i$ typically represents a noise-free channel, which connects a single transmitter node with a corresponding receiver node. All network

Table 2.1.: Frequently used graph-related notations.

description	notation
node i	V_i
set of nodes	$\mathcal{V}$
cardinality of the set of nodes $\mathcal{V}$	$\|\mathcal{V}\|$
directed edge from V_1 to V_2	(V_1, V_2)
path from V_1 to V_3 via V_2	$\{(V_1, V_2), (V_2, V3)\}$
edge i	$\vec{e}_i$
set of edges	$\mathcal{E}$
set of outgoing edges of node V_i	$\mathcal{E}^{(out)}(V_i)$
set of incoming edges of node V_i	$\mathcal{E}^{(in)}(V_i)$
set of outgoing edges of the set of nodes $\mathcal{V}$	$\mathcal{E}^{(out)}(\mathcal{V})$
set of incoming edges of the set of nodes $\mathcal{V}$	$\mathcal{E}^{(in)}(\mathcal{V})$
edge capacity of edge i	$c(\vec{e}_i)$
allocated resources on edge i	$g(\vec{e}_i)$

operations are subject to a finite field of transmission symbols, i.e. the signal vectors at both the transmitter and receiver nodes are elements of the same finite field symbol alphabet. Siddharth Ray refers to such a channel as "finite field adder channel" [RMA03]. A path within the network is thereby defined as a set of concatenated edges, where the head of one edge is tail of a subsequent edge.

According to the definitions in [Die97], the graph model of the wireless network is characterized by several properties. All edges are directed such that each edge has a tail symbolizing the transmitter and a head representing the receiver. Each directed edge of the graph has an inverted correspondent edge. This underlies the assumption that the wireless channel is symmetric. Due to the symmetry, the graph naturally contains cycles. The wireless network is supposed to be connected at any time, i.e. a path between any pair of the wireless terminals is always feasible. The resulting graph is therefore connected.

2.5.2. Numerical Optimization

Since a few years, modern computer technology allows the solution of complex numerical optimization problems. In the following, we will take a closer look on linear optimization problems as a special class of optimization problems. A linear problem is defined by a set of constraints $\mathbf{A}\vec{x} = \vec{b}$, where $\mathbf{A}$ is a (m x n) matrix and $\vec{b}$ a vector, both with constant coefficients. The vector $\vec{x}$ represents the set of optimization variables. In dependence on these variables, we refer to the problem as linear optimization problem (LP) for a continuous set of variables x_i, as integer linear problem (ILP) for a set of discrete integer variables x_i, and finally to a mixed integer linear problem (MILP) for a set of partly discrete and partly integer variables. An associated optimization objective is formulated by a linear function $\Omega = \vec{c}^{\,T} \cdot \vec{x}$ with again constant coefficients c_i. The overall problem statement therefore matches the following form:

$$\begin{array}{ll} \textit{optimize} & \Omega = \vec{c}^{\,T} \cdot \vec{x} \\ \textit{subject to} & \mathbf{A}\vec{x} = \vec{b}. \end{array}$$

The set of linear equations spans a convex polyeder. Any linear function is a convex function. Each local optimum represents also a global optimum, whenever the admissible region is convex, as stated by the Kuhn-Tucker conditions [KT51].

Simplex Algorithm Linear problems are typically solved by the simplex approach. The simplex approach is based on the idea to determine a start solution and iteratively improve this solution. We assume that the matrix of the linear problem is characterized by rank$(\mathbf{A}) = m \leq n$. Consequently, there is a (m x m) nonsingular submatrix $\mathbf{A}_B$ and a (m x n-m) matrix $\mathbf{A}_n$ such that $\mathbf{A}\vec{x} = \vec{b}$ can be rewritten as $\mathbf{A}_B\vec{x}_b + \mathbf{A}_n\vec{x}_n = \vec{b}$. A corresponding solution to $\mathbf{A}\vec{x} = \vec{b}$ is therefore given by $\vec{x}_B = \mathbf{A}_B^{-1}\vec{b}$ and $\vec{x}_n = \vec{0}$. $\mathbf{A}_B$ is referred to as basis and (x_b, x_n) as basic solution. The idea of the simplex algorithm consists in the initial determination and a subsequent reformulation of the basis, until the optimum basis is found. Changes to the basis can be accomplished by a pivot transformation. Provided that the objective is calculated by $\vec{c}^{\,T} \cdot \vec{x} = \vec{c}_b^{\,T} \cdot \vec{x}_b + \vec{c}_n^{\,T} \cdot \vec{x}_n$, it can be shown that the basis is optimum, if $\vec{c}_b\mathbf{A}_b^{-1}\mathbf{A}_N \geq \vec{c}_n$. For further details please refer to [WN99] or related literature.

Branch-And-Bound Algorithm The simplex approach is not sufficient to solve large problems due to the increasing complexity with raising number of variables. In this case, the simplex approach can be combined with more sophisticated strategies to split the original problem into feasible subproblems. The Branch-and-Bound algorithm as proposed in [LD60] is one possible approach for these problems. The idea of the Branch-and-Bound algorithm is to identify solvable subproblems and to concentrate on these subproblems first. This proceeding is referred to as branch step and results in a recursive, tree-like problem structure. The second part is to calculate lower and upper bounds for distinct vertices of the tree. If the lower bound of some tree vertex surpasses the upper bound for another vertex, then the initial vertex needs not be considered in the further calculation. It is impossible that this vertex contributes to the optimum solution in this case and therefore can be omitted. The Branch-and-Bound algorithm is especially useful for the solution of a MILP. Given the fact that an optimization problem with discrete solution variables is bounded by the same problem with continuous solution space, the relaxation to a continuous solution is a promising technique to step-by-step approach the discrete solution. Intelligent strategies are characterized by the property that large parts of the original solution space are segregated early by the branch and subsequent bound step. If applied successfully, only a small part of the original space has to be considered. Numerical solver software packages use this property and apply intelligent selection heuristics in order to even solve complex problems successfully.

2.5.3. Simulation

Simulation is a method to verify theoretical expectations with practical results. Simulation can also be used as a proof of principle to illustrate the operation and efficiency of a proposed algorithm in a network with realistic protocol stack. In this work, two different simulation methods are used:

- *graph-oriented approach*: In order to understand the technical challenges and problems, graph-based models of the wireless network are developed. These models are implemented with GRAPH [Gru07] and ILOG CPLEX [ILO07]. The graph-oriented approach primarily serves as indicator of potential performance margins.
- *protocol-oriented approach*: In order to illustrate the feasibility of a distinct approach, a realistic protocol stack is implemented in the event-based network simulator ns-2 [Fal00]. The proof of principle of the proposed solution approach is the primary objective of this type of simulation. Apart, the protocol simulation is suitable to characterize the trade-off between signaling efforts and subsequent performance gains.

Graph-Oriented Approach Graph theory has been presented in Section 2.5.1 as a tool to model wireless communication networks. We use the proprietary C++ library GRAPH to implement the graph representation of the cellular networks. GRAPH allows the object-oriented modeling of multi-layer graph structures and offers a C++ interface to the numerical solver software CPLEX. ILOG CPLEX is able to cope with a MILP as described in Section 2.5.2. The exact solution strategy is a corporate secret. However, it is known that the numerical solver is based on the principle of the above described Branch-and-Bound algorithm. Provided the convergence of the optimization algorithm, CPLEX guarantees the optimality of the solution for any linear problem statement. We use the combination of these two software libraries to calculate optimum transport strategies in cellular networks.

Protocol-Oriented Approach The basis for later protocol simulations is the event-based network simulator, version 2.27, (ns-2.27) as available at [Fal00]. The network simulator is embedded in a simulation framework. The simulation process as sketched in Figure 2.11 works as follows: The software tool ScenarioGenerator with graphical user interface has been developed to create static scenarios and building descriptions [Sei05]. This input information can be processed by CityProp 2.0 to calculate the connectivity and to add mobility as described in Appendix A.1 to the static initial scenario. CityProp 2.0 is an extended version of the software tool CityProp as proposed in [GL04b]. The protocol stack of the wireless devices is created by a Tool Command Language (TCL) configuration script, which links ns-2 specific C++ modules dynamically. The simulation flow is steered by this configuration script. Accordingly, the TCL script also feeds the topology and mobility information from the scenario files into the simulator. Finally, an adjusted statistic module keeps record of the simulation process and traces concentrated simulation data. A visualization of the simulation process is feasible a posteriori by the AdHocNam tool [Zöl03]. The evaluation of all simulations includes a 10 s start-up phase, which is not accounted in the evaluation results. In the following, plotted confidence intervals illustrate the 95% confidence level. The confidence intervals underlie the assumption that the sample results are distributed according to the student distribution.

A detailed description of the simulation tool is provided in Appendix A. At this point, no more than a rough overview about the different components is given. The simulation environment itself results from four models as sketched in Figure 2.12(b): a channel model based on the Friis free space propagation formula [Fri46], a traffic model with a constant bit rate traffic generator, a mobility model according to the random direction mobility model as proposed in [RMSM01], and finally the Manhattan grid topology as obstacle model. The node itself is composed in a layered structure according to the OSI protocol stack as shown in Figure 2.12(a). Starting from the PHY layer bottom-up, any transmitted packet is received correctly, when a minimum signal power and a minimum SINR requirement are surpassed. An optimum link adaptation process without feedback and packet loss is accomplished by adjusting a packet's physical transmission data rate to the current received signal strength. The ns-2 MAC module has been replaced by a generic component referred to as ZERO-MAC, which models an optimum and collision-free, but capacity-restricted medium access. The working principle of ZERO-MAC is very similar to a conventional Time Division Multiple Access (TDMA) MAC. A radio resource management unit to distribute transmission resources among all wireless devices is included, similarly as the Address Resolution Protocol (ARP), which is responsible for the mapping between layer 2 MAC and layer 3 IP addresses. The network layer establishes Internet Protocol (IP) connectivity and therefore disposes of the IPv4 protocol stack, the mobility extension Mobile IPv4, and a supplementary routing and clustering unit. This unit is focused within the following chapters, as it takes control about the logical connectivity within the network. Finally, a User Datagram Protocol (UDP) client is located on top of the IP layer and fed with input data by the traffic generator as virtual application.

The overall simulation model is not sufficient to get a detailed understanding of physical layer issues, especially concerning inter-channel and intra-channel interference, correlation, and fading effects. This means that it is not possible to indicate symbol-based delays and chunk-based scheduling schemes. The given model is detailed enough to understand the characteristics of routing and handover issues, to determine signaling overhead, and to estimate the required average resource consumption. It represents a reasonable trade-off between accuracy and complexity. Such a compromise is typically necessary to make system level simulations feasible.

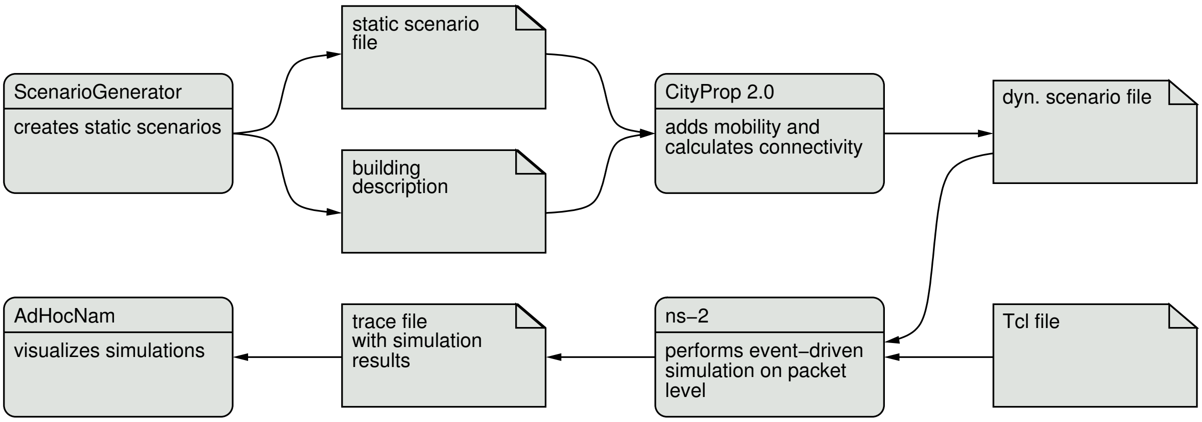
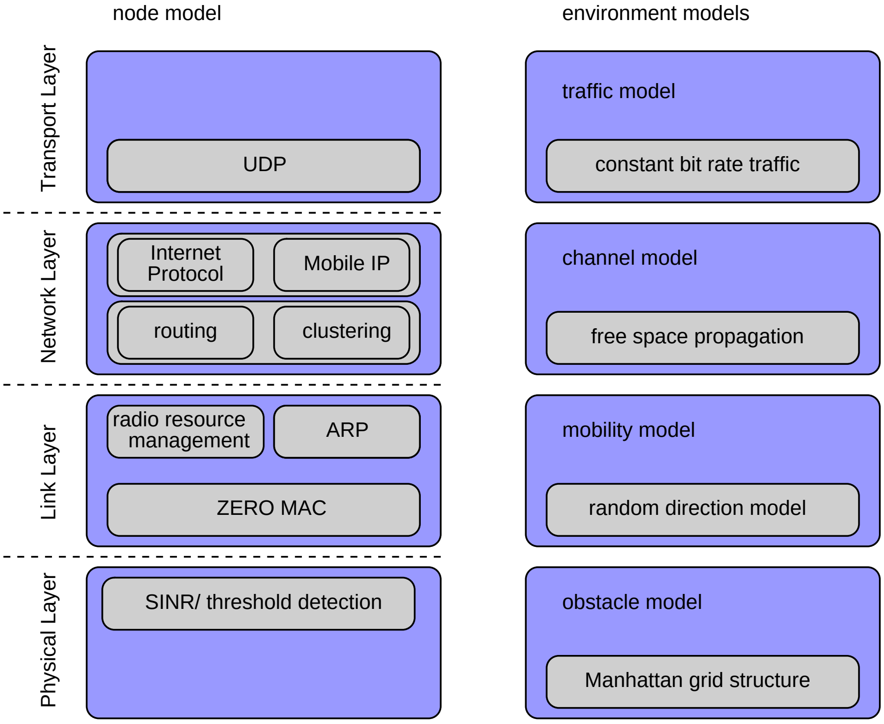


Figure 2.11.: The overall simulation flow around the network simulator ns-2.



(a) Protocol stack of node model.

(b) Environment models for simulation.

Figure 2.12.: Layered node implementation of the wireless devices and extrinsic models, which influence the nodes' interaction.

3. Dynamic Cell Clustering in Cellular Multi-Hop Networks

3.1. Introduction

Imagine a wireless infrastructure network to be given by a set of access points and relay stations. In order to operate such a network as a cellular network, a method to cluster access points and relay stations into multi-hop cells is required. This chapter describes possible algorithms, whose basic idea is the dynamic accomplishment of this clustering process in dependence on the current load situation. In doing so, the clustering process fulfills two tasks: The clustering process assigns a relay station to a corresponding access point. This process is also referred to as gateway discovery in the literature. Additionally, the clustering process enables the balancing of traffic load among different access points and is therefore a way to extend the capacity of the multi-hop network.

The methodology is as follows: At first, related approaches from literature and state-of-the-art techniques to solve this combined load balancing and gateway discovery problem are presented. After that, a system model is created to describe the problem mathematically. The subsequent section compares a centralized and a distributed solution to the problem. The later analysis also involves topology-related issues and examines potential performance gains. Finally, a practical realization in a protocol stack is presented and evaluated by simulation.

3.2. Motivation

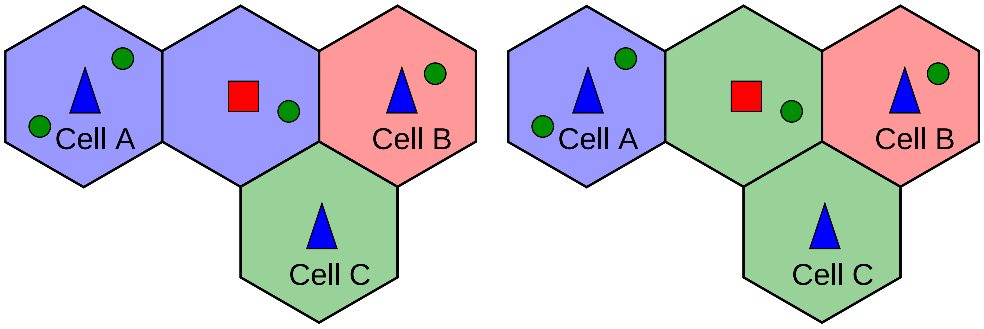
Relay stations within a cellular multi-hop network are dependent on distinct access points, as only access points dispose of wired connections to the backbone network. The assignment of relay stations to access points can be accomplished in a manual one-time gateway assignment at the beginning of the network operation. This approach is easy, but inflexible. Therefore, literature proposes dynamic gateway discovery mechanisms for wireless access networks with access points and wireless ad hoc devices. The challenge of gateway discovery consists in the exploration of mutual link conditions and the subsequent establishment of a routing path from a relay station to a distinct access point. The aggregation of wireless nodes around a privileged node is also referred to as clustering process. Clustering typically results in the formation of a local subnetwork. If relay stations are assigned to access points in a cellular network, the created logical unit is a multi-hop cell.

Cellular networks are also characterized by the fact that different cells are often subject to unequal traffic loads. This is related to the natural issue that mobile users are not uniformly distributed on the coverage area in practice and have a time-dependent traffic demand. Congestion situations may appear at locations of major public interest, for instance sport events or festivals, but also conferences or airports. Typically, a set of transmission resources is assigned to each cell. Due to the unbalanced traffic load, resources may not be sufficient to support the traffic load in one cell, while neighboring cells may still have left over remaining resources. A mechanism, which balances the user load per cell in a way that all cells have about equal load, will extend the network's capacity. Generally, load balancing is known as a technique to distribute load among different components in a system such that all parts achieve the same load level. In the following, load balancing is defined as a operation strategy to distribute a given network-wide traffic load on different multi-hop cells about equally.

We propose Dynamic Cell Clustering (DCC) as a joint solution to the gateway assignment and load balancing task. The idea is to accomplish a dynamic assignment of relay stations to access points. The assignment depends on the load situation and can be changed during the network's operation. As access point and relay stations form a common multi-hop cell together, this assignment procedure represents a cell clustering mechanism. The coverage range of a cell will change after each reassignment. Any change in the coverage range of a cell generally also applies to the user load of this cell. If the assignment is performed in an intelligent way, the average load per cell can be equalized by this mechanism. As previously mentioned, this equalization process is also known as load balancing. Dynamic Cell Clustering is therefore a technical approach, which offers a combined solution to the gateway assignment and load balancing task in a cellular multi-hop network.

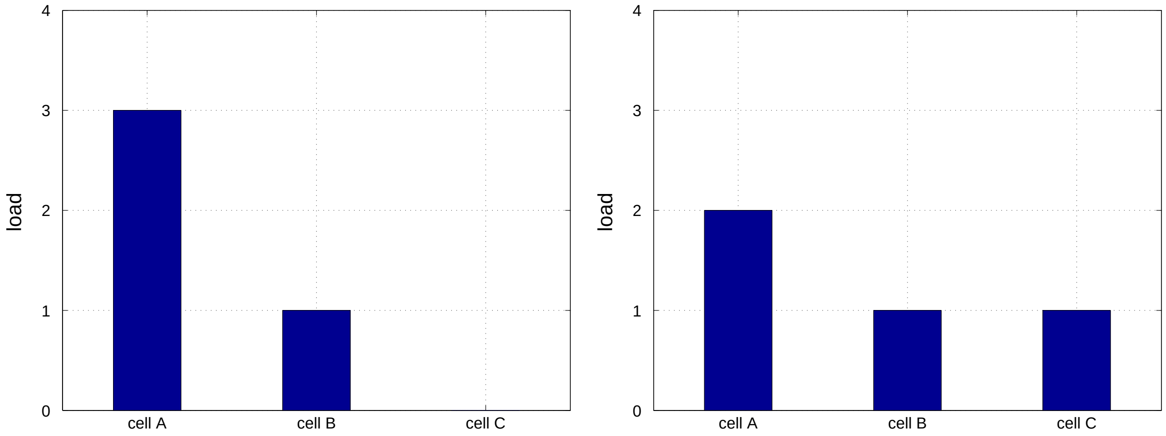
Let us consider a network with three cells A, B, and C as shown in Figure 3.1(a). As illustrated, there is a relay station attached to the access point of cell A, while the infrastructure in cell B and cell C consists of no more than access points. It is also visible that there are three mobile terminals within the coverage range of cell A. However, there is only a single terminal in cell B and no terminal at all in cell C. This means that cell A has a by far higher traffic load than cell C as sketched in Figure 3.1(c). It is obvious that a reassignment of the relay station from cell A to cell C will reduce the coverage range of cell A and extend the coverage range of cell C. Consequently, the maximum traffic load per cell will be decreased and the traffic load per cell be equalized according to Figure 3.1(d). The final cellular structure after the reassignment is depicted in Figure 3.1(b). The change in the coverage range is also observable in this figure.

Dynamic Cell Clustering is not a substitute for some radio resource management entity, but a complementary tool in addition to conventional radio resource management schemes. Dynamic Cell Clustering is a process, which we could imagine to be performed on an hourly basis, possibly also within minutes, but probably not on a second scale or even beyond that. In doing so, Dynamic Cell Clustering can be considered as one component of a radio resource management tool kit, which acts on a slower time scale than the conventional scheduler of the radio resource management unit. Alternatively, Dynamic Cell Clustering can be interpreted as a handover procedure of a relay station from one cell to another. Whenever a relay station is reassigned



(a) Initial cell configuration with relay-extended cell A.

(b) Cell configuration after reclustering with relay-extended cell C.



(c) Overload situation at cell A.

(d) Improved load distribution after reclustering.

Figure 3.1.: Example network configuration before and after the reclustering process: load is transfered from cell A to cell C.

from one cell to another cell, it is assumed that the relay station has to keep its attached mobile terminals. Therefore, the proposal can be considered as a handover approach in a hierarchically organized network. The handover is motivated by load imbalances and takes place on an intermediate level of the network hierarchy, i.e. the relay station level. The proposed Dynamic Cell Clustering mechanism is beneficial in multiple regard. We will present these advantages in the subsequent sections.

Extension of Network Capacity It is known that access points run into danger to become possible bottlenecks during the network's operation. Bottlenecks may be created in the fixed part of the network as well as on the air interface. While in former times mainly the air interface has been considered as capacity limitation, this might change in future high data rate mobile communication networks. Assuming wireless transmission capacities in the gigabyte domain, the fixed line connection — today mostly a rented DSL line — may indeed become a bottleneck for the network's operation. This is not due to technical reasons, but due to economic issues, as the rental fee is rate-dependent and costly. It is consequently very expensive in terms of operational costs to dimension network connections according to some possible peak load.

Balancing the traffic among different cells is therefore an economically reasonable and necessary issue to extend the capacity of future mobile networks. The overall capacity of the network is raised in two cases, namely

- *for a wireless bottleneck* — if the wireless transmission resources of one cell are not sufficient to support all the traffic within a cell. Relay stations allocate and receive resource subsets from their corresponding access points. By assigning a relay station to another access point, the resource demand at the original access point is decreased.
- *for a wired bottleneck* — if the fixed line connection of a single access point is not sufficient to carry the local peak traffic. Equalizing the traffic load among different cells automatically balances the traffic among different backbone connections.

For this reason, Dynamic Cell Clustering can help to extend the capacity of a cellular multi-hop network. This is especially the case for unforeseen and temporary local peak loads due to some external event. The traffic will be balanced by distributing the traffic to surrounding cells then.

Deployment and Administration It has been already mentioned that CAPEX and OPEX per wireless infrastructure router are key parameters for the economically successful operation of a network. Dynamic Cell Clustering can reduce OPEX significantly. This is especially important for a large number of infrastructure stations. Assuming for instance an inner-city coverage area with a radius of 2 km and a transmission range of a station of about 100 m, at least 400 stations are required to achieve coverage. Technically, one GSM base station is sufficient to cover the 2 km radius, although — for capacity reasons — the number of base stations is often higher. The problem is that such a cellular multi-hop network cannot be planned and maintained the same way as, e.g., a current UMTS or GSM network, where the number of stations is by far less. Relay stations and access points shall be mounted and maintained by service personnel without deeper technical expertise. It is very obvious that a sophisticated cell planning process, as it is currently accomplished for GSM sites, is probably unrealistic in case of a 5 GHz network with a tremendous number of access points and relay stations. As already proposed, we assume a potential contract between a cell phone network provider and the local public services company or city government. In doing so, the city agrees that the network provider is allowed to place relay stations for instance on traffic lights or street lighting. This in turn guarantees a convenient and permanent access to electricity for the transmission devices and a large amount of potential sites with a single two-party contract. Accordingly, the service staff of the network provider is able to allocate comparably quickly a large number of relay stations practically anywhere. Additionally, some access points are deployed at places, where fiber connections are easily available. Nevertheless, the deployment process of relay stations shall be reduced to a simple installation and plug-in procedure of electricity. This means that it is desirable to just have a self-configuring cell topology with autonomously acting relay stations. After being switched on, a relay station is supposed to start an auto-configuration mode in order to get an assignment to an existing cell of the network. This mechanism could be part of the Dynamic Cell Clustering scheme. Therefore, Dynamic Cell Clustering does not only facilitate the deployment of new networks, but also makes the administration and extension of these networks easier.

Maintenance and Service It is already common practice in the operation of today's networks to switch off GSM and UMTS base stations from time to time for maintenance purpose. Typically, the surrounding base stations are reconfigured to take over the coverage range of the maintained station then. Given the large amount of stations, it is not clear, what to do with the attached relay stations in the maintenance case of an access point in a cellular multi-hop network. The easiest — but probably not the best — option is to simply switch off all of them. A more advanced, but also more complex and laborious proceeding is realized by the manual reconfiguration of surrounding stations in case of maintenance. The same solution applies in case of a failure of one component and the corresponding reassignment of stations into the regular network service, after the failure is fixed again. Due to the large amount of stations, the risk of a failure of one device is by far higher than in today's networks. Dynamic Cell Clustering is a possible solution to the question, how to react on and to ease these maintenance and failure scenarios. The clustering mechanism could help to alleviate manual reconfiguration processes, as additional intelligence is added to the network in order to enable automatic reconfiguration. To put the motivation in short, Dynamic Cell Clustering shall be developed to enable the construction of cost-efficient, resource-efficient, and flexible networks.

Open Issues of Dynamic Cell Clustering

It is not apparent, whether and how much Dynamic Cell Clustering will improve the capacity of a specific cellular multi-hop network topology. There is a set of issues to be investigated and answered for that purpose:

- *How does the network topology influence Dynamic Cell Clustering?* The inherent ability of a network to balance its load depends on topology aspects like the access point density and the mesh degree. We want to examine the relationship between infrastructure topology and the options of the clustering algorithm to accomplish load balancing. The results give advice for the greenfield planning of such networks.

- *How to determine and resolve an overload situation efficiently?* A suitable measure to characterize a load situation is necessary for that purpose. Additionally, appropriate criteria, how to start and execute the reconfiguration process have to be defined.

- *What will happen in case of dynamics in the network's traffic load behavior?* Network dynamic is an important factor. Given the fact that the load changes very quickly within the network, the load distribution is a time-dependent quantity. For that purpose, the success of the balancing approach depends heavily on its ability to also deal with dynamic load changes.

- *How can these tasks be accomplished in terms of a protocol stack?* Getting a theoretical solution to an idealized model of a process is one issue. However, it is a different challenge to provide evidence that the proposed solution can be realized in a real environment. The acquisition of characteristic parameters, the embedding of corresponding components into the standard protocol stack, as well as the reasonable exchange of protocol messages is thereby the central issue.

We want address these questions in the following and start with a review of related approaches at first.

3.3. Related Work

Dynamic Cell Clustering is a new approach, which combines strategies from cellular networks with strategies from multi-hop networks. A literature survey has identified these two areas as key components of Dynamic Cell Clustering:

- *load balancing* as a strategy to increase capacity in cellular networks and
- *clustering* as a strategy to increase the efficiency of data transport in multi-hop networks.

Dynamic Cell Clustering is a joint strategy, which introduces the variable assignment of relay stations to access points in order to equalize inhomogeneous load distributions throughout cellular networks. Figure 3.2 shows that it is a concept, which combines various elements of existing approaches.

3.3.1. Load Balancing in Cellular Networks

Load Balancing by Adjusting Coverage

Beamforming The first option to balance user load is to artificially either extend or decrease the cell coverage range and thus to increase or to reduce the number of subscribers within the transmission range. Several methods to artificially adjust the cell size and coverage area are known from today's cellular networks. An adaptation of the cell coverage area can be achieved by mechanically turning antenna angles [Bun99] and electronically by changing or focusing the antenna beam [Har02]. [SVY06] proposes an automated scheme to optimize the coverage of the cell for 3G networks. Assuming site locations like traffic lights and low-cost transmission devices, beamforming and antenna tilting are most likely not applicable for router-to-terminal connections in cellular multi-hop networks. Nevertheless, electronic beamforming could be a reasonable approach to optimize the connectivity of the relay-based infrastructure network itself, for instance according to [SVY06], and thus become a valuable supplement of Dynamic Cell Clustering.

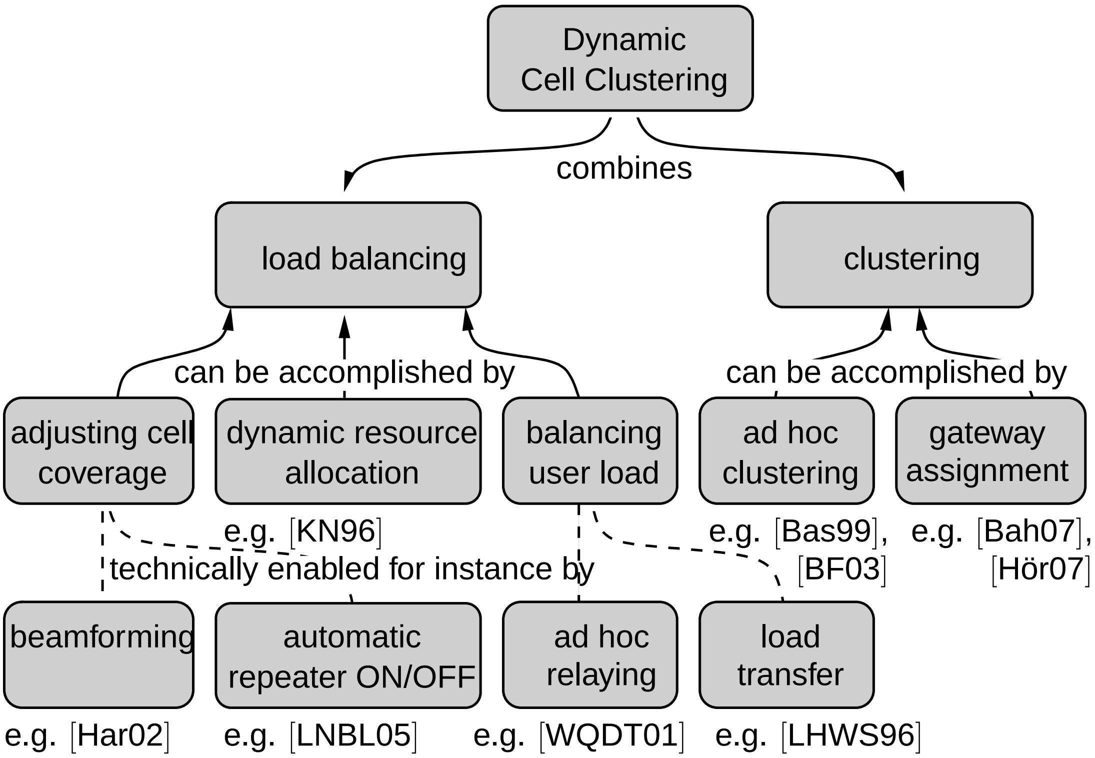


Figure 3.2.: Visualization of the Dynamic Cell Clustering strategy as a combination of load balancing and clustering in a cellular multi-hop network.

Repeaters Repeaters can be used to extend a base station's coverage range by amplifying incoming signals [RSW06]. Repeaters are also able to increase the capacity of CDMA networks in hot-spot areas ([RE04] or [CYCWB01]). A very interesting proposal in [LNBL05] suggests an "automatic on−off switching" of repeaters in dependence on the traffic load situation. Although this approach does not intend any load balancing between different stations, this proposal emphasizes another aspect, which is also important in Dynamic Cell Clustering: According to [LNBL05], the cell area and capacity of a single, relay-extended multi-hop cell can be adjusted dynamically to the load by turning repeaters automatically on and off.

Load Balancing by Dynamic Resource Allocation

Dynamic resource allocation is the trial to adjust resources at any base station to the network's need and thus to equalize load differences at different base stations. [KN96] provides an overview about different channel allocation schemes. It is generally distinguished between Fixed Channel Allocation (FCA) and Dynamic Channel Allocation (DCA). Typically, FCA works well for networks with low traffic situation. On the contrary, DCA performs efficiently for systems with moderate up to high load, as the inflexible assignment of channels for FCA raises the blocking probability. The future LTE architecture intends a full frequency reuse in each cell such that classical

FCA and DCA mechanisms are not applicable anymore. However, the LTE standard proposes mutual communication between adjacent base stations over the so-called x2 interface. This interface can be used to coordinate reuse partitioning strategies among neighboring stations to prevent interference problems at the cell border and is therefore also suitable for the signaling of Dynamic Cell Clustering. Instead of shifting resources between different cells, the Dynamic Cell Clustering approach is adjusting the cell borders and consequently balancing the traffic between cells. The amount of available resources per cell stays constant as foreseen by the LTE standard.

Dynamic User Load Balancing

Balancing by Ad Hoc Relaying [WQDT01] introduces a proposal, where mobile terminals based on IEEE 802.11 technology are used as ad hoc relays to increase coverage and capacity of a cellular network. This proposal is extended in [YTM$^+$02] and [YT04] in order to balance load between base stations of difference cells. Similarly, [SLKS03] presents a load balancing approach for heterogeneous networks with interworking between different networks. The idea to transfer load by relaying from one cell to another is also part of the Dynamic Cell Clustering mechanism. Fundamental differences between [WQDT01] and Dynamic Cell Clustering are found in two aspects: By using fixed instead of ad hoc relays, the Dynamic Cell Clustering proposal considers the load transfer as an integral, permanent, and reliable service component of the ongoing network activity. Furthermore, access point and relay station operate with a single interface technology such that our proposal can be classified as load transfer in a homogeneous network environment, while the scenario of [YT04] defines a heterogeneous network environment.

Balancing by Load Migration [LHWS96] investigates load balancing strategies in computing clusters, in which neighboring processors are mutually interconnected like nodes in a network. [LHWS96] compares the results of locally operating heuristics with the results of a central scheduler. The central scheduler is given by a queuing system with a predefined number of service units. A central scheduler may thus achieve some global load balancing goal as shown for instance in [NH85]. [LHWS96] additionally presents a heuristic, in which a preassigned load distribution, the physical connectivity between neighboring processors, and the local load transfer between neighbors are key elements. The problem formulation of [LHWS96] is therefore very similar to the balancing task in a cellular multi-hop network. Interestingly, it is shown that simple, local load balancing strategies, according to which tasks are just shifted to a neighboring processor in case of overload, can indeed perform good global balancing results. An important observation is the dependence of the overall performance on the connectivity of the processor network and on the variance of the load distribution.

3.3.2. Clustering in Wireless Multi-Hop Networks

Ad Hoc Networks The main purpose of clustering in ad hoc networks is to reduce signaling overhead by determining a speaker or so-called clusterhead of groups of geographically concatenated nodes. One of the early contributions is the simple Distributed Mobility-Adaptive Clustering (DMAC) for ad hoc networks, which relies on two messages types JOIN to join a cluster and CLUSTERHEAD to announce a clusterhead [Bas99]. The statistic properties of DMAC have been examined in [BF03] and [Bet04a]. The efficiency of a clustering algorithm is characterized by cluster stability or life time and membership time of a node within a cluster. The performance of clustering algorithms is limited by the node density within the network. The interesting issue is the simplicity and efficiency of this approach. We can keep this in mind for the development of a corresponding protocol for cellular multi-hop networks. However, different from ad hoc networks, the cellular topology does not foresee equality among the nodes, as access points and relay stations are local clusterheads by definition. A numerical optimization of the cluster number and connectivity given a predefined metric and using beamforming has been proposed in [HKV08]. Again, such beamforming approaches could also optimize the connectivity of the wireless infrastructure network instead of optimizing the number of clusters.

Mesh Networks Gateway discovery in mesh networks is a problem very similar to the relay attachment to access points in cellular multi-hop networks. Terminology distinguishes between so-called mesh points (= fixed relay stations, which do not provide access to mobile terminals), mesh access points (= fixed relay stations with access point functionality), and mesh portal points (= access points with wired backbone connection). IEEE 802.11s [IEE06a] [HMZ$^+$07] proposes a Hybrid Wireless Mesh Protocol (HWMP), which consists of two parts, namely a proactive component, as explicitly chosen mesh portal points advertise periodically their location to proactively build up routing trees [Bah06], and a reactive component, which is a reactive routing protocol. It is very similar to AODV [Bah07] and derived from the Cisco[1] proprietary Adaptive Wireless Path Protocol [Hör07]. The interesting aspect about this approach is the tree-like structure, which is built from the mesh portals to other nodes and the consideration of the physical data rate as routing metric in the Adaptive Wireless Path Protocol.

3.3.3. Summary

We have shown that Dynamic Cell Clustering is an operational strategy for cellular multi-hop networks, which incorporates ideas as the variable adjustment of cell ranges and the adaptation of user loads from existing proposals. Furthermore, we have explained, why the shifting of transmission resources between cells is difficult in future network architectures, and illustrated that Dynamic Cell Clustering will also

---

[1] Cisco Systems, Inc.

help to deal with this issue. Additionally, we have seen that the idea to cluster networks dynamically and to build up tree structures around access points is related to approaches from ad hoc and mesh networks. In the following, we want to examine expected performance gains of Dynamic Cell Clustering in dependence on the network topology and therefore start with the definition of an adequate graph-theoretical model to describe cellular multi-hop networks.

3.4. A Conceptional Study of Feasibility and Impacts of Dynamic Cell Clustering

3.4.1. System Model

Network Topology

Graph Let us develop a model to illustrate the impacts of the network topology on potential clustering strategies for cellular multi-hop networks. For that purpose, assume a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, c)$ of a network topology with nodes $V \in \mathcal{V}$, bidirectional links symbolized by a set of two oncoming directed edges $\vec{e} \in \mathcal{E}$[2], and corresponding edge capacities $c(\vec{e})$. The overall set of graph nodes $\mathcal{V}$ is composed by a set of access points $\mathcal{V}_A \subseteq \mathcal{V}$ and relay stations $\mathcal{V}_R \subseteq \mathcal{V}$ such that $\mathcal{V}_A \cup \mathcal{V}_R = \mathcal{V}$. A multi-hop cell $\mathcal{C}_n = (\mathcal{V}_n, \mathcal{E}_n)$ is defined as a subgraph $\mathcal{C}_n \subseteq \mathcal{G}$. We assume that the maximum number of cells N within the network equals to the number of access points

$$N := |\mathcal{V}_A|, \tag{3.1}$$

as each access point defines its own cell. The set of nodes $\mathcal{V}_n$ belonging to the same cell $\mathcal{C}_n$ is a joint set of a single access point $V_A \in \mathcal{V}_A$ and several attached relay stations $V_R \in \mathcal{V}_R$. The set of cell edges $\mathcal{E}_n$ defines an inverse tree structure from all attached relay stations to the access point as common root node. The clustering process of the overall network into a set of cells $\{\mathcal{C}_1, \ldots, \mathcal{C}_N\}$ is defined by the subgraph $\mathcal{S} = (\mathcal{V}, \mathcal{E}_s, g)$:

$$\mathcal{S} = \bigcup_{n=1}^{N} \mathcal{C}_n \subseteq \mathcal{G} \tag{3.2}$$

and

$$\mathcal{E}_s = \bigcup_{n=1}^{N} \mathcal{E}_n \subseteq \mathcal{E}. \tag{3.3}$$

A resource allocation $g(\vec{e})$ can be accomplished for each edge $\vec{e} \in \mathcal{E}$ to enable the transport of data. Naturally, the feasible resource allocation is bounded by the capac-

---

[2] The use of directed edges alleviates the mathematical examination. In contrast, we decided to rely on undirected edges in the graphical illustration of this chapter to avoid confusion due to an enormous number of edges in the figures.

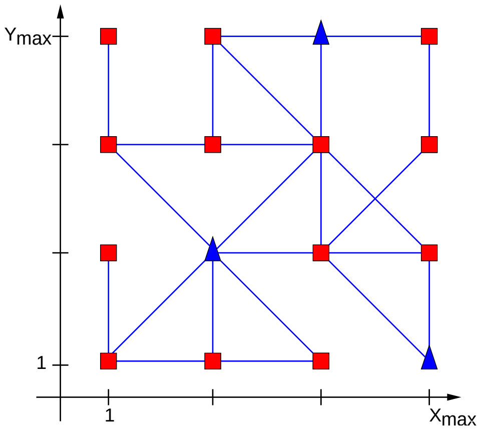


Figure 3.3.: Randomly meshed network topology with 16 nodes, mesh degree $\gamma = 3$, and $\rho = \frac{3}{16}$.

ity constraints of the overall network graph $\mathcal{G}$,
$\forall \vec{e} \in \mathcal{E}$:

$$0 \leq g(\vec{e}) \leq c(\vec{e}). \tag{3.4}$$

We presume that any edge $\vec{e} \in \mathcal{E}_s$ is equipped with resources $g(\vec{e}) > 0$, as an edge with $g(\vec{e}) = 0$ can not contribute to the solution of the clustering problem. Thus, $\mathcal{E}_s \subseteq \mathcal{E} : g(e) > 0 \ \forall \vec{e} \in \mathcal{E}_s$. While the initial graph $\mathcal{G}$ is connected, the subgraph $\mathcal{S}$ consists of N partitions.

Mesh Degree Access points and relay stations are placed randomly on a plain Manhattan grid structure such that each node has eight neighbors, i.e. two horizontal, two vertical, and four diagonal neighboring nodes as illustrated in Figure 3.3. The node positions are defined by a set of coordinates (X, Y) with $X \in \{1, \ldots, X_{max}\}$ and $Y \in \{1, \ldots, Y_{max}\}$. The mesh degree γ of a network is defined by the following expression

$$\gamma := \frac{|\mathcal{E}|}{|\mathcal{V}|} \tag{3.5}$$

as the quotient of the number of edges and the number of all nodes. For that purpose, bidirectional links described by a pair of edges are added at a random position in the network graph, until the network reaches the predefined mesh degree. The maximum mesh degree is constrained by the number of neighbors to $\gamma_{max} = 8$.

Access Point Density The network topology is created randomly with the probability of a node to become access point being uniformly distributed. The network graph is therefore given by a two-dimensional simulation grid structure with random mesh links and random access point positions. There is also a lower bound for the minimum mesh degree. This bound results from the requirement that all relay stations have to be connected to at least one access point. This lower bound is given by

$$\gamma_{min} = \frac{|\mathcal{E}_{min}|}{|\mathcal{V}|} = \frac{2 \cdot |\mathcal{V}_R|}{|\mathcal{V}_A| + |\mathcal{V}_R|} = 2 \cdot (1 - \rho), \tag{3.6}$$

as the minimum number of edges is two per relay station. Consequently, ρ defines the access point density according to

$$\rho := \frac{|\mathcal{V}_A|}{|\mathcal{V}_A| + |\mathcal{V}_R|}, \tag{3.7}$$

as the ratio of access points to all stations.
Imagine a large network with many relay stations and only a single access point. The access point density $\rho \rightarrow 0$, while the minimum mesh degree $\gamma_{min} \rightarrow 2$ in this case. This is explained by the fact that relay stations need at least one link, i.e. one outgoing and one incoming edge, to connect to another station and thus to have an opportunity to get access to an access point. For $\rho = 1$, there is no relay station at all. As access points build up their own cells, this means that no additional wireless link is required such that the minimum mesh degree $\gamma_{min} = 0$ then.

Torus-shaped Simulation Grid Considering the described plain field scenario, there is one major drawback, as access points and relay stations at the border of the simulation grid area face less neighbors than nodes in the middle of the simulation field. This phenomenon that simulation conditions are location-dependent is also known as border effect. To avoid any border effects, the two-dimensional network is extended to a boundless three-dimensional network structure by folding the plain simulation grid into a torus-like structure. This is achieved by a transformation of the plain field coordinates (X, Y) into a three-dimensional Cartesian system of coordinates (X', Y', Z'). For that purpose, the angles θ and ψ are defined as $\theta := \frac{X}{X_{max}} \cdot 2\pi$ and $\psi := \frac{Y}{Y_{max}} \cdot 2\pi$ based on the original plain field coordinates. Afterwards, the following transformation is applied:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} \Big(K_1 + K_2 \cdot \cos(\theta)\Big) \cdot \cos(\psi) \\ \Big(K_1 + K_2 \cdot \cos(\theta)\Big) \cdot \sin(\psi) \\ K_2 \cdot \sin(\theta) \end{pmatrix}, \tag{3.8}$$

whereby K_1 and K_2 are constant factors to define the overall radius of the torus itself and the radius of the torus tube, respectively. An example topology, which will be used as default configuration, is shown in Figure 3.4.

Traffic Model

Wireless infrastructure networks are deployed to provide service for mobile users, which connect to the network in an unpredictable manner. The resulting traffic load at the wireless router is therefore variable and represents the aggregated traffic of connected users. As we consider only the infrastructure network at this point, we refer to the envelope of user traffic at a distinct station when talking about traffic in the following.

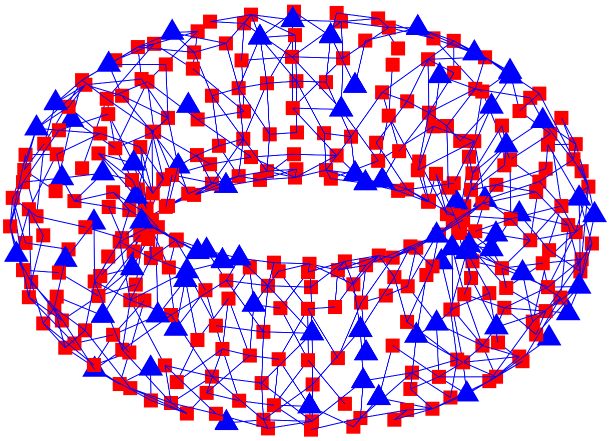


Figure 3.4.: Torus-shaped example network topology with 360 nodes ($X_{max} = 10$, $Y_{max} = 36$), mesh degree $\gamma = 4$, and access point density $\rho = 0.2$.

Initial Traffic For that purpose, all access points and relay stations are assigned traffic load values $\tau(V)$ to represent the aggregated random user traffic. Let $\tau(V)$ be a uniformly distributed random variable with $\tau(V) \in [0; \tau_{max}]$. Subsequently, all traffic values of access point and relay stations within a certain multi-hop cell $\mathcal{C}_n$ sum up to the cell load parameter λ_n:
$\forall \mathcal{C}_n = (\mathcal{V}_n, \mathcal{E}_n) \in \mathcal{G}$:

$$\lambda_n := \sum_{V \in \mathcal{V}_n} \tau(V). \tag{3.9}$$

Time-discrete Suite of Traffic Values Realistic communication networks are characterized by the property that the traffic configuration is not static, but dynamic. Traffic changes are caused by the mobility of users and time-dependent service requests with limited duration. To map these dynamics, the traffic model is extended to a time-discrete function $\tau(V, t)$, which provides traffic values in fixed time intervals Δt for a corresponding station V. $\tau(V, t)$ is calculated by a recursive relationship from the initial random values. Let us therefore introduce another uniformly distributed random variable $\eta(t) \in [-\eta_{max}; \eta_{max}]$ with $0 \leq \eta_{max} \leq 1$. $\eta(t)$ defines the percentage of maximum traffic t_{max}, according to which the traffic load may change after an interval Δt. Assuming a static network configuration, η_{max} is chosen to be $\eta_{max} = 0$ such that

$\tau(V, t + \Delta t)$=$\tau(t) + \eta(t) \cdot \tau_{max}$=$\tau(V, t)$ in the static case. Whenever the new traffic value $\tau(V, t + \Delta t)$ surpasses the maximum value τ_{max} or underpasses 0, the result will be bounded as well by these extreme values:

$$\tau(V, t + \Delta t) := \begin{cases} 0 & \text{for } \tau(V, t) + \eta(t) \cdot \tau_{max} \leq 0 \\ \tau_{max} & \text{for } \tau(V, t) + \eta(t) \cdot \tau_{max} \geq \tau_{max} \\ \tau(V, t) + \eta(t) \cdot \tau_{max} & \text{else.} \end{cases} \tag{3.10}$$

This model corresponds to the idea that there are periodic measurement reports in a real network, which update the network load status. The sequence of this measurement reports is usually not a memoryless value chain, but a correlated sequence of measurement values. This is due to the fact that the traffic does not originate from a single user, but represents an aggregated traffic sum of several user loads. This sum is typically less dynamic than a corresponding single user's load. The change of network load within a small time scale is therefore unlikely to be uncorrelated. The degree of correlation is expressed by the choice of η_{max}. For values close to zero, the chain of traffic values is highly correlated, while the mutual correlation decreases for larger values.

Frequency of Load Change Following the idea that the inertness of the network is driven by these periodic measurement reports, the duration of Δt depends on network characteristics. In case of load changing only slowly, Δt will be a large value. In case of load changing very fast, Δt will be considerably small. Let us take a look at Figure 3.5, which shows the load configurations of two exemplary cells A and B within a network with multiple cells. We can observe the periodic change of load $\lambda_A(t)$ and $\lambda_B(t)$, which results from corresponding measurements reports after Δt time units. We assume the current load profile to be evaluated by some balance indicator, which acts as a network-wide indicator for load inequalities at different cells. A high indicator value expresses equality among all loads, while a small value characterizes a network with high imbalances. The load changes are periodic in time, but unpredictable in amplitude. As a consequence of resulting imbalances, the balance indicator takes a hit every Δt time units. Suppose that there is an algorithm, which performs subsequent reassignments of relay stations from one cell to another cell in order to increase the balance indicator. For instance after $t = (2k+1) \cdot \Delta t_{switch}$ time units, some change in the load profile of cell A and B is observed. The increase of load at cell A equals to the decrease of load in cell B. This means that a reconfiguration of cell A and B has been accomplished. The reconfiguration yields an increase of the balance indicator.

Imagine the overall time interval to determine a local overload situation, to start the reclustering process, and finally to reroute a relay station to another cell to be constant and independent on the network dynamics. Let us assume this time interval to be given by Δt_{switch} as shown as well in Figure 3.5. If we suppose that reconfigurations have to be executed sequentially, one after another, the inequation

$$\Delta t \geq k \cdot \Delta t_{switch} \tag{3.11}$$

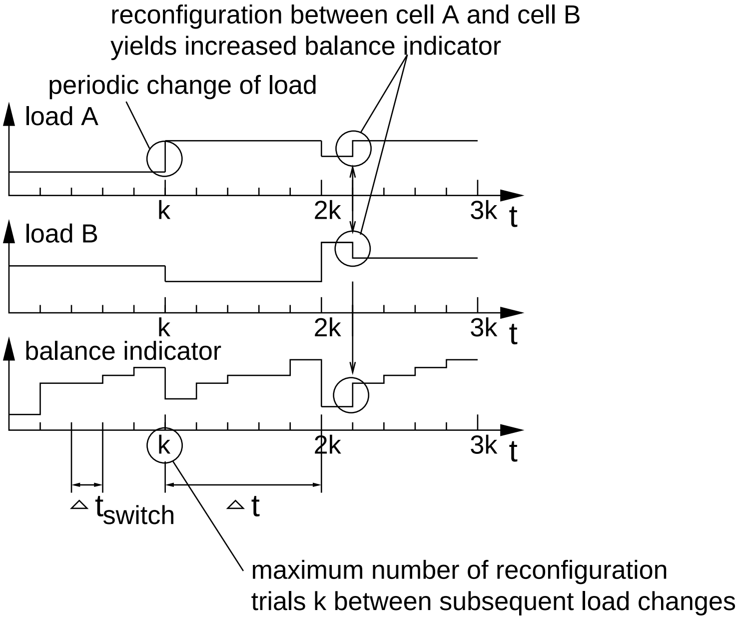


Figure 3.5.: Sketch of the influencing parameters for load balancing and reconfiguration in a network with dynamic traffic load.

determines the number of possible reconfigurations steps between two load changes. Given a constant time interval Δt_{switch}, Δt can be characterized by the amount of relay stations $k \in \mathbb{N}$, which may accomplish subsequent reclustering trials, until the load changes again. It is therefore possible to define a frequency of load change

$$f_{load} := \frac{1}{k} \tag{3.12}$$

as the reciprocal value of the maximum number of feasible reconfigurations steps k in the interval Δt. This means that the frequency $f_{load} \rightarrow 0$, if $k \rightarrow \infty$. This case corresponds to a static configuration with the possibility of an infinite number of reclustering steps to achieve the optimum clustering state. By definition, the worst case is given by $f_{load} = 1$, i.e. only a single station receives the opportunity to perform a reclustering trial, before the network load status changes again. It does not make sense to allow higher frequencies $f_{load} \geq 1$, as it is obvious that reclustering is not reasonable, if the load distribution changes much faster than the network can react. Please keep in mind that the absolute values of Δt and Δt_{switch} do not matter. The essential relationship between the network's dynamic and its intrinsic ability for self-reconfiguration is expressed by k.

Performance Evaluation Criterion

We have already addressed the need of a balance indicator, which makes a measurable statement, whether a network is well-balanced or not. Load balancing is a well-known problem for decades such that a large number of performance metrics exist. We have chosen a corresponding definition of a balance index β as introduced in [JCH84]:

$$\beta := \frac{\left(\sum\limits_{n=1}^{N} \lambda_n\right)^2}{N \cdot \sum\limits_{n=1}^{N} \lambda_n^2}. \tag{3.13}$$

This balance index β has already been applied successfully to wireless access networks in [HHKV01] and [Sei05]. Please recall that λ_n is defined as the load of a cell n. Referring to [JCH84], β is independent on the size, scale, and the metric of λ. To illustrate these properties, let us consider a totally imbalanced load situation. Without loss of generality, assume a tremendous overload $\lim\limits_{\lambda_1 \to \infty}$ in cell $\mathcal{C}_1$, while all other cells $\mathcal{C}_2, \ldots, \mathcal{C}_N$ are free of any load: $\lambda_2, \ldots, \lambda_N = 0$. This leads to

$$\lim_{\lambda_1 \to \infty} \beta = \frac{\left(\sum\limits_{n=1}^{N} \lambda_n\right)^2}{N \cdot \sum\limits_{n=1}^{N} \lambda_n^2} = \frac{\lambda_1^2}{N \cdot \lambda_1^2} = \frac{1}{N}. \tag{3.14}$$

The equation above reveals a lower bound for overload situations in dependence on the number of cells N. Additionally, it illustrates that β approaches asymptotically zero for an infinite number of cells. Given the other case of a fully balanced network with $\lambda_n = \lambda_{mean}$ for all $n \in \{1, ..., N\}$ we obtain

$$\beta \bigg|_{\lambda_n = \lambda_{mean}} = \frac{\left(\sum\limits_{n=1}^{N} \lambda_n\right)^2}{N \cdot \sum\limits_{n=1}^{N} \lambda_n^2} = \frac{(N \cdot \lambda_{mean})^2}{N \cdot N \cdot \lambda_{mean}^2} = 1. \tag{3.15}$$

Thus, total balance implies $\beta = 1$, while high imbalance is expressed by β close to $1/N$. It is possible that two different network configurations achieve the same balance index. In this case, the configuration, which requires less wireless hops to supply all relay stations, is preferable. Therefore, the required hop count of the configuration is used as a secondary decision criterion.

3.4.2. Numerical Optimization based on Global Knowledge

Problem Formulation We want to determine an optimum solution to the clustering process. Thereby, it is our intention to equalize the traffic load among all the cells. For that purpose, the load deviation of a cell $\Delta\lambda_n$ is defined as absolute value of the difference of a load λ_n of cell n and the average cell load $\tilde{\lambda}$
$\forall n \in \{1, \ldots, N\}$:

$$\Delta\lambda_n := |\lambda_n - \tilde{\lambda}|, \tag{3.16}$$

whereby the average cell load $\tilde{\lambda}$ is determined by

$$\tilde{\lambda} := \frac{1}{N}\sum_{n=1}^{N} \lambda_n. \tag{3.17}$$

To describe the absolute value $|(\cdot)|$ by linear equations, it is enough to characterize $\Delta\lambda_n$ by its lower bound given by
$\forall n \in \{1, \ldots, N\}$:

$$\Delta\lambda_n \geq \lambda_n - \tilde{\lambda} \tag{3.18}$$

and
$\forall n \in \{1, \ldots, N\}$:

$$\Delta\lambda_n \geq \tilde{\lambda} - \lambda_n. \tag{3.19}$$

The lower bound is sufficient, as a minimization of $\Delta\lambda_n$ is intended. The delta terms are aggregated in the objective function Ω_{load}, which describes the overall minimization goal:

$$\Omega_{load} := \sum_{n=1}^{N} \Delta\lambda_n. \tag{3.20}$$

We want to formulate the problem as a LP according to

minimize	aggregated deviations of cell loads from average cell load
subject to	
	clustering constraints

and therefore look for a set of clustering constraints to describe and restrict the clustering task mathematically.

Clustering Constraints We have already introduced the notion of a subgraph $\mathcal{S} = (\mathcal{V}, \mathcal{E}_s, g)$ to describe the clustered network. For that purpose, the resource allocation on the outgoing edge of a relay station is defined by the load of the distinct station and the sum of allocated resources on all incoming edges from other attached relay stations:
$\forall V \in \mathcal{V}_R$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} g(\vec{e}) = \tau(V) + \sum_{\vec{e} \in \mathcal{E}^{(in)}(V)} g(\vec{e}). \tag{3.21}$$

The cell load λ_n of each multi-hop cell is determined accordingly at all access points:
$\forall \mathcal{C}_n, V_A \in \mathcal{C}_n$:

$$\lambda_n = \tau(V_A) + \sum_{\vec{e} \in \mathcal{E}^{(in)}(V_A)} g(\vec{e}). \tag{3.22}$$

Please recall that each relay station is connected in an inverse tree structure to exactly one other relay station or an access point alternatively. This can be expressed in mathematical terms by
$\forall V \in \mathcal{V}_R$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} \operatorname{sgn}\{g(\vec{e})\} = 1, \tag{3.23}$$

with sgn$\{\cdot\}$ known as the signum function. It is possible to bypass the signum function by a binary variable $g_h(\vec{e}) \in \{0,1\}$ with
$\forall \vec{e} \in \mathcal{E}$:

$$g_h(\vec{e}) \cdot C \geq g(\vec{e}) \tag{3.24}$$

in order to keep a linear problem structure. Provided C as a large integer constant, $g_h(\vec{e}) = 1$ for all $g(\vec{e}) \neq 0$. The replacement of Equation (3.23) by
$\forall V \in \mathcal{V}_R$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} g_h(\vec{e}) = 1 \tag{3.25}$$

enables each relay station to connect to exactly a single, other station. In order to assure that all relay stations are finally attached to access points and loops are avoided, the overall load in the network is equated with the sum of all individual traffic loads $\tau(V)$:

$$\sum_{V \in \mathcal{V}} \tau(V) \overset{!}{=} \sum_{n=1}^{N} \lambda_n. \tag{3.26}$$

The set of Equation (3.21) to Equation (3.26) defines the clustering process such that each relay station is assigned to exactly one access point. Let us therefore summarize the clustering task as MILP:

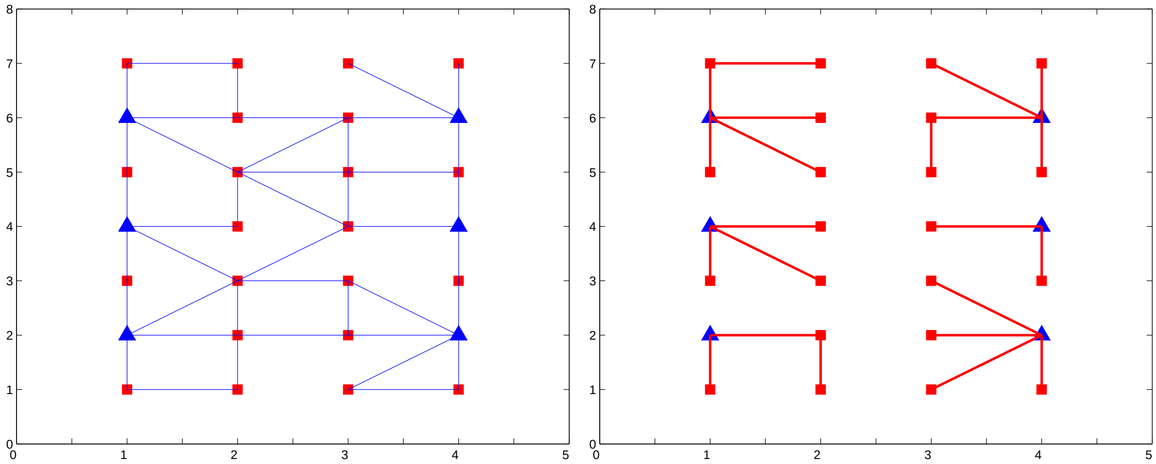
$$
\begin{array}{lll}
\textit{minimize} & \Omega_{load} := \sum_{n=1}^{N} \Delta\lambda_n & \\
\textit{subject to} & & \\
 & & \tilde{\lambda} = \frac{1}{N} \sum_{n=1}^{N} \lambda_n \\
 & \forall n \in \{1, \dots, N\} & \lambda_n - \tilde{\lambda} \leq \Delta\lambda_n \\
 & \forall n \in \{1, \dots, N\} & \tilde{\lambda} - \lambda_n \leq \Delta\lambda_n \\
 & \forall V \in \mathcal{V}_R : & \sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} g(\vec{e}) = \tau(V) + \sum_{\vec{e} \in \mathcal{E}^{(in)}(V)} g(\vec{e}) \\
 & \forall n \in \{1, \dots, N\}, V_A \in \mathcal{C}_n & \lambda_n = \tau(V_A) + \sum_{\vec{e} \in \mathcal{E}^{(in)}(V_A)} g(\vec{e}) \\
 & \forall \vec{e} \in \mathcal{E} & g(\vec{e}) \leq c(\vec{e}) \\
 & \forall \vec{e} \in \mathcal{E} & g(\vec{e}) \leq g_h(\vec{e}) \cdot C \\
 & \forall V \in \mathcal{V}_R & \sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} g_h(\vec{e}) = 1 \\
 & & \sum_{V \in \mathcal{V}} \tau(V) = \sum_{n=1}^{N} \lambda_n \\
 & & g(\vec{e}), \lambda_n, \Delta\lambda_n, \tilde{\lambda} \in \mathbb{R}_0^+, g_h(\vec{e}) \in \{0; 1\}.
\end{array}
$$

This optimization problem can be solved by numerical solver software. However, please take a note that the solution requires global topology knowledge and its complexity depends on the overall node number $|\mathcal{V}|$. These issues make the problem difficult to handle, especially for large networks. We therefore want to develop an alternate distributed solution based on local knowledge, before we present some evaluation results for both approaches.

3.4.3. Greedy Algorithm based on Local Knowledge

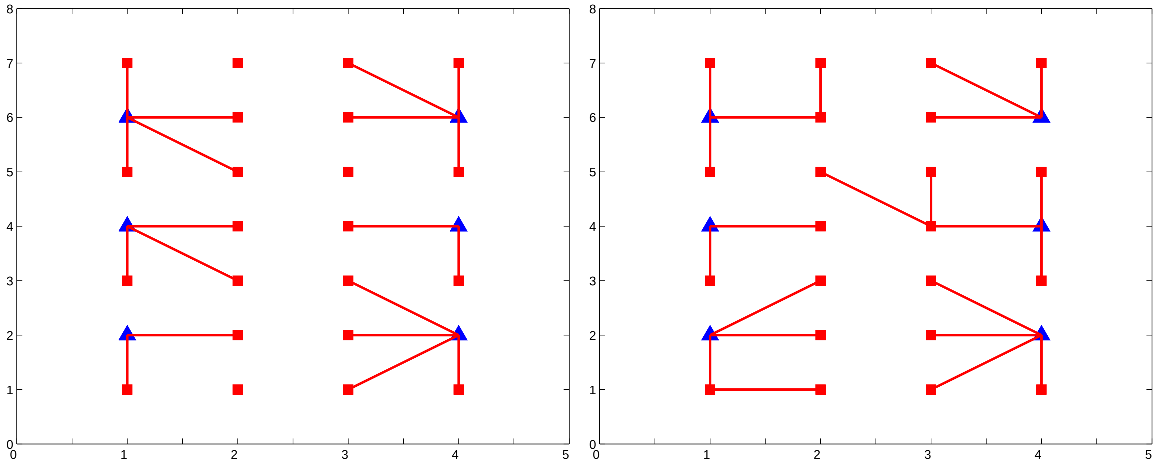
Greedy Concept The larger a network and the more nodes involved, the more difficult it is to determine an optimum solution with respect to clustering, routing, or similar tasks. This due to the fact that the number of possible solutions is increasing with each additional node. Consequently, the numerical approach is not applicable anymore for networks with medium size (∼200 nodes) for reasons of complexity. We therefore propose a greedy heuristic to replace the numerical optimization for large networks. The advantage of this approach is its full scalability, independently on the network size. In addition to that, the greedy algorithm is constructed in a way that fast convergence and a good approximation to the optimum solution can be achieved on the same time. Unfortunately, the closeness to the optimum solution is not granted, even if probable. Typically, suboptimal solutions show up, whenever the heuristic is caught in a local optimum and thus prevented from achieving a global optimum.

Heuristics like for instance simulated annealing prevent the process of getting stuck by allowing the step backward to an inferior configuration. The probability of accept-



(a) Connectivity of example network.

(b) Logical start configuration determined by greedy algorithm.



(c) Start configuration after first initialization of greedy algorithm.

(d) Final configuration after convergence of greedy algorithm.

Figure 3.6.: Plain field example network topology with 28 nodes and mesh degree $\gamma = 3$.

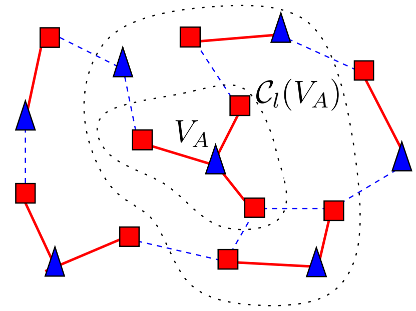


Figure 3.7.: Local environment $\mathcal{C}_l(V_A)$ of access point V_A.

ing a worse configuration than the given one is referred to as temperature in simulated annealing and decreases over time such that the solution is stabilized. We intentionally renounce such mechanisms in here, as the traffic load dynamics will automatically get the algorithm back on track over time. A temporary suboptimal solution will never result in a permanent state for that purpose and the greedy algorithm will recover after some time. Fast convergence and especially stability are more important criteria at this point.

Start Solution First of all, a start solution has to be found. In doing so, the list of access points is taken. Subsequently, access point by access point, all outgoing links are examined. If some outgoing link leads to a relay station and this relay station is not assigned to any other access point yet, then the relay station is assigned to the initial access point, using the distinct outgoing link. A corresponding illustration of the topology after the initial assignment is given in Figure 3.6(c) for the physical topology of Figure 3.6(a). In a second step, the list of all relay stations is examined and in a similar way as previously, all outgoing links are checked. Whenever the link leads to a station not yet assigned, the link is activated and the corresponding relay station is assigned to the other relay station's access point. This process is repeated until all relay stations are assigned to an access point. Thus, a start solution is created. The situation after the start-up phase looks like the scenario shown in Figure 3.6(b).

Step-by-step Improvement Subsequently, the start solution has to be improved step by step. For that purpose, let us define $\mathcal{C}_l(V_R)$ as the set of cells within the local environment of relay station V_R. The local environment is determined as the set of all neighboring cells, which can be reached in a wireless single-hop transmission by the relay station V_R. Accordingly, the local environment of an access point V_A is defined as the set of all neighboring cells, which can be reached in a single-hop transmission by the access point itself or any of its attached relay stations as sketched exemplarily in Figure 3.7. Let us consider a relay station $V_R \in \mathcal{C}_n$, which is picked randomly from the list of all relay stations. This station is supposed to find a cell $\mathcal{C}_k \in \mathcal{C}_l(V_R)$ in its local environment, which fulfills

$$\lambda_n \overset{?}{>} \lambda_k + \lambda_{thres} + \tau(V_R). \tag{3.27}$$

An estimation of the load of a neighboring cell $\mathcal{C}_k$ after a potential reclustering step is given by the sum of λ_k and $\tau(V_R)$. An artificial threshold λ_{thres} load is added to stabilize the clustering process. Reclustering makes sense in all cases, where the estimation for the load of the target cell after reclustering is smaller than the current cell load. This relationship is expressed by Equation (3.27). It is important to emphasize that the inequality successfully prevents oscillations, as only an improved overall load situation is accepted such that the reclustering process is naturally stopped at some time. λ_{thres} can be considered as a kind of hysteresis value. Its amount defines the sensitivity of the algorithm toward load imbalances. It helps to prevent network reconfigurations with minor improvement with respect to the reduction of cell load. The larger λ_{thres} is chosen, the less imbalances within the network will be detected and the more stable the current network configuration will be. Various trials have shown that the greedy algorithm with $\lambda_{thres} = 0.2 \cdot \tau_{max}$ converges faster, but its overall performance is limited, as not all options are checked. To illustrate the maximum performance, $\lambda_{thres} = 0$ is assumed in the following.

It is the greedy algorithm's final target to maximize the aggregated load deviation from the average cell load. However, the aggregated sum can only be determined with global network knowledge. Nevertheless, the greedy algorithm can optimize the different summands locally. Therefore, the local maximization criterion is defined by

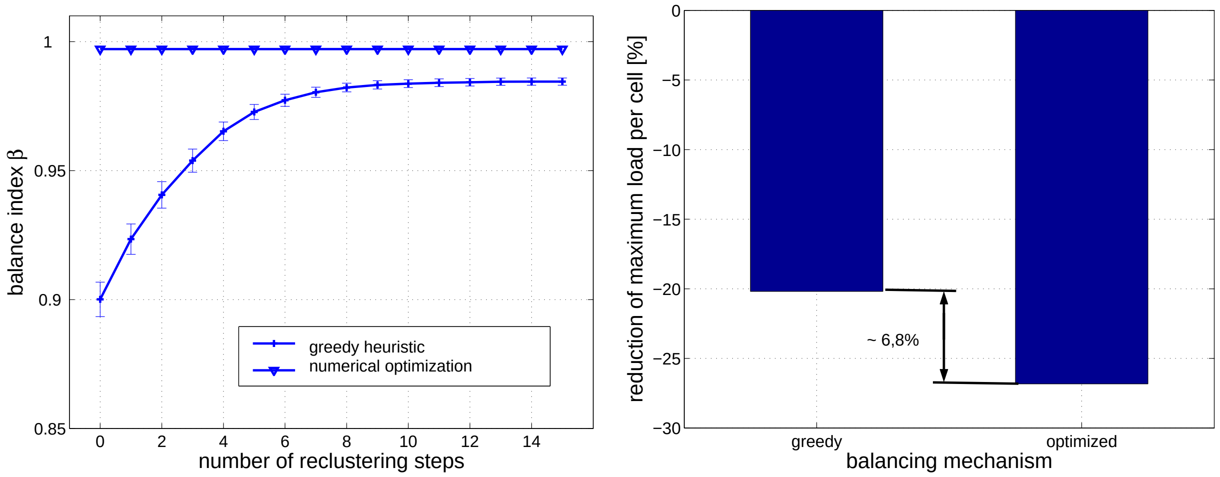
$\mathcal{C}_n = (\mathcal{V}_n, \mathcal{E}_n), V_R \in \mathcal{V}_n :$

$$\Omega_{gr} = \max_{\mathcal{C}_k \in \mathcal{C}_l(V_R)} \{\lambda_n - \lambda_k\}. \tag{3.28}$$

This equation chooses a distinct cell with maximum load difference among all neighboring cells, which fulfill Equation (3.27). Afterwards, the relay station V_R is assigned to the new cell. In doing so, the cell load is decreased in the original cell and is increased in the new cell such that the maximum cell load within the network is equalized over time to an average value.

Successful Reclustering It is important to emphasize the difference between a reclustering trial and a successful reclustering step. As mentioned initially, a relay station randomly picked from the list of all stations will start the reclustering process. The reclustering is successful, if and only if the relay station V_R determines a potential target cell $\mathcal{C}_k \in \mathcal{C}_l(V_R)$ such that Equation (3.27) is fulfilled. Otherwise, the reclustering process remains a trial and is abandoned. Another station is chosen then to restart the reclustering process.

We have presented a numerical approach and a heuristic to transfer an unorganized set of access points and relay stations into a cellular multi-hop network. We will continue by an analysis, how the topology of a network impacts the performance of the clustering process.



(a) Greedy algorithm with upper bound given by numerical evaluation applied to plain field scenario.

(b) Decrease of maximum cell load as effect of clustering.

Figure 3.8.: Self-organization of cell clustering process in plain field example network with 28 nodes.

3.4.4. Simulative Study

Self-organization of Cell Clustering

Recalling the initial example of Chapter 2, about 400 access points and relay stations are required to cover a considerably small inner-city area of 2 km radius. It is fairly obvious that the large number of stations arises the question, whether a centrally managed network is feasible for the ease of network management and maintenance. Thinking about WLAN technology, topology and frequency management is a decentralized and self-organizing process, as the different access points detect mutual beacon signals to figure out a potential transmission channel with the least interference possible. Recalling these issues, it is an interesting question to examine, whether self-organization in a cellular clustering process is generally possible. For that purpose, let us mutually compare the solution calculated by the numerical optimization provided global knowledge about the load status with results of the greedy algorithm based upon a local estimation of the load.

The analysis will be restricted to static scenarios at first, i.e. $f_{load} = 0$. We address a basic scenario with 6 access points and 22 relay stations as shown in Figure 3.6(a). The results printed in Figure 3.8(a) indicate the mean values of 300 independent simulations with different traffic loads, whereby the traffic values are assigned uniformly distributed random variables $\tau \in [0; 1]$. The result of the numerical optimization can be performed in a single, network-wide optimization step and serves as upper boundary for the greedy algorithm's result. In contrast to the numerical optimization, the greedy algorithm requires several subsequent and distributed configuration steps to optimize the balance index. About 6 - 8 reconfiguration steps on average are necessary to balance the load near the optimum, before the process enters the zone of saturation. Given the fact that there are 22 relay stations available, this means that a large percentage of all stations is involved in the reconfiguration process. Simulation results have shown that every further reclustering step does not lead to any significant performance gain after around 10 successful reclustering steps.

The analysis in Figure 3.8(b) shows that load balancing may indeed help to equalize the load distribution in cellular multi-hop networks. For that purpose, we consider the cell with maximum traffic load λ_{max} within the network. The maximum traffic load λ_{max} is a parameter of special interest in practical networks, as this peak value is interesting for network planning purpose and also represents a capacity limiting factor. We therefore define λ_{max} as the maximum of the loads λ_n of all cells $\mathcal{C}_n$,

$$\lambda_{max} := \max_{n \in \{1,\dots,N\}} \{\lambda_n\}. \tag{3.29}$$

The major intention of the applied load balancing mechanism is an equal distribution of traffic load within the different cells of the network. Therefore, the normalized change $\Delta\lambda_{max}$ of the maximum load per cell before and after the reclustering process is an indicator of practical relevance. For that purpose, the load λ_{max} of the cell with maximum load is evaluated before and after the application of the balancing mechanism:

$$\Delta\lambda_{max} := \frac{\lambda_{max}^{(before)} - \lambda_{max}^{(after)}}{\lambda_{max}^{(before)}}. \tag{3.30}$$

Please be aware of the fact that the cell with maximum load can differ before and after the reclustering process. It is possible that the Dynamic Cell Clustering algorithm causes changes to the load distribution in a way that the maximum load after reclustering is determined by a completely different cell in the network.

According to Figure 3.8(b), the average load reduction of the maximum load per cell is about 20% for the greedy algorithm and about 27% for the numerical optimization in the example network. The analysis has shown even one case, where the numerically evaluated maximum load per cell has been reduced by 56% compared to the start configuration of the greedy algorithm. Nevertheless, it requires quite a number of reclustering steps (6 - 8) to achieve significant performance gains. Given the comparably small network size (28 nodes), the number of required reclustering steps appears high. However, the big advantage of the greedy algorithm is the fact that it can be applied in a decentralized environment. The evaluation result is really astonishing, as the start configuration is — as Figure 3.6(c) indicates — a very intuitive setup. One might imagine the start configuration of the greedy algorithm as shown in Figure 3.6(c) as a reasonable setup for a multi-hop network with fixed relay assignment. Given the enourmous reduction in maximum load by Dynamic Cell Clustering, the tremendous potential of this technique becomes visible.

Summary The results of this initial study can be summarized by two aspects: The evaluation indicates that there is indeed a significant reduction of the maximum load by Dynamic Cell Clustering such that the approach itself is promising. Furthermore, the greedy approach achieves good performance results in smaller networks in comparison to the numerical evaluation. Self-organization of the clustering process in cellular networks seems to be possible according to these results. This is an important issue, as the envisioned target network with about 400 nodes cannot be evaluated numerically due to its complexity.

Mesh Degree

Imagine the situation of a network provider, who is willing to invest some money in a cellular multi-hop infrastructure network. Clearly, a certain number of wireless routers is unavoidably needed to achieve coverage. However, after a distinct coverage level is reached, further added infrastructure routers will result in a higher mesh degree of the network topology without any increase in coverage. The higher mesh degree in turn influences the inherent ability of load balancing within the network. Please recall that the mesh degree of a network is defined as the average number of outgoing links at the network nodes. It is presumable that the higher the mesh degree, the more options in terms of routing paths will exist to attach a relay station to a given access point. Consequently, there will be more variations to form different sets of multi-hop cells and thus to distribute load within the different cells. The entrepreneur is now focused on the question, whether the expenses for the additional infrastructure are justified for instance by a higher level of customer satisfaction or some minor costs for example in minimizing the necessary peak capacities on rented wired access lines. The higher costumer satisfaction as well as the minor costs could result from an increased overall network capacity, simply due to an intelligent load balancing mechanism, which avoids congestion and local bottlenecks efficiently.

We therefore want to take closer look on the question, whether a higher mesh degree really helps to achieve better load balancing results. A torus network with 360 nodes as exemplary illustrated in Figure 3.4 will serve as example network. The size of the network implies that only the greedy algorithm is applicable. 30 different randomly meshed network topologies with an access point density $\rho = 0.2$ are determined and 10 different traffic load distributions are assigned to each topology. In doing so, every measurement point of the simulation results is based on 300 different simulations. This procedure is repeated for a set of mesh degrees from $\gamma = 2$ up to $\gamma = 8$.

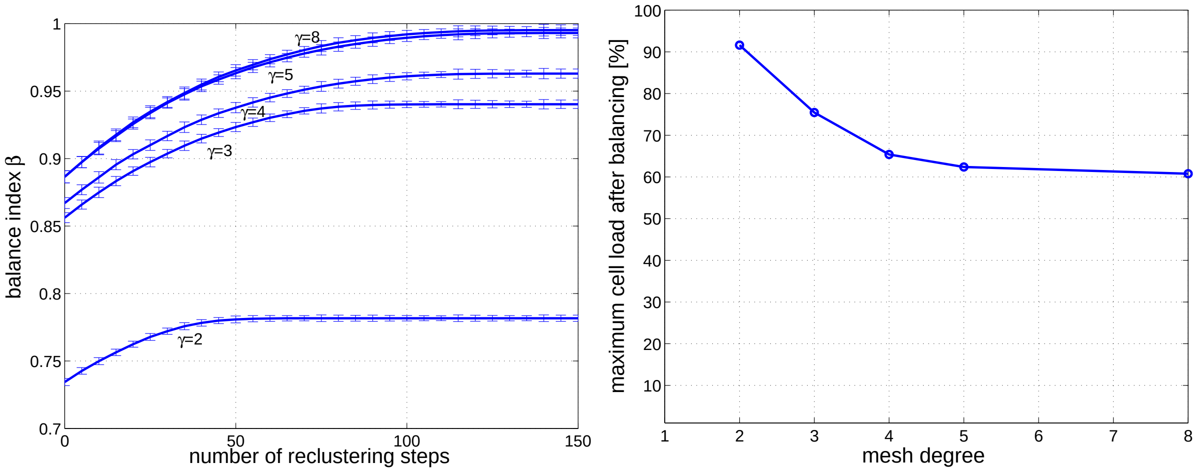
Figure 3.9(a) shows the evaluation results for the application of the greedy algorithm. After the initialization procedure, we observe an initial performance gap of about 0.15 units between the topology with mesh degree $\gamma = 2$ and $\gamma = 8$. Recalling the start-up procedure for the greedy algorithm, there is an initial assignment round, where relay stations are subsequently attached to an arbitrary access point within the one-hop neighborhood. The analysis shows that most relay stations are assigned to an access point after the initial start-up procedure for a network with high mesh degree. The remaining relay stations are recursively attached to other relay stations in a second step. It turns out that the number of remaining relay stations for the second step is much higher for networks with low mesh degree, as the chance of getting a direct connection to an access point is simply smaller in this case. The risk of an unequal load configuration becomes implicitly higher then, as each relay station assigned in the first

round increases the likelihood of an access point to receive further relay stations in the second round. Thus, the initial cellular assignment tends to become imbalanced with decreasing mesh degree.
Similarly, it is observed that the level of optimization after many iterations of the greedy algorithm is reduced by more than 20% for mesh degree $\gamma = 2$ in comparison to $\gamma = 8$. The explanation can be found — similar to the start configuration — again in the lack of suitable link choices to optimize the cell configurations. It also important to point out that the simulations of Figure 3.9(a) reveal that a nearly optimum performance can be achieved for high mesh degrees, as β is very close to its upper boundary. It can be also seen that the performance gain by a relative increase of the mesh degree decreases asymptotically for higher mesh degrees. While — as already explained — the performance results are fairly poor for $\gamma = 2$, there is practically no difference between the configuration for mesh degree $\gamma = 5$ up to $\gamma = 8$. This fact appears reasonable, as once there are enough physical link options for logical routing paths available (apparently $\gamma \geq 5$ in the presented scenario), any additional physical link will not make a contribution to broaden the variety of routing paths. Although any physical link is very valuable to the network topology in the beginning, the importance of a single link decreases intuitively with the overall number of links.

Summary The results of these simulations show that the application of Dynamic Cell Clustering will only achieve significant performance gains in networks with a mesh degree $\gamma \geq 3$. This means that a corresponding average mesh degree shall be ensured by the network deployment process, otherwise the application of Dynamic Cell Clustering will only have minor effects.

Reduction of Maximum Load

The balance index β is a performance indicator, which is theoretically interesting due to its properties as for instance scalability and comparability of different topologies. However, it has been already mentioned that the practical change of the maximum load per cell within the network as a direct consequence of the load balancing mechanism is of even greater relevance for a real network operator.
For these reasons, the maximum load per cell after balancing and the influence of the mesh degree shall be investigated one more time in a torus network with 360 nodes. Please take a look at Figure 3.9(b) for that purpose, which illustrates the impact of the mesh degree on the maximum load per cell in the network. Given the same set of simulations as presented above, it turns out that the maximum load per cell λ_{max} is decreased by 40% for $\gamma = 8$, which is a very substantial reduction. The performance gains are only about 8% compared to otherwise 40% for a low mesh degree of $\gamma = 2$. Thus, a very fundamental result for the design of those networks is the fact that a mesh degree $\gamma \geq 3$ would be desirable in order to achieve reasonable performance gains of more than 25%. Please recall the fact that the small network example 3.8(b) has also mesh degree $\gamma = 3$ and indicates performance gains of 20% for the greedy strategy and 27% for the numerical optimization. A comparison reveals that results are thus comparable and the approach seems to be scalable to large network sizes.



(a) Balance index β in dependence on the mesh degree γ.

(b) Percentage of average maximum traffic load after reclustering in comparison to initial state.

Figure 3.9.: Mesh degree as topologically defined performance boundary for load balancing in torus-shaped network topology with 360 nodes.

Summary To sum up, two major results have to be pointed out: On the one hand, a minimum mesh degree ($\gamma \geq 3$) is necessary to fulfill the topological prerequisites, which enable the load balancing mechanism to result in significant performance gains. On the other hand, the application of load balancing may reduce the maximum load dramatically, i.e. 40% for high mesh degrees.

Access Point Density

Although it might sound superfluous to be mentioned, it is important to consider the access point density of the network as a key parameter in the network planning process. Access points play an important role in the establishment of a multi-hop network, as the access points are wired with the backbone network and therefore act as lifeline of the mobile communication network. Unfortunately, the wiring of the access points is a costly business with twofold regard: Firstly, deployment costs are fairly high due to the fact that the wire line has to be connected with the access point. Secondly, the connections are also mirrored in monthly operational costs. Typically, the capacity of the network is increased with raising access point density ρ. However, the more dominant the access points are with respect to the overall number of nodes in the network, the less flexible the network will behave toward load balancing. This is related to the property that — by definition — each access point establishes its own multi-hop cell such that the number of cells is therefore equal to the number of access points.

According to the initial definition, the load in each multi-hop cell is given by the sum of access point's and relay stations' traffic. Load balancing is infeasible in the extreme case of $\rho = 1$, as all nodes of the network are access points. The load distribution is strictly defined by the initial traffic values of the access points in this case. Given a fixed overall number of nodes, the average number of relay stations per multi-hop cell is increasing, when decreasing ρ. This means in turn that the number of possible cell variations will raise and so the probability of the balancing algorithm to distribute load about equally will increase as well.

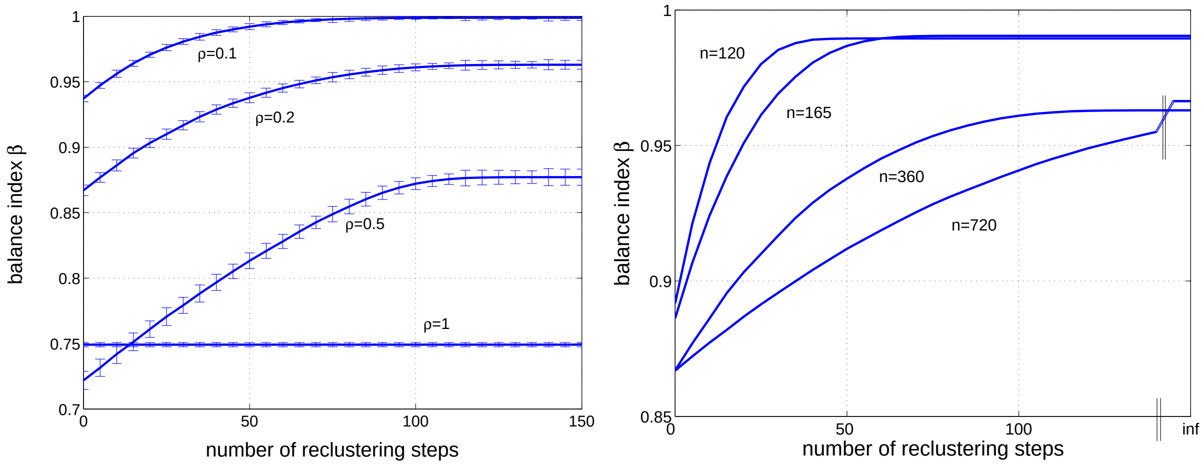
30 torus-shaped random network topologies with 360 nodes similar to Figure 3.4 have been evaluated for different access point densities each. The average results for these topologies are shown in Figure 3.10(a). This figure clearly illustrates that the balance index β is constant for $\rho = 1$. The final value $\beta = 0.75$ corresponds to the theoretical expectation value. Provided $\rho = 1$, the load value of each cell is given by the uniformly distributed random variable τ. For a large number of cells N, we receive:

$$\beta = \frac{\left(\sum\limits_{n=1}^{N} \lambda_n\right)^2}{N \cdot \sum\limits_{n=1}^{N} \lambda_n^2} = \frac{\left(\frac{\sum\limits_{n=1}^{N} \lambda_n}{N}\right)^2}{\frac{\sum\limits_{n=1}^{N} \lambda_n^2}{N}} = \frac{(E\{\tau\})^2}{E\{\tau^2\}} = \frac{\left(\int_0^1 \tau d\tau\right)^2}{\int_0^1 \tau^2 d\tau} = \frac{3}{4}. \tag{3.31}$$

The fact that the graph for $\rho = 0.5$ starts below the expectation value of the graph with $\rho = 1$ is a consequence of the initialization phase of the greedy algorithm. It does not take into account load and therefore may lead to severe imbalances, if only few relay stations are available.

While ρ is decreasing, the performance curves approach the theoretical maximum value $\beta = 1$. Given a density $\rho = 0.1$, which corresponds to an average of 9 relay stations per access point, an optimum performance of $\beta = 1$ is almost achievable. Thus, Figure 3.10(a) can be seen as illustration of the natural performance boundary of the balancing algorithm for a fixed access point density. A good performance is practically infeasible for very high access point densities, as the application of the algorithm is limited by the topology in this case. A low density leaves plenty of room for optimization instead.

Summary Please do not make the mistake to interpret this figure with respect to the question, whether a low access point density is generally desirable or not. When considering load balancing, the access point density should be taken rather as an extrinsic impact factor than a design parameter. Usually, a high access point density has major advantages with respect to spatial reuse of resources and the overall network capacity, which are not considered in this case. Nevertheless, a high access point density ρ also limits the options and thus the benefit of balancing, which has to be emphasized at this point. The access point density therefore implicitly defines an upper performance boundary for the overall load distribution.



(a) The access point density limits the margin of load balancing between different cells.

(b) An investigation of random sample networks with different node number ($\rho = 0.2$) emphasizes the scalability of the balancing approach.

Figure 3.10.: Influence of access point density and network size on load balancing in torus-shaped random network topologies with $\gamma = 4$.

Scalability

It is difficult to determine the numerical optimum of the load balancing process for a larger number of nodes. The example has pointed out that about 400 wireless routers are a realistic number to achieve coverage in a city-center area. The scalability of the load balancing approach for larger networks cannot be neglected therefore and plays a key role for the practical applicability of the algorithm.

So let us consider a set of different example networks with increasing overall node number as shown in Figure 3.10(b). Starting with a node number of 120 nodes, the size is increased up to a maximum number of 720 nodes. As already shown mathematically in Equation (3.14), the lower bound for the balance index β depends on the number of cells N. This explains the decreasing start values for larger networks with increased N. Nevertheless, the figure reveals a similar final performance independently on the network size, although the gradient of the performance curves seriously depends on the network size itself. The load balancing algorithm will converge faster in small networks, while in larger networks more reconfiguration steps are necessary to achieve the same performance. The result of final convergence is comparable for all presented network scenarios, although the results for large network size seem to be slightly worse. This might be partly explained by the fact that the chosen network topology and traffic distribution are just random samples, which are subject to statistical variations.

Summary The interesting and important message of this paragraph is given by the statement that the balancing potential is scalable and almost independent on the network size provided a fixed mesh degree.

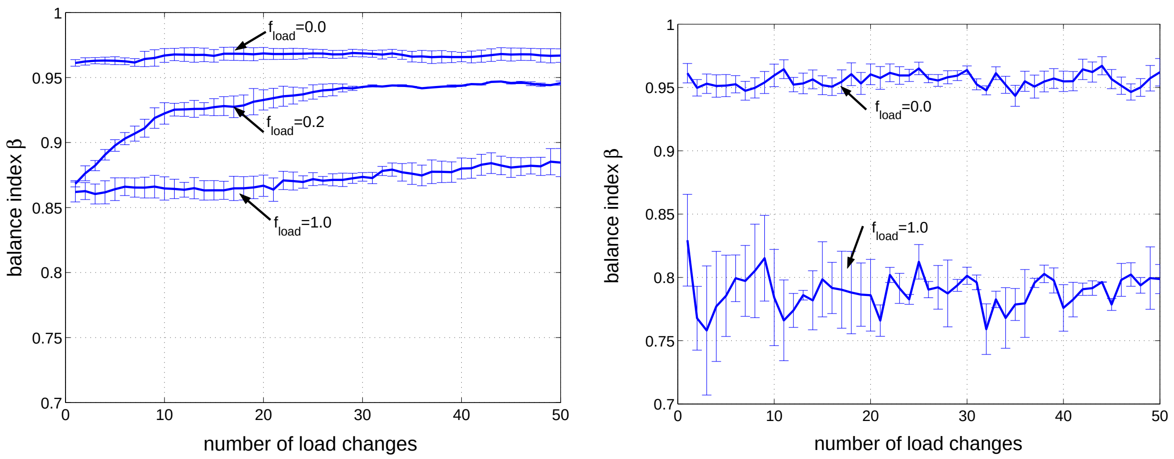
Effects of Dynamic Load Characteristics

While the convergence of the load balancing mechanism has been examined for a static traffic load situation in the previous sections, the load profile is subject to dynamics in real networks such that a static load cannot be reasonably justified over a longer period of time. For that purpose, any balancing mechanism with practical relevance will definitively have to deal with the dynamic characteristics of the network load. The dynamic in mobile communication networks is introduced by the mobility of its users. Typically, users are equipped with mobile devices, which implies that their location as well as the corresponding channel conditions are changing. On top, users may have time-dependent demands, which create a large variety of scenarios. A very common phenomenon is also the formation of groups. Due to some external trigger, for instance a show event or also an accident, a large crowd meets at a very small area. The traffic load will increase dramatically in this area, but only for a limited period of time. A good load balancing mechanism will take those effects into account.

The previous sections have examined a load situation, where the greedy algorithm has been offered an unlimited number of iterations to finally adapt to the optimum load state of convergence. The question is now, what will happen, if the balancing algorithm can only imperfectly converge, due to the dynamics of the traffic load. We know from control theory that there are three major constellations, namely the state of convergence, the state of non-convergence or chaotic behavior, and the state of oscillation, where system oscillates periodically between different configurations. It is obvious that there is nothing to be done from the technical point of view, whenever the interval between two subsequent load changes is even higher than the period between start and termination of a single reclustering step of one relay station from one cell to another ($f_{load} \geq 1$). In this case, the dynamic of the system is definitely too high to allow any reasonable adaptation such that the system will result in chaotic behavior. The case of oscillation is impossible due to the design of the greedy strategy, as any change in the network configuration without improvement is impossible.

In the following, we want to figure out the necessary number of reclustering steps between two subsequent load changes, for which a stable — i.e. in this context also balanced — network operation is feasible. For evaluation purpose, let us consider a torus-shaped network topology with 360 nodes ($\gamma = 4$ and $\rho = 0.2$). A series of 10 different random topologies is investigated, where the load profile is changed in a correlated suite of 50 load values per station. This means that subsequent load values within the same series are correlated, while values of different series are statistically independent.

Starting with Figure 3.11(a), a highly correlated suite of values ($\eta = 0.05$) is examined. The frequency of load change is varied from $f_{load} = 0.0$ (= full convergence between two subsequent load changes), over an intermediate frequency $f_{load} = 0.2$ (= five reclustering trials between two subsequent load changes) up to $f_{load} = 1.0$ (= one potential reclustering trial between two subsequent load changes). It can be observed that the network configuration is stable in all three cases. While the best performance is obviously achieved for $f_{load} = 0.0$, it is also illustrated that the network configura-



(a) Dynamic load change with $\eta = 0.05$. (b) Dynamic load change with $\eta = 1.0$.

Figure 3.11.: Influence of load dynamics on balancing performance in torus-shaped random network topology with 360 nodes.

tion stabilizes and improves over time for the other two cases. The configuration for $f_{load} = 0.2$ has almost reached the optimum after 50 iterations. The positive tendency is clearly visible for $f_{load} = 1.0$ as well, nevertheless the network configuration is still far away from the optimum case. The load balancing mechanism only proceeds fairly slowly due to the high frequency of load changes.

Figure 3.11(b) illustrates that the achievable maximum balance index for $f_{load} = 0.0$ is similar for $\eta = 0.05$ and $\eta = 1.00$. This means that the correlation between load values does not matter, as long as the balancing mechanism can converge, which is an expected result. The slight differences in the final values can be explained by statistical variations. Due to the different correlation factors in both cases, the optimum balanced index for both sets of random sample sequences is not the same. It is also visible that the variance of the final β value is much higher for $\eta = 1.0$. 10 simulations are apparently not enough to get fully rid of those statistical side effects. Nevertheless, Figure 3.11(b) reveals another important aspect: In cases, where the frequency of load change is assumed $f_{load} = 1.0$, a study of 50 changes of load iterations does not result in any stabilization. In this case, the dynamic of the load is too fast for the balancing mechanism such that the result is nothing else, but a statistical noise around an expectation value and no significant performance gains can be achieved. The results for $f_{load} = 0.2$ are slightly better, but all in all comparable to the results of $f_{load} = 1.0$. To avoid confusions, this graph is not printed, but one can assume it to be similar to the graph of $f_{load} = 1.0$.

Summary The very important result of this section is that small, periodical changes on the network load ($\eta \leq 5\%$) do not impede the stable operation of the greedy load balancing mechanism. It has been illustrated that the application of Dynamic Cell Clustering results in performance gains and depends on f_{load} and the random factor η. While the amount of η_{max} determines the maximum amount of the load change, f_{load}

limits the ability of the algorithm to recover again and to rebalance the network. The necessary value of f_{load} depends on the network size, as the number of nodes is related to the inertness of the network. However, the diagrams illustrate that performance gains can be achieved for reasonable values of η and f_{load}. The results therefore verify that the algorithm cannot only be applied to static networks, but can also deal with dynamic network configurations.

3.5. Towards a Reasonable Protocol Stack

Dynamic Cell Clustering helps to reduce the maximum load per cell significantly in cellular multi-hop networks. While the preceding sections are based on a graph-theoretical and mathematical model, we want to get some practical insights, how to include the presented ideas in a precise and practically feasible network protocol stack.

3.5.1. Preliminary Thoughts on the Implementation

As generally known, the routing task is traditionally part of the network layer. The clustering process can be considered as a special form of hierarchical routing, where a relay station including all mobile terminals attached to this station are rerouted from one cell to another cell. The clustering unit is ideally part of the routing unit in the wireless device. Current network architectures range from highly centralized cellular networks up to self-organizing WLANs. Similarly, one could imagine different organizational control units of the clustering process in dependence on the overall network architectures. In detail, three major categories can be distinguished, namely

- *a centralized clustering approach*, where a central network management unit steers the clustering process for the whole or at least a larger area of the network,
- *a hybrid clustering approach*, which means that for instance the access point as the head of a local multi-hop cell organizes the cell configuration in exchange and collaboration with neighboring cells, and finally
- *a decentralized clustering approach*, where each relay station acts independently and tries to optimize the cell configuration by its non-selfish behavior.

The centralized approach involves global knowledge of the traffic situation within the network, but allows global optimization of the cell clusters in return. In addition to that, potential reassignments of relay stations can be performed in one network-wide configuration step, while the hybrid or decentralized approach require several distributed optimization steps, until a final configuration is reached, which is considered to be optimum by all nodes of the network. The big disadvantage of the centralized approach in combination with numerical optimization is the complexity related to the determination of the optimum solution. Indeed, complexity is increasing dramatically for larger network size, so is the signaling overhead for the collection of up-to-date measurement reports. The centralized solution obviously guarantees a high level of control in practical realizations of communication networks, especially if it comes to mechanisms like call admission control to constrain or to temporarily limit the access to the network. The hybrid as well as the decentralized approach involve distributed responsibilities and are therefore scalable with the network size. Global optimization is infeasible in these cases, simply due to lack of global knowledge. The optimization process toward the optimized solution includes several local balancing steps and therefore possibly does not lead to the global optimum. Nevertheless, the decentralized solution involves a lot of flexibility and higher independence against failures, as all units are acting autonomously.

3.5.2. Centralized Approach with Numerical Optimization

Let us first of all consider the centralized approach, where a central network management unit is assumed to be available in the wired core network. This could be for example the access gateway node in the LTE network architecture, which acts as centralized unit for a local set of base stations [EFK$^+$06]. In this case, the central network management unit is in charge of two major tasks, namely the surveillance of the current load situation and the triggering of necessary reclustering procedures.

In order to get a current view on the load situation, the clustering unit of the central unit has to demand periodically for load measurement reports to its corresponding access points. For this reason, the central unit sends out requests ASK_LOAD to all connected access points, in which the access points are pleased to hand in a load measurement report. As shown in the message sequence chart of Figure 3.12(a), the access points answer these measurement requests with a corresponding measurement report CURR_LOAD, which includes information about the connectivity of attached relay stations and the current traffic load situation. According to the LTE system architecture, the base stations already receive periodical measurement reports from all mobile terminals. In a similar way, this information is provided to the access points of a cellular multi-hop network, as each access point has a radio resource management unit, which is in charge of the distribution of radio resources to the access point and all attached relay stations.

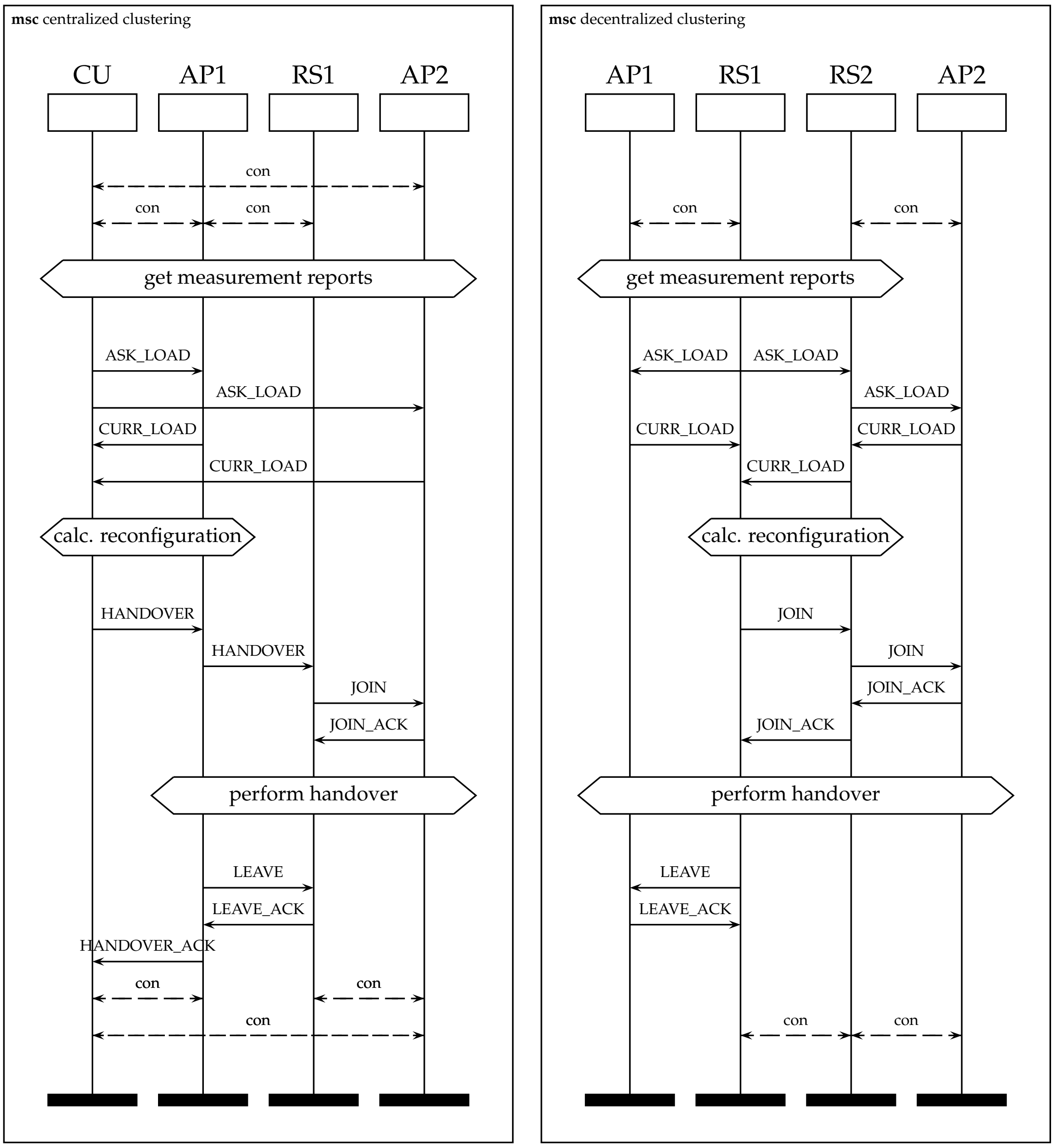
The central unit feeds all these input data into either the numerical optimization solver program or operates the greedy algorithm with centralized knowledge, until an intermediate or even optimum configuration is achieved. Whenever a reclustering process is suggested, the central unit sends out a HANDOVER request message to the distinct station. This message initiates a handover process of the corresponding relay station to a station in another cell, which is defined in the HANDOVER request message itself. After receiving this message, the relay station tries to join the distinct target cell, which is symbolized by a JOIN message. This process could look alike the HiperLAN/2 association procedure [ETS02] or the similar LTE association procedure (see [BHKV07]). Once the handover related steps like association and authentication are finished and confirmed by a JOIN_ACK acknowledgment, the relay station leaves the former cell by a LEAVE message. This message is in turn acknowledged by the corresponding access point again.

3.5.3. Decentralized and Hybrid Approach with Greedy Algorithm

While the approach presented in the previous section involves a centralized steering unit, another approach working in a fully decentralized manner is illustrated in the following. In this case, each relay station decides on its own, which cell to join such that a self-organizing network architecture is established. Clearly, we cannot rely on the numerical optimization algorithm in this case due to the lack of global knowledge. But we can alternatively deal with the proposed greedy algorithm. The good news about this concept is the fact that there is no single point of failure. This property turns the approach into a very robust and failure-tolerant one. In principle, the proceeding is similar to the centralized solution, but the decision is only taken by local and not by global knowledge.

Again, in order to start a handover process, a situation of network imbalance has to be recognized first of all. For that purpose, a relay station has to send out ASK_LOAD queries to receive corresponding load reports CURR_LOAD from all cells with direct connectivity to the station. This is also shown in the message sequence chart of Figure 3.12(b). The question, how to determine this corresponding station can be solved for example by a randomly determined back-off timer similar to the IEEE 802.11 MAC. The timer triggers the mentioned load report and thus starts a potential handover procedure. The correct choice of the back-off interval depends on the network dynamics and is clearly out of focus at this point. After the load situation is investigated and — eventually — the handover is performed, the station switches into back-off mode again. A potential problem appears, whenever neighboring stations start the load inquiry at the same time. Although this probability is small, if the back-off interval is chosen large enough, it makes sense to prevent that by locking a station, which receives an ASK_LOAD message for some time. Being locked means in this context that the corresponding station is not allowed to either answer any other load inquires or to start an inquiry on its own. Thus, the blocking effect of an ASK_LOAD message is comparable to the function of a ready-to-send (RTS) message in the IEEE 802.11 standard and the CURR_LOAD message comparable to a clear-to-send (CTS) message. The blocking time equals to the typical duration of a handover procedure plus some guard interval. The clustering process can also be executed in a distributed manner with simultaneous reconfigurations in different parts of the network as a consequence of the application of the proposed scheme.

The relay station shows its willingness to enter a new cell by a corresponding JOIN message. Received JOIN messages are forwarded to the access point, in case they are received by a relay station. After a positive acknowledgment of the JOIN message is received, the distinct relay station signals with a LEAVE message that it is connected to another cell right now. Of course, the LEAVE message is acknowledged by the former host access point. Any potential implementation requires an answer to the question, whether an access point is allowed to prevent a relay station from joining the new cell, i.e. to reject the JOIN request. Under the assumption that all stations act



(a) Message sequence chart of centralized reclustering approach.

(b) Message sequence chart of decentralized reclustering approach.

Figure 3.12.: Message sequence charts for centralized and decentralized reclustering approaches.

in a cooperative and constructive manner, this rejection mechanism makes no sense and becomes obsolete.

Hybrid Mode The previously mentioned hybrid clustering mechanism is realized as a combination of the centralized and the fully decentralized approach. In contrast to the decentralized approach, the load detection is triggered by the access point only in the hybrid mode. It can be therefore alternatively signaled over the wired backbone network, as all access points are attached to that. Once a local overload is detected, it is — similar to the centralized mechanism — not the central unit's, but the access point's task to send out a handover message to the concerned relay station. The handover itself is executed the same way as in the decentralized case such that the major difference between decentralized and hybrid clustering is the additional handover trigger by the access point.

Auto-configuration We have explicitly pointed out in Chapter 2 that the ease of maintenance is a major issue in relay-based networks. Given the large amount stations, replacements due to technical failures or extensions of the given infrastructure play an important role in the daily operation of the network. Whenever a station is supposed to join the network, there is the question, how to embed the new station in the regular network operation. The decentralized cell clustering algorithm can be easily extended by some auto-configuration mode. The auto-configuration mode is suited to enable for instance a relay station to automatically join the network. For that purpose, a suitable link detection mechanism is needed, which scans for friendly relay stations or access points within transmission range. Once the link topology is explored, the relay station could join an existing cell with a simple JOIN message very similar to the regular handover procedure in the decentralized clustering concept. The proposed auto-configuration mechanism alleviates the deployment of new relay stations and the maintenance of existing relay stations substantially.

3.5.4. Simulative Study

Implementation

In order to prove the general feasibility of the presented load balancing proposal, the hybrid balancing approach has been implemented in a network environment with real protocol stack. For that purpose, the routing unit of the network simulator as explained in Appendix A has been extended by a clustering component to accomplish the clustering process.

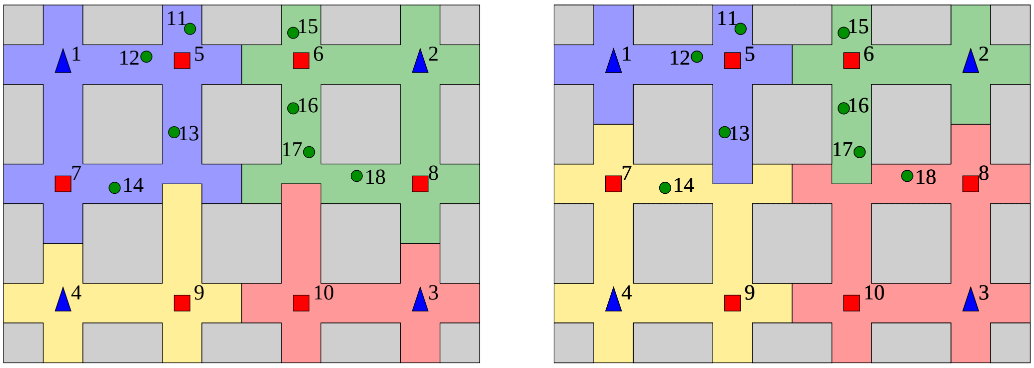
Mobile IP Routing in the backbone part of 3G networks from the SGSN to the GGSN is performed by the GPRS Tunneling Protocol (GTP) [3GP07], which is responsible for the routing from a central backbone node over the base station to the mobile terminal. To operate a cellular multi-hop network, a similar mechanism to bind mobile terminals to corresponding access points is needed. It is possible to therefore use Mobile IP, which is not directly comparable to GTP, but provides IP-based mobility to terminals for instance in a WLAN. Each mobile terminal is assigned to a home agent by Mobile IPv4. This home agent acts as permanent anchor point and reroutes the data traffic for the mobile terminal to the foreign agent. The mobile terminal is temporarily attached to these foreign agents, which represent the terminal's current location. In order to deal with multi-hop cells, Mobile IP has to be slightly changed. The standard Mobile IPv4 is modified in a way that the relay stations forward the agent advertisement messages of the access points to potential mobile terminals. Similarly, relay stations forward the agent reply messages in the other direction from the mobile terminal back to the access point. Each multi-hop cell can be considered as the coverage range of one Mobile IP agent then, i.e. a relay station is not acting as an agent on its own. In doing so, it is possible for the mobile terminals to perform inter-cell handovers between different multi-hop cells. Thereby, it does not matter, whether the mobile terminal is directly attached to an access point or indirectly over a relay station. Mobile IPv4 can be therefore used as inter-cell routing protocol to handle user mobility in cellular multi-hop networks.

Topology The scenario is based on a Manhattan topology with four access points, six relay stations and eight mobile terminals as shown in Figure 3.13. All access points are connected by wire to a central node, which acts as a common source of constant bit rate data streams to all mobile terminals. A proactive intra-cell routing protocol WOSPF as presented in Chapter 4 is used in each multi-hop cell to explore the topology and to establish the multi-hop routing paths. The clustering component is integrated into the routing unit and takes traffic information from the intra-cell routing protocol.

Discussion of Results

Start-up Phase Initially, the network traffic consists of only proactive routing and Mobile IP traffic without any data traffic. At time $t = 40$ s a constant bit rate traffic ($r = 100$ kbyte/s) is sent from the central node to each mobile terminal. Figure 3.14 shows that the balance index β decreases significantly after 40 s - 60 s. This is approximately the time frame, which is needed to receive load measurement reports. These reports are a precondition to evaluate the load situation and requested in a 10 s time interval. Thus, it takes about 20 s, until the traffic situation is detected at the access points. There is no scientific justification for the 10 s duration of the trigger. It is an arbitrary value, which has been chosen for demonstration purpose and could be reduced to smaller values. Nevertheless, the time trigger should not be chosen too small in a real implementation, as in this case the subsequent overhead by frequent measurement reports becomes large.

Clustering Phase An implemented timer function periodically calls the clustering unit and causes the access points to start a load inquiry process. Please take note that the reclustering interval is by far smaller than the overall simulation time,



(a) Start configuration at $t = 0$ s. (b) Configuration after $t = 75$ s.

Figure 3.13.: Effects of reclustering process in a Manhattan scenario.

$\Delta t_{switch} \ll T$. As shown in Figure 3.14, access point (1) and (2) detect a local overload. Therefore, these nodes try to assign one of their relay stations to another cell. This is achieved by reassigning relay station (7) including mobile terminal (14) to access point (4) and relay station (8) including mobile terminal (18) to access point (3) as indicated in Figure 3.13. In doing so, the coverage range of access points (3) and (4) is extended. Consequently, the load balancing causes the balance index to recover. In contrast, the balance index stays at the low level for the simulation with fixed relay station assignment. Please keep in mind that the balance index is calculated just for the purpose of analysis from an artificial god perspective.

Shut-down Phase The traffic load is switched off at $t = 160$ s. The remaining traffic is caused by fixed signaling traffic of the proactive routing protocol. The balance index raises in both, the simulation with and without cell clustering. This is simply due to the signaling traffic, as the distribution of the signaling traffic in absence of load is even more equal than the balanced situation with load. For this reason, the cell configuration remains constant and the balance index increases again toward the end of the simulation.

Summary This section has shown in a prototypical implementation that Dynamic Cell Clustering is indeed a suitable and also practically feasible mechanism to balance load among different cells in a cellular multi-hop network. The proposed implementation carries a prototypical character, as the assignment of the mobile terminal via Mobile IP to the access points is not an optimum replacement for GTP or similar protocols. Therefore heavy handover delays in case of mobility are expected. Nevertheless, the potential has been illustrated.

3.6. Conclusion

This chapter introduces a new cell clustering approach to balance load in cellular multi-hop networks. The mechanism has been examined and derived in a graph-

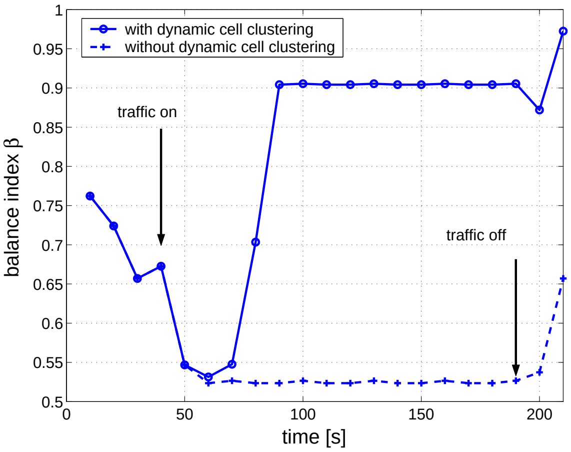


Figure 3.14.: Evaluation of the balance index β in a Manhattan example scenario.

theoretical model at first. Later on, a proposal for a prototypical implementation with practical feasibility has been given and simulated with the network simulator ns-2. The study has shown several major results: The most obvious, but probably also the most important result is the proof of principle that Dynamic Cell Clustering can indeed reduce the maximum load per cell in a network significantly. Optimum results can be achieved by numerical optimization, nevertheless the simple greedy heuristic is already sufficient to obtain convincing results. The main consequence of this is that a central entity to coordinate the clustering process is not necessarily needed. The decentralized greedy strategy can be even applied, if the network's load is dynamic. In this case, the performance depends heavily on the frequency of the load changes and the corresponding network size. In addition to these investigations about the balancing mechanism, some general statements about the topological requirements of a network to enable load balancing have been accomplished. It has turned out that the performance gain depends seriously on the mesh degree of the network and that a minimum mesh degree $\gamma \geq 3$ is recommended.

4. Intra-Cell Routing in Cellular Multi-Hop Networks

4.1. Introduction

While the previous chapter has shown several ways to cluster the infrastructure components of the network into different multi-hop cells, the question, how to handle user mobility within a single multi-hop cell, is not clear yet. The challenge is to assure that the mobile terminals are connected to the infrastructure network permanently as well as efficiently. The permanent connection is enabled by a suitable intra-cell handover approach, which passes the terminals from one to another router. Efficiency implies a resource-aware routing metric, which determines a routing path with small resource efforts.

The first part of the following chapter explains the need for a corresponding intra-cell routing protocol and takes a closer look at selected routing protocols for wired and wireless networks. We investigate these protocols regarding their suitability to act as intra-cell routing approach for multi-hop cells. Subsequently, we examine some exemplary routing metrics. The intention is to figure out with a simple graph-theoretical model, whether a more intelligent and therefore more complex routing strategy will yield significant resource savings. The second part deals with the question, how to realize such an intelligent routing protocol in practice. Thereby, the routing problem can be split into two subproblems, namely the topology exploration and the data delivery. To begin with, we present a topology exploration mechanism to announce local topology changes globally within the network. The consecutive data delivery requires prior topology exploration and includes the addressing of routing paths. For that purpose, we suggest a label switching scheme to realize the data delivery within the multi-hop cell. The combined proposal for topology exploration and data delivery represents a possible concept to enable intra-cell mobility in a single multi-hop cell. Before explaining the approach in detail, let us firstly highlight the demand for such an intra-cell routing protocol in the following.

4.2. Motivation

Joint Routing-Scheduling Problem Today's cellular networks are single-hop networks, where the spectrally efficient transmission of signals from the base station to the mobile terminal is of major interest. For that purpose, research in cellular networks is historically concentrated on efficient scheduling algorithms instead of routing issues. Scientific research has brought up many different scheduling approaches.

Round robin scheduling and proportional fair scheduling are two prominent examples for such schedulers. An excellent overview can be found in [FL02]. While routing is restricted to the wired part of today's cellular networks, routing between wireless stations is an integral part of cellular multi-hop networks. This means that routing, especially choosing a resource-efficient path to the final destination, is equally important to a spectrally efficient link transmission for each hop on that path. The development of an intra-cell routing approach is therefore a major prerequisite to allow a resource-efficient operation of the multi-hop cell.

Routing Metric Typically, there is more than one routing path to provide access for a mobile terminal to the wired backbone network of the operator. It is not really clear, how to choose the corresponding routing path, as there are several, contradicting optimization goals. A very obvious solution to the problem is for instance the attachment of the mobile terminal to the wireless router with the best incoming signal strength. However, this might occur at a router with a dense traffic situation such that following the strongest signal may lead to overload situations. Beyond signal strength and traffic load level, there are also other parameters to consider as for example the available transmission resources, potential processing and transmission delays, the hop number to the final destination, and also the related signaling overhead. The set of physical parameters and the user requirements represent constraints to a general optimization problem, which the network provider is subject to. For this reason, the network provider has to define an optimization objective or measure, which is generally referred to as routing metric. As the routing metric is the core of each routing protocol, it is a very apparent motivation to figure out, how the choice of the routing metric influences performance indicators such as spectral efficiency. A comprehensive routing metric obviously demands high standards of measurement data and also of computational complexity. The trade-off between the potential performance gain of a metric and the subsequent signaling overhead of required measurement data is therefore a very interesting research objective from the network provider's perspective.

Intra-Cell and Inter-Cell Handover Capability One key challenge of the routing protocol is a design, which supports the mobility of the user terminals within the multi-hop cell. Although mobility is restricted to typical inner-city velocities, i.e. a speed range between 0 m/s and 20 m/s at the absolute maximum, the number of expected intra-cell handovers is considerably high. This is related to the fact that the transmission ranges of the wireless routers in the 5 GHz frequency band are small compared to the ranges of GSM or UMTS base stations. Therefore, the realization of intra-cell mobility between wireless routers of the same cell is a major challenge for the routing protocol. This is particularly the case, as the compliance to the existing LTE inter-cell handover scheme has to be guaranteed at the same time. This calls for a proposal, in which a corresponding intra-cell handover is embedded in the LTE inter-cell handover framework such that both schemes provide supplementary and not substitutional mobility support.

Quality of Service The provisioning of mobility support is one element in a set of properties to characterize quality of service. Please recall quality of service as the allocation of previously negotiated resources. Seen in this sense, it becomes understandable that quality of service can be provided more easily, if the access point is also in charge of the cell-wide routing activities. An access point focused approach guarantees among other issues the local control of all connections by the access point and enables the installation of a call admission control. Business models with premium services like advanced voice or multimedia services will desperately rely on the efficient implementation of such policies, as costumers will only pay for real service differentiation. Therefore, any routing protocol of practical relevance has to at least notify the access point about ongoing routing activities, or even better has to provide a fast access and control opportunity regarding the routing activities for access point related components such as call admission control. The compatibility to today's cellular networks is evidently hard to handle otherwise.

After these general comments about the demand for and the requirements to a distinct intra-cell routing approach, we will take a look on selected existing routing protocols. In doing so, we want to understand the assets and drawbacks of different routing mechanisms.

4.3. Related Work

Certainly the easiest and not necessarily an inefficient way of routing is the definition of a permanent or static routing table. We can find such examples in local area networks, where a fixed gateway is defined in a static routing table. Static routing however cannot deal with the dynamic challenges of a modern wireless communication network, for instance with link breaks. Therefore, dynamic routing is the standard approach in networks with dynamics as for example the Internet or mobile ad hoc networks.

Historically, the problem of dynamic multi-hop routing firstly appeared in the wired network domain. The most prominent representatives of routing protocols are Open Shortest Path First (OSPF)[Moy98] and Border Gateway Protocol (BGP)[RL95][Cis06], but also Multi-Path Label Switching (MPLS) [RVC01]. The major difference between the wired and wireless link characteristics consists in the fact that wired links have usually a fixed, constant bit rate, while wireless links can be allocated dynamically and are subject to the instantaneous conditions of the wireless channel. Later on, the wireless community started research on ad hoc networks. Being a popular research topic, especially in academics, a tremendous amount of routing protocols has been invented over the last years in order to support multi-hop routes in ad hoc networks. Optimized Link State Routing (OLSR) [JCL$^+$01], Destination-Sequenced Distance Vector Routing (DSDV) [PB94], Ad hoc On demand Distance Vector routing (AODV) [PR99], and Dynamic Source Routing (DSR) [JMB01] are typically named as well-known examples for wireless ad hoc routing protocols.

We want to introduce some general concepts at this point, which characterize the properties of a routing protocol: Literature primarily distinguishes proactive, reactive, and also hybrid combinations of routing protocols. Proactive routing involves the periodic and traffic-independent exploration of the network topology, as observed in OLSR and OSPF. On the contrary, reactive routing is based on the philosophy to only react by on demand trouble shooting, whenever new events like the initiation of a connection or a link break occur. We can observe such a behavior for instance in AODV. Routing protocols typically rely on link information, which results from the connectivity of the network. Apart, there is also a large number of related approaches, which use further information, for example geographical positions. We will concentrate on purely link-based routing protocols in the subsequent sections.

There are different ways to communicate link conditions. Link state protocols (e.g. OSPF and OLSR) publicly announce local link information between neighboring nodes in the global network environment, while distance vector protocols (e.g. DSDV) only provide a node's distance to a final destination. The distribution of topology information to all stations within the network is not necessarily required, as the corresponding routing information can be included alternatively into the data packet header itself. The distinct approach is known as source routing. Source routing works in a way that any intermediate node just follows the guidelines in the packet header to determine the next-hop destination. DSR is a well-known example for a source routing approach. Label switching can be considered as a related approach, in which the address-based path information is replaced by a suite of path-defining labels. Intermediate routers use the attached labels and distinct switching tables to determine the next hop of an incoming packet. Some important conceptional design issue is the hierarchy imposed upon a network by the routing protocol. Addresses can cluster parts of the network into subnetworks. Certain nodes can be elected as gateway nodes with special responsibilities. Hierarchical routing concepts are usually introduced to deal with large networks, as otherwise the network architecture is not scalable. The corresponding complement is flat routing, which is especially suited for smaller networks. Flat routing is characterized by the lack of hierarchies such that all nodes are equal from the addressing point of view.

After this brief overview, we address the most prominent representatives of routing protocols in the wired and wireless domain. It is the main objective to develop a clear understanding of the major characteristics and steering parameters of those protocols. Obviously, their potential suitability for the operation of a multi-hop cell is particularly focused during this inspection.

4.3.1. Routing Protocols for Wired Networks

Open Shortest Path First

Open Shortest Path First (OSPF) [Moy98] is a proactive link state routing protocol, which is widely deployed in the domain of fixed networks. The OSPF scheme concentrates on two main features: At first, neighborhood topology exploration is accomplished by a simple topology exploration protocol based on HELLO messages. Secondly, the exchange of link state update messages represents an efficient mechanism to globally exchange locally collected topology information among all routers within the same routing area. Although the basic working principle of OSPF is simple, the complexity of the protocol relies on its highly sophisticated details, which are fixed in 37 RFC[1] standards.

Topology Exploration According to OSPF's basic functionality, a link connection between two adjacent routers is established by the HELLO protocol. For that purpose, routers regularly send out HELLO messages to all surrounding routers. Neighboring terminals answer the receipt of a HELLO message promptly by a corresponding HELLO reply message. A network-wide defined expiration deadline causes a link connection to time out after a certain period, once the mutual reception of HELLO messages has stopped. The local topology information, which is gathered by HELLO messages, is distributed in the network by the use of link state updates. Each link state update includes information about incremental changes in the network topology. As the link state update messages are flooded through the whole local area of the network, each router's link state database gets a synchronized view of the overall network topology.

Traffic Engineering Based upon the link state database, each router calculates a possible routing path to any other router in the same area. Typically, Dijkstra's shortest path algorithm [Dij59] or the more comprehensive Bellman-Ford [For56] algorithm are applied to determine such a routing path. Congestion is supposed to "frequently occur, when the shortest paths of multiple traffic streams converge on specific links" [AMA+99]. For that purpose, traffic engineering is used to enable efficient and reliable network operations, while simultaneously optimizing the network resource consumption [AMA+99]. The traffic engineering extension of OSPF is specified in [KKY03]. The major improvement of this extension is the additional definition of header fields, which allow bandwidth reservation in the link state update messages. It has been shown in [FRT02] and [FT00] that optimum traffic engineering belongs to the family of nondeterministic polynomial-time hard problems such that heuristics are reasonable approaches toward the optimum solution of those problems.
All in all, OSPF represents a proactive link state protocol for wired networks. The extensive efforts within OSPF to support traffic engineering turn this protocol into an interesting study object for the processing of advanced routing metrics.

Multi-Path Label Switching

Multi-Path Label Switching (MPLS) [RVC01] is a packet transport mechanism, which is especially useful for wide area fixed networks. MPLS allows a path-oriented packet transport as an alternative to destination-based forwarding. MPLS thereby acts in a regionally separated part of the overall network denoted as autonomous system. The autonomous system has several interaction routers to other parts of the network. Packets may enter the MPLS area at these routers, which are also named ingress routers. Complementary, there are egress routers, at which packets may leave the

---

[1] Request For Comments (RFC)

autonomous system. All routers within the autonomous system are called label-switched routers. During their travel through the autonomous system, only the assigned labels of each packet are checked and processed to determine the further path. Two different types of MPLS can be distinguished after [Law01] in dependence on the protocol stack, namely

- *a packet-based MPLS type*, according to which the label routers have full header- and content-decoding capability and can cope with layer 3 header information, and
- *a switch-based MPLS type*, which provides label routers with only layer 2 functionality and reduces forwarding to a switching process through the networks.

Label Distribution A label distribution protocol [ADF+01] is applied for the assignment of labels. MPLS is a connection-oriented data transport approach, in which a path is established before the first data packet is delivered. Typically, the path is composed by the reservation of several labels for individual path segments. Label-switched routers have local mapping tables for that purpose. The label is embedded into a 4 byte MPLS header (see Figure C.1(a)), which is added at the ingress router, evaluated at all intermediate routers, and finally removed at the egress router. MPLS can be combined with resource reservation schemes as for instance the Resource Reservation Protocol (RSVP) [BZB+97] and its traffic engineering extensions Resource Reservation Protocol - Traffic Engineering (RSVP-TE) [ABG+01] to fulfill distinct quality of service criteria. It is an important aspect that MPLS cannot act as a stand-alone routing protocol, but operates on top of an interior gateway protocol, for instance OSPF [Law01].

There are two major observations to record at this place: Firstly, MPLS represents an approach to provide connection-oriented transport services in a packet-oriented network. Secondly, the assignment of labels is an efficient path-based signaling mechanism.

4.3.2. Routing Protocols for Wireless Networks

Optimized Link State Routing

Optimized Link State Routing (OLSR) [CJ03] is driven by the desire to develop a link state protocol adjusted to the specific conditions of a mobile ad hoc network. For that purpose, [JCL+01] has come up with a so-called multi-point relay approach. The authors consider OLSR as an "optimized" link state protocol, which arises the question, how a "non-optimized" link state protocol could look like. According to [JLMV01], OSPF can be seen as a "non-optimized" link state protocol, as each node is equally committed to the distribution of topology information around the network. The optimization procedure refers to the choice of special nodes within the network, the multi-point relay nodes. The exchange of topology information throughout the network is restricted to these relay nodes, which are chosen by a voting procedure known as multi-point relay selection mechanism.

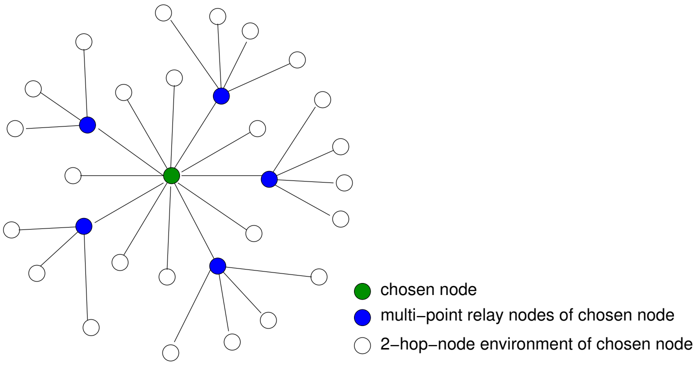


Figure 4.1.: Each node choses distinct multi-point relay nodes to decrease the overhead for flooding the network.

Multi-Point Relays The idea of using multi-point relay stations is related to the development of the HIPERLAN/1 standard [ETS96]. The basic motivation behind relay stations is the reduction of flooding overhead by choosing a subset of distinct forwarding relay stations from the set of all nodes. Each node determines its own set of multi-point relay stations in a way that the set covers all nodes, which are two hops away [JCL$^+$01] as exemplarily shown in Figure 4.1. In addition to that, the set is supposed to be chosen in such a manner that the number of nodes included is minimum. In doing so, a node will only forward a received flooding packet, if it has not received the same packet before and if the node has been chosen as multi-point relay node by the sending node. This mechanism guarantees a huge reduction in terms of signaling overhead, as only a percentage of all nodes is taking part in the flooding process. Nevertheless, due to the two-hop rule, it is still guaranteed that all nodes receive the flooded message.

Topology Control OLSR uses a topology table — also referred to as intra-forwarding database — to determine a possible routing path between two nodes. The topology table is constructed from topology control (TC) messages. These messages are periodically sent out by each multi-point relay node to announce the set of nodes, which have chosen this distinct node as their multi-point relay. Topology control messages are broadcast through the entire network. Clearly, this implies that multi-point relay nodes only are allowed to generate topology control messages. The ensemble of all received topology control messages will help each node to build its topology table. It is understandable that the topology table is sufficient to calculate a potential route between two arbitrary network nodes. This can be accomplished by iteratively constructing a backward chain of multi-point relay nodes, starting from the receiver node back to the sender node.

To sum up, the major characteristic of OLSR consists in the fact that OLSR is a proactive routing protocol, in which the signaling overhead is reduced by the choice of distinct relay stations for the flooding operation. Another aspect to put on record is the protocol operation with two important message types: a HELLO message for the local topology exploration and a topology control message for the distribution of link state information around the network.

Ad Hoc On Demand Distance Vector Routing

Ad hoc On demand Distance Vector Routing (AODV) [PBRD03] is a reactive routing protocol. This means that a routing path will be established, if and only if there is a corresponding traffic demand. While proactive routing protocols will also generate signaling overhead in the absence of any demand traffic, AODV signaling overhead is always directly linked to a certain demand.

Path Establishment Whenever traffic demand occurs within the network, the source node sends out a route request (RREQ) message, which is flooded through the network. Each RREQ message contains a specific, incrementally increased destination sequence number. Any intermediate node has to rebroadcast the RREQ to its other neighboring nodes, if the sequence number of a received message is larger than a cached destination sequence number to the same destination. Given the case that the received sequence number is smaller or equal, this is a sign that the intermediate node potentially knows a valid connection to the destination node. The intermediate node is able to reply instantly to the request with a corresponding route reply (RREP) then. The RREP is sent in a chain of unicast messages back to the initial source node of the RREQ. The RREP moves along the inverse path, which has been generated by the RREQ broadcast message. For that purpose, each node evaluates all information from incoming signaling packets and stores this information in a local routing table. The reverse path for RREP messages can be reconstructed from those tables.

Path Maintenance Although not included in the original protocol specification of AODV [PR99], later publications in [PRDM01], [PRDM00], and the RFC specification [PBRD03] propose a combination of time stamps and route error (RERR) messages to prevent the network from permanent link failures. According to [PBRD03], a RERR message is optionally either unicast or broadcast. A node transmits a RERR message in case a link failure is detected during an active transmission, in case a data packet is mislead and an appropriate route is not known to the node, or in case a RERR message is received from another node. In addition to that, each entry in the routing table is constrained by a life time field. Whenever a route is not used for a predefined period of time, it will automatically expire. In order to keep track of the current link conditions, nodes may optionally use one-hop reply messages in absence of other broadcast packets as for instance RREQ messages. This special form of a RREP message is also denominated as HELLO message.

The on demand philosophy distributes responsibilities locally and requires equal effort from each participant. This makes communication flexible and resistant to failures, which is an important characteristic, especially for networks with high node mobility. One of the big advantages is that — as long a path is stable — there is no additional signaling overhead, as ongoing communication will be used for the link monitoring and path maintenance. However, AODV requires more time than a proac-

Table 4.1.: Classification and comparison of different routing protocols.

type	AODV	OLSR	OSPF
classification	reactive	proactive	proactive
routing concept	flat	multi-point relaying	hierarchical
topology exploration	route request (RREQ) message	HELLO message	HELLO message
update frequency	on demand	periodically	periodically
topology announcement	route reply (RREP) message	topology control message (TC)	link state update message
error recognition	route error (RERR) message	-	error message
topology view	local	global	global
scalability	none	partly (by clusters)	full (by areas)
quality of service metric	difficult proposals available e.g. [SBR04]	possible proposals available e.g. [BMM07]	OSPF quality of service extension

tive approach for the initial establishment of a connection, because only the proactive approach provides ongoing network monitoring and permanently updated routing tables. In addition to that, it is difficult to establish and maintain a global view and common routing practice for the whole network, as each AODV node acts on its own, independently from other nodes.

4.3.3. Summary

The most important attributes and properties of AODV, OLSR, and OSPF as presented in the preceding sections are summarized in Table 4.1. MPLS is not considered in this table, as MPLS is generally acting on top of another interior gateway protocol and thus is not directly comparable. It has been shown that proactive protocols are characterized by a beacon frequency, according to which bidirectional link connections are maintained periodically. The accuracy of this topology exploration mechanism raises with an increasing beacon frequency, but the signaling overhead does accordingly. Therefore, proactive routing protocols yield large signaling overhead even in static scenarios. There is a typical trade-off for that purpose between the accuracy of the topology exploration and the signaling overhead, which proactive protocols are subject to. Obviously, there is no beacon frequency in a reactive protocol. As a consequence, the reactive protocol is free of signaling overhead in the static case. This also implies that ongoing link state information is not available. For this reason, it is difficult for entities like the access point to preserve a consistent overview on permanent network activities in this case.

Standard compliant proactive and reactive routing protocols evaluate the hop count to determine the shortest path as routing path between two nodes. This underlies the primary assumption that every further hop involves delay and represents a potential source of failure. Thus, the shortest path is assumed to be the fastest and the most reliable connection at the same time. Apparently, this assumption is not always true. A good example is given by the cellular multi-hop network itself, as for instance the relay links are considered to be very reliable, while the links to the user terminals are subject to mobility. The probability of a link break is therefore not equally distributed. To reflect the network's properties in a more realistic way, research develops more complex routing metrics (see [SBR04] and [BMM07]) and includes exemplarily link quality or user velocity as measurement parameters. The routing metric also impacts the resource demand of a connection. The better the metric accounts for the topology, the more efficiently a routing path can be established. For this reason, we want to investigate the influence of the routing metric on the resource demand in the following and therefore develop a corresponding graph model first of all.

4.4. Influence of Routing Metrics on Resource Demand

4.4.1. System Model

Assume a network topology to be defined by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, c)$ with nodes $V_i \in \mathcal{V}$, directed edges $\vec{e}_i \in \mathcal{E}$, and edge-dependent capacities $c(\vec{e})$. The network graph consists of a single access point V_A, a set of relay stations $\mathcal{V}_R$, and mobile terminals $\mathcal{V}_M$ with $\{V_A\} \cup \mathcal{V}_R \cup \mathcal{V}_M = \mathcal{V}$. Each node V_i is assigned a geographical position in terms of a set of coordinates $V_i \leftrightarrow (X_{V_i}, Y_{V_i})$. Correspondingly, the Euclidean distance $d(V_1, V_2)$ between two nodes is defined as:

$$d(V_1, V_2) = \sqrt{(X_{V_1} - X_{V_2})^2 + (Y_{V_1} - Y_{V_2})^2}. \tag{4.1}$$

Connectivity The question, whether two nodes are connected by an edge, depends on the distance $d(V_1, V_2)$, the distinct antenna gain factor a, and finally a predefined threshold distance d_{max}. The distance d_{max} is the maximum communication range of a transmission station with omnidirectional antennas at sender and receiver. This value is primarily dependent on the transmission power and the channel conditions. The antenna gain factor is related to hardware properties of the equipment. We therefore distinguish between router-to-router transmissions, for which we assign a^{static}, and mobile-terminal-to-mobile-terminal transmissions, for which a^{mob} is taken into account. Given this basic parameter set, two different link constellations are distinguished:

$\forall V_1, V_2 \in \mathcal{V}$:
$\exists \vec{e} = (V_1, V_2) \Leftrightarrow$

$$\begin{cases} \text{(I) } \vec{e} \in \mathcal{E}^{mob} : & V_1 \in \mathcal{V}_M \quad \wedge \quad V_2 \notin \mathcal{V}_M \quad \wedge \quad \Big(d(V_1, V_2) \leq d_{max} \cdot a^{mob}\Big) \\ \text{(II) } \vec{e} \in \mathcal{E}^{static} : & V_1 \notin \mathcal{V}_M \quad \wedge \quad V_2 \notin \mathcal{V}_M \quad \wedge \quad \Big(d(V_1, V_2) \leq d_{max} \cdot a^{static}\Big). \end{cases}$$

While the access point V_A and the set of relay stations $\mathcal{V}_R$ are fixed, the set of mobile terminals $\mathcal{V}_M$ is subject to mobility. There is an edge between a mobile terminal and

a wireless router, if the mutual distance is smaller than the product of the threshold distance d_{max} with the antenna gain factor. This set of edges is referred to as $\mathcal{E}^{mob}$. Similarly, there is a subset of all edges $\mathcal{E}$ with quasi-static link connections $\mathcal{E}^{static}$. In this case, the links are established between two wireless routers. The physical propagation conditions are comparable, but a higher antenna gain is assumed for this constellation. Inversely, the case of an edge connecting two mobile terminals is infeasible by definition. Let us consider a piecewise linear link adaptation function f_a to define a relationship between the topological properties of a link $\vec{e} = (V_1, V_2)$ and a corresponding physical data rate $r_{phy}(\vec{e}) = r_{phy}(V_1, V_2)$:
$\forall \vec{e} \in \mathcal{E}$:

$$r_{phy}(\vec{e}) := \mathrm{f}_a(d(V_1, V_2), a). \tag{4.2}$$

The physical implications of the channel limit the capacity of a link such that the physical data rate and the edge capacity are equivalent,
$\forall \vec{e} \in \mathcal{E}$:

$$c(\vec{e}) \overset{!}{=} r_{phy}(\vec{e}). \tag{4.3}$$

Resource Efficiency Assume a subgraph $\mathcal{S} \subseteq \mathcal{G}$ with $\mathcal{S} = (\mathcal{V}, \mathcal{E}, g)$ to provide the solution of a unicast routing problem, whereby some data is routed from given source nodes to distinct destination nodes. The routing process requires an edge-dependent resource allocation $g(\vec{e})$. As $\mathcal{S}$ is a subgraph of $\mathcal{G}$, the resource allocation $g(\vec{e})$ is constrained by the edge capacities,
$\forall \vec{e} \in \mathcal{E}$:

$$0 \leq g(\vec{e}) \leq c(\vec{e}). \tag{4.4}$$

We define the ratio between allocated resources $g(\vec{e})$ and maximum available capacity $c(\vec{e})$ as resource efficiency indicator $\nu(\vec{e})$ for the resource consumption on a specific link
$\forall \vec{e} \in \mathcal{E}$:

$$\nu(\vec{e}) := \frac{g(\vec{e})}{c(\vec{e})}. \tag{4.5}$$

The efficiency indicator is bounded by $\nu(\vec{e}) \in [0.0; 1.0]$. This relationship directly follows from the restriction in Equation (4.4). The indicator decreases with raising efficiency. The most efficient realization does not require any resources at all, i.e. $g(\vec{e}) = 0$. The most inefficient realization requires full capacity $g(\vec{e}) = c(\vec{e})$ and $\nu = 1.0$. The summation of all individual performance indicators $\nu(\vec{e})$ to an aggregated value $\tilde{\nu}$ provides information about the resource efficiency of the overall routing solution:

$$\tilde{\nu} = \sum_{\vec{e} \,\in\, \mathcal{E}} \nu(\vec{e}) = \sum_{\vec{e} \,\in\, \mathcal{E}} \frac{g(\vec{e})}{c(\vec{e})}. \tag{4.6}$$

Let us explain this relationship with a short example. Assume a set of data connections with an envelope data rate of $r = 9$ MBit/s to be alternatively realized over a two-hop connection with a capacity of $c(\vec{e_1})$ = $c(\vec{e_2})$ = 54 MBit/s each or a single-hop connection with $c(\vec{e_3})$ = 18 MBit/s. A mutual comparison of $\nu_{th} = \frac{g(\vec{e_1})}{c(\vec{e_1})} + \frac{g(\vec{e_2})}{c(\vec{e_2})} = 2 \cdot \frac{9\,\text{MBit/s}}{54\,\text{MBit/s}} = \frac{1}{3}$ and $\nu_{sh} = \frac{g(\vec{e_3})}{c(\vec{e_3})} = \frac{9\,\text{MBit/s}}{18\,\text{MBit/s}} = \frac{1}{2}$ reveals that $\nu_{th} < \nu_{sh}$. Therefore, the two-hop connection is considered to be more efficient than the single-hop connection in this case. $\tilde{\nu}$ is an important indicator to compare the different routing realizations with respect to their individual resource consumption and will be used for evaluation purpose.

Mobility Due to mobility, $\mathcal{G}$ and its physical properties change over time. We consider a interval of T seconds. While the set of nodes $\mathcal{V}$ and the set of edges $\mathcal{E}^{static}$ remain unchanged, the set of dynamic edges $\mathcal{E}^{mob}$ is time-dependent. We can describe the topological changes by a series of time-discrete snapshot graphs $\mathcal{G}^t = (\mathcal{V}, \mathcal{E}^t, c^t)$, whereby each graph is an instantaneous view of the topology at time t. The corresponding time-dependent routing solution is again described by a subgraph $\mathcal{S}^t \subseteq \mathcal{G}^t$. Consequently, we can also assign time-dependent values to the resource allocation $g(\vec{e}, t)$ and the capacity $c(\vec{e}, t)$. The solution of the overall problem is thus given by a comprehensive survey of all graphs $\mathcal{G}^t$. For the investigation of the resource efficiency, the summation of all individual edge-dependent indicators is consequently extended to edge- and time-dependent indicators $\nu(\vec{e}, t)$. The resulting average value is given by

$$\begin{aligned}\tilde{\nu}_{avg} &= \frac{1}{T}\int_0^T \sum_{\vec{e}\,\in\,\mathcal{E}^t} \nu(\vec{e}, t)dt = \frac{1}{T}\int_0^T \sum_{\vec{e}\,\in\,\mathcal{E}^t} \frac{g(\vec{e}, t)}{c(\vec{e}, t)}dt \\ &\approx \frac{1}{T}\sum_{t\,\in\,\mathcal{T}}\sum_{\vec{e}\,\in\,\mathcal{E}^t} \frac{g(\vec{e}, t)}{c(\vec{e}, t)} \cdot \Delta t.\end{aligned} \tag{4.7}$$

As it is impossible to solve the routing problem continuously, the integral can be approximated by a summation of time-discrete solutions. For that purpose, the overall time T is divided into a set of equally spaced time intervals Δt. We consider snapshot views of the graph $\mathcal{G}^t$ at the beginning of each interval. The set of these moments is given by $\mathcal{T}$. The average resource efficiency of the time-dependent overall routing problem is thus approximated by the term in Equation (4.7).

4.4.2. Optimum Solution to the Routing Problem

Flow Approach

Problem Formulation We want to develop a mathematical model to determine an optimum solution to the routing problem. The routing task includes the problem, how to assure a connection between source node and destination node for a given set of traffic flows. For that purpose, we consider a set of data flows $\mathcal{F}$ with $M = |\mathcal{F}|$. Each flow $f^m \in \mathcal{F}$ is heading from a source node V_s^m to a destination node V_d^m with a fixed

rate r^m. The set of source nodes is referred to as $\mathcal{V}_s^m$. The overall flow is composed of individual edge-dependent flow elements $f^m(\vec{e})$. Any flow element is bound to a directed edge and represents the transport of data such that negative flow elements are physically impossible:
$\forall m \in \{1, \dots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$f^m(\vec{e}) \geq 0. \tag{4.8}$$

Typically, the network topology allows different solutions to the routing problem. The routing problem itself is described by a set of linear equations, which we refer to as routing constraints. Apart, an operator-defined optimization criterion defines a distinct routing objective. In doing so, the problem itself can be formulated as LP

minimize routing objective
subject to
routing constraints.

As discussed previously, the scenario with mobile terminals is described by a series of time-discrete snapshot graphs $\mathcal{G}^t$. The routing problem is solved individually for each graph. The corresponding solution of the routing problem is presented by a subgraph $\mathcal{S}^t \subseteq \mathcal{G}^t$. For the clearness of presentation, we restrict our analysis to a single graph $\mathcal{G}$ only and demonstrate a potential solution approach. The overall solution can be determined by the repeated application of this approach to all graphs $\mathcal{G}^t$.

Routing Constraints Under the assumption that each flow f^m has to be routed, we formulate a linear problem description based on a mathematical concept, which is commonly known as flow approach. The flow approach is based on the assumption that the equality of incoming and outgoing flows must hold at intermediate nodes. Flows are created at source nodes and finally disappear at the corresponding destination nodes as sink. Starting at the source node V_s^m, there must be a net outflow of rate r^m such that the difference between incoming and outgoing flows equals the rate,
$\forall m \in \{1, \dots, M\}, \forall V_s^m \in \mathcal{V}_s^m$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V_s^m)} f^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V_s^m)} f^m(\vec{e}) = r^m. \tag{4.9}$$

For all intermediate nodes $V \in \mathcal{V} \backslash \{V_s^m, V_d^m\}$, which are neither source nor destination node V_d^m, the equilibrium of flows must be valid. According to the principle of flow preservation, the sum of outgoing flows at each intermediate node must be equal to the sum of incoming flows. In doing so, the amount of a distinct flow can neither be artificially increased nor misleadingly decreased,
$\forall m \in \{1, \dots, M\}, \forall V \in \mathcal{V} \backslash \{V_s^m, V_d^m\}$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} f^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V)} f^m(\vec{e}) = 0. \tag{4.10}$$

Accordingly, flows can only be created at the corresponding source. Subsequently, flows are terminated at the destination terminal acting as sink. Additionally, we introduce a binary auxiliary variable $f_h^m(\vec{e}) \in \{0, 1\}$ with
$\forall m \in \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$f_h^m(\vec{e}) \cdot C \geq f^m(\vec{e}), \tag{4.11}$$

where C is a large integer constant. Consequently, $f_h^m(\vec{e}) = 1$ for all $f^m(\vec{e}) \neq 0$ and $f_h^m(\vec{e})$ is indifferent for $f^m(\vec{e}) = 0$. We want to concentrate on single-path routing such that we restrict the number of binary outflows $f_h^m(\vec{e})$ to one flow per router at maximum,
$\forall m \in \{1, \ldots, M\}, \forall V \in \mathcal{V}_{\mathcal{R}} \cup \{V_A\}$:

$$\sum_{\vec{e} \,\in\, \mathcal{E}^{(out)}(V)} f_h^m(\vec{e}) \leq 1. \tag{4.12}$$

As the set of mobile terminals $\mathcal{V}_M$ is not intended to act as wireless routers, outgoing flows from any mobile terminal are generally prevented. The sole exception is given by cases, in which a distinct mobile terminal V_M acts as source node for a flow f^m,
$\forall m \in \{1, \ldots, M\}, \forall V_M \in \mathcal{V}_{\mathcal{M}} \backslash \{V_s^m\}$:

$$\sum_{\vec{e} \,\in\, \mathcal{E}^{(out)}(V_M)} f_h^m(\vec{e}) = 0. \tag{4.13}$$

All the different flow elements on an edge sum up to the edge-dependent resource allocation $g(\vec{e})$ as introduced in the previous section,
$\forall \vec{e} \in \mathcal{E}$:

$$g(\vec{e}) = \sum_{m=1}^{M} f^m(\vec{e}). \tag{4.14}$$

Equation (4.9) to Equation (4.14) and the additional capacity restriction Equation (4.4) represent a set of linear equations, which are referred to as routing constraints. The given set of equations fully describes the routing approach. Potential resource allocation vectors fulfilling the set of equations are valid solutions to the routing problem.

Objective Function

A metric to measure the costs for the operator typically acts as objective function for a distinct routing operation. Although most likely more than one solution to a given routing problem exists, not all solutions are equally appealing to the network operator. The operator's main objective is therefore put into a cost function. This function can take into account for instance the hop count, the delay of a packet transmission, the available physical data rate, processing delays, or the load status of the current network configuration. As we have seen, any route assignment results in a corresponding resource allocation, which can be measured by this cost function. By minimizing the distinct objective function, the route assignment follows the operator's strategy and satisfies the defined optimization objective in the best feasible way. For that purpose, let us take a look at possible routing strategies.

Minimum Hop Strategy Historically, simple metrics like the hop count originate from the wired network domain, where link topologies are hardware-defined by fixed wires. As processing at intermediate nodes is in most cases more critical than the wired transmission, the binary accounting of a link connection between two devices as one hop is a very intuitive issue. Under the assumption that the routing process is performed on the network layer, this process is also logically separated from physical layer information, following a strict layering principle. For that purpose, the choice of a minimum hop metric Ω_{hop} is one reasonable of the very limited set of options on the network layer. Ω_{hop} is a well-known metric, which is also applied in ad hoc network routing protocols, for instance standard-compliant AODV. In wireless ad hoc networks, links are generally unstable such that the risk of a link break increases with each additional hop. In using the minimum hop metric, the risk of a link failure is minimized in that respect. As only the last hop is subject to mobility in cellular multi-hop networks, this advantage is not accounted for. Usually, minimizing the hop count also minimizes the delay, although this is not generally valid. Thus, chances of a routing path being stable and reliable are increased — under the assumption that no knowledge about user velocity and user-specific propagation characteristics are available — when applying the minimum hop metric. The corresponding objective function is

$$\Omega_{hop} = \sum_{m=1}^{M} \sum_{\vec{e} \in \mathcal{E}} \operatorname{sgn}\{f^m(\vec{e})\}, \tag{4.15}$$

whereby sgn$\{\cdot\}$ represents the signum function. The signum function can be avoided by the auxiliary variable $f_h^m(\vec{e})$ such that the given minimization goal can be therefore alternatively formulated by

$$\Omega_{hop} = \sum_{m=1}^{M} \sum_{\vec{e} \in \mathcal{E}} f_h^m(\vec{e}). \tag{4.16}$$

The summation accounts for all binary flow elements over all edges of the network for all individual flows. Due to the linearity of the metric, the overall sum of hops is minimum, if all individual flows are routed with a minimum number of hops.

Minimum Resource Strategy The second metric to be introduced is a metric Ω_{res}, which considers the required transmission resources. The resource indicator $\nu(\vec{e})$ has been introduced as an efficiency measure for the resources spent on a link. The summation of all individual measures $\nu(\vec{e})$ allows a statement on the overall efficiency of a routing solution:

$$\Omega_{res} = \sum_{\vec{e} \in \mathcal{E}} \nu(\vec{e}) = \sum_{\vec{e} \in \mathcal{E}} \frac{g(\vec{e})}{c(\vec{e})}. \tag{4.17}$$

While a small link capacity will lead to large individual summands, an additional hop will increase the overall number of summands in the resource indicator sum. A minimization of the given metric will therefore result in the most resource-efficient transmission.

Best Connection Strategy Another metric Ω_{con} guarantees that a mobile terminal will establish a connection to the wireless router with best signal strength. The important difference between the previously presented metric Ω_{res} and Ω_{con} is the property that Ω_{con} takes only the last hop into account, while Ω_{res} evaluates all links on the path between access point and mobile terminal. The corresponding mathematical formulation of Ω_{con} is given by
$\mathcal{E}(\mathcal{V}_M) = \mathcal{E}^{(in)}(\mathcal{V}_M) \cup \mathcal{E}^{(out)}(\mathcal{V}_M)$:

$$\Omega_{con} = \sum_{\vec{e} \in \mathcal{E}(\mathcal{V}_M)} \frac{g(\vec{e})}{c(\vec{e})}. \tag{4.18}$$

Please recall that the idea behind Ω_{con} is close to the practice in today's WLANs. Typically, users try to connect to the access point with best signal strength, which is a reasonable approach from the terminal's perspective without further knowledge.

Balancing Strategy The last proposal is a metric, which balances the resource allocation of each wireless router:

$$\Omega_{bal} = \max_{V \in \mathcal{V}_R \cup \mathcal{V}_A} \sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} g(\vec{e}). \tag{4.19}$$

For that purpose, the maximum resource allocation $\max(\cdot)$ on the sum of outgoing edges is determined at each router among the set of all wireless routers $\mathcal{V}_R \cup \mathcal{V}_A$. This metric is mainly based on the assumption that decreasing the maximum resource allocation per router helps to avoid bottlenecks and will alleviate the spatial reuse of frequency resources.

Routing MILP Let us sum up the presented routing problem. The overall problem formulation includes

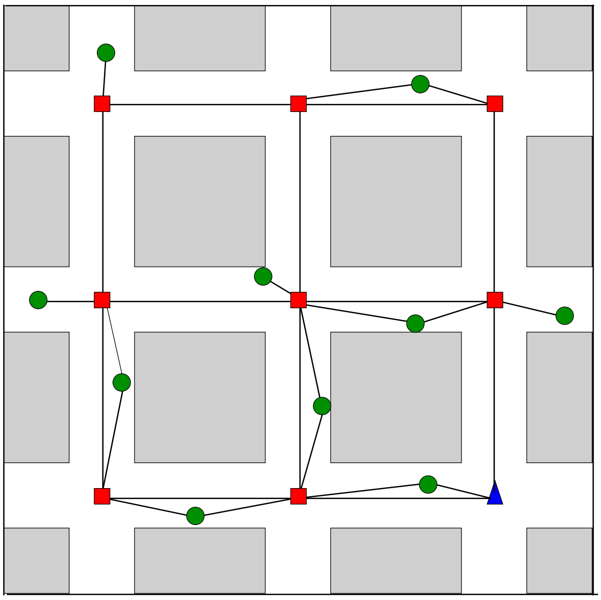


Figure 4.2.: Multi-hop scenario with single access point, 8 relay stations, and — for illustration purpose — only 10 mobile terminals.

$$
\begin{array}{ll}
\textit{minimize} & \Omega = \mathrm{f}(f^m(\vec{e}), f_h^m(\vec{e})) \\
\textit{subject to} & \\
\forall m \in \{1, \ldots, M\}, \forall V_s^m \in \mathcal{V}_s^m : & \displaystyle\sum_{\vec{e} \in \mathcal{E}^{(out)}(V_s^m)} f^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V_s^m)} f^m(\vec{e}) = r^m \\
\forall m \in \{1, \ldots, M\}, \forall V \in \mathcal{V} \backslash \{V_s^m, V_d^m\} : & \displaystyle\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} f^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V)} f^m(\vec{e}) = 0 \\
\forall m \in \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E} : & f_h^m(\vec{e}) \cdot C \geq f^m(\vec{e}) \\
\forall m \in \{1, \ldots, M\}, \forall V \in \mathcal{V}_{\mathcal{R}} \cup \{V_A\} : & \displaystyle\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} f_h^m(\vec{e}) \leq 1 \\
\forall m \in \{1, \ldots, M\}, \forall V_M \in \mathcal{V}_{\mathcal{M}} \backslash \{V_s^m\} : & \displaystyle\sum_{\vec{e} \in \mathcal{E}^{(out)}(V_M)} f_h^m(\vec{e}) = 0 \\
\forall \vec{e} \in \mathcal{E} : & \displaystyle\sum_{m=1}^{M} f^m(\vec{e}) \leq c(\vec{e}) \\
 & f^m(\vec{e}) \in \mathbb{R}_0^+, f_h^m(\vec{e}) \in \{0, 1\},
\end{array}
$$

where the objective function Ω is defined by one of the four presented metrics. A solution of this MILP leads to a valid routing solution within the multi-hop cell. We will examine and compare the different metrics in the following section.

4.4.3. Simulative Study

Imagine a simulation scenario as illustrated in Figure 4.2. The setup includes one access point $|\mathcal{V}_A| = 1$, relay stations $|\mathcal{V}_R| = 8$, and mobile terminals $|\mathcal{V}_M| = 30$. Access point and relay stations are located in the middle of corresponding street crossings, while mobile terminals move around randomly in the street tunnels. A mobile node cannot enter the building area, its movement radius is restricted to the streets. Each building has a block size of 120 m x 120 m, the street width is 60 m. Consequently, the mutual distance between two routers is 180 m. Terminal nodes are uniformly distributed on the simulation area in the start configuration. An initial placement within the buildings is not possible. The terminals move according to the random direction mobility model as specified in Section A.1. Each terminal is assigned a triple vector of movement angle, speed, and duration. The movement duration is uniformly distributed with a maximum duration t_{max} = 30 s, the movement speed is uniformly distributed with a maximum user speed of $v_{max} = 10$ m/s. Five runs with independent terminal movements and a duration of $T = 250$ s each have been performed to get statistically meaningful results. The evaluation results represent average values of these scenarios. The snapshot resolution is chosen to be $\Delta t = 10$ ms. The default transmission range of the wireless routers is $R = 120$ m and can be adjusted by varying the necessary received signal threshold. The maximum overlapping range in the coverage area of two neighboring routers is thus 60 m on the connecting line of those routers. The antenna gains for relay links between wireless routers are chosen in a way that these links are operated at a maximum data rate of 54 MBit/s in the default configuration. The link adaptation function is equivalent to Equation (A.6). Contrarily, mobile devices are small low-cost terminals and therefore supposed to have no antenna gain, i.e. $a^{mob} = 1$ on links between routers and mobile terminals. A constant bit rate connection with $r^m = 16$ kbit/s is established from the access point to each of the 30 mobile terminals.

The absolute amount of required resources of a certain metric is not meaningful, as this value can be scaled arbitrarily with different traffic loads. The interesting issue is the relative comparison of the resource requirements for different metrics. The minimum resource strategy is therefore taken as 100% benchmark. The values of other strategies are normalized accordingly.

Resource Efficiency

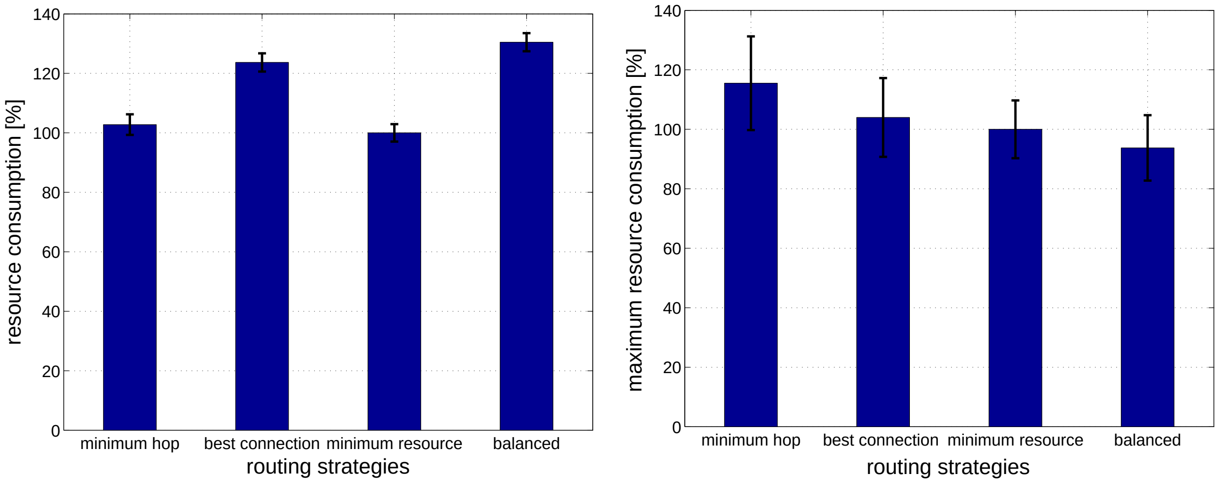
The efficient use of available transmission resources is a major issue in cellular networks. Imagine the case of a provider, who has spent billions into a license for a frequency band. It is clearly understandable that the spectral efficiency of the cellular network will influence the profitability of the investment. As soon a network operation mode with high efficiency is applied, the network's operation will require less resources to supply the same amount of subscribers such that the overall investment costs will be reduced.

The spectral efficiency on a single link is mainly related to the physical layer conditions. The spectral efficiency of a multi-hop network is also dependent on the routing strategy. This is due to the fact that different strategies result in different routes. As the links' capacities are variable, some strategies will require more resources than other strategies. The following simulations will illustrate this relationship. For that purpose, the normalized value of the aggregated resource consumption $\tilde{\nu}$ is calculated for different routing strategies.

Minimizing Resources As expected, the simulation results of Figure 4.3(a) reveal that the minimum resource strategy uses the least amount of resources and thus defines the 100% benchmark in the chart. More interestingly, it is shown that the minimum hop strategy has comparable results for this scenario. This can be explained by the limitation of the transmission range, due to which long-distance connections with very small PHY modes are infeasible. These long-distance links are typical examples, where a minimum hop strategy could potentially rely on a very inefficient one-hop connection instead of a better two-hop connection. As these unfavorable constellations are prevented by the physical layer, the routing unit is forced to establish a new route as consequence of a physical link break. This link break is independent on the routing strategy. Therefore, there is practically no difference between both strategies in this scenario. The simulations also show that the best connection strategy requires about 25% more resources than the comparable minimum resource strategy. This is a very interesting result. Please keep in mind that a decision according to the best connection strategy is taken for instance by today's WLAN users, which simply try to connect to the next hot spot access point with best signal quality. Even worse behaves the load balancing strategy, where about 30% more resources are needed than for the minimum resource strategy.

Load Balancing The advantages of the balancing strategy are illustrated in Figure 4.3(b). This result indicates that the minimum hop strategy tends to create a bottleneck at the access point, as all feasible devices are connected in a one-hop connection to the access point. The maximum resource consumption per router is decreased by the balancing strategy by about 20% compared to the minimum hop strategy and about 10% to the minimum resource strategy. However, this decrease in maximum load is ongoing with an increase in overall load as shown in the previous figure. Nevertheless, any decrease in the maximum amount of required resources may facilitate the spatial reuse of frequency resources. We currently assume that resources cannot be reused within the same cell such that potential savings due to spatial frequency reuse are not accounted. From this perspective, there is no real advantage of intra-cell load balancing such that we will renounce on further investigations with the balancing strategy. Interestingly, the minimum resource strategy and the best connection strategy have a similar maximum resource consumption.

Summary These initial results indicate that the very easy minimum hop strategy is indeed an efficient approach to operate a multi-hop cell, which is comparable to the minimum resource strategy with optimum resource allocation, but leads to peak loads around the access point. It is also shown that the probably most obvious approach from the user's perspective, the best connection strategy increases the overall resources significantly. Therefore, the application of this approach in a cellular multi-hop network is not recommendable at all. Balancing the user load will help to reduce



(a) Illustration of the overall resource consumption for the different routing strategies.

(b) Illustration of the maximum resource consumption per router for the different routing strategies.

Figure 4.3.: The figures indicate that the choice of the routing strategies has serious impact on the overall resource consumption and the maximum resource consumption within the multi-hop cell.

the maximum load per router by about 10%. However, an increased overall resource requirement has to be accepted in return for this peak load reduction.

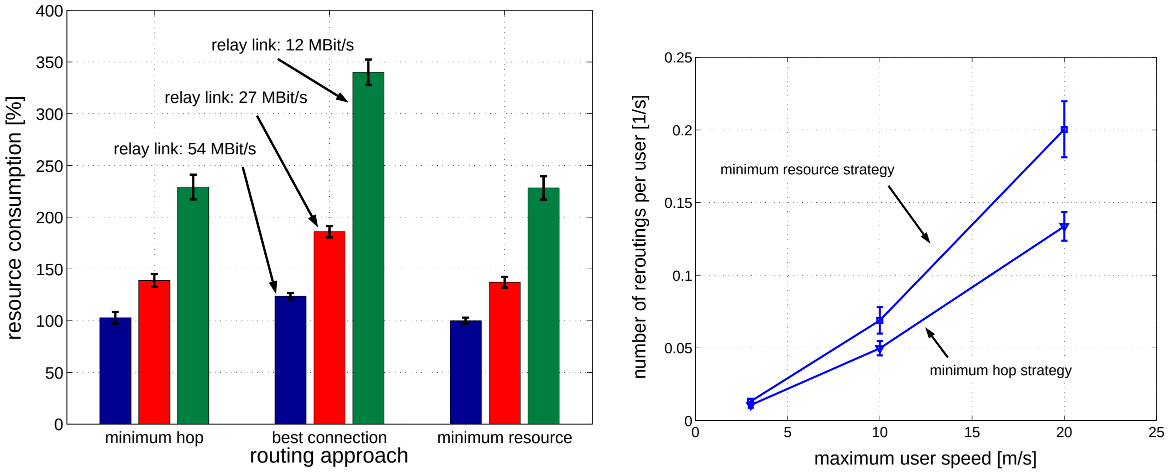
Dependency on Relay Link

In the deployment and planning of a cellular multi-hop network, the links between wireless routers are a technical key impact factor. The density of the infrastructure network as well as the choice and the quality of the technical equipment influence the link capacity of these wireless relay links. In order to get a better understanding of the influence of these wireless backbone links, the relationship between relay link capacities and the resulting overall resource consumption in the network is investigated in the following. This analysis is motivated by the desire to emphasize the importance of the topological design of the infrastructure network for the overall performance.

Antenna Gain In practice, the physical data rate of a relay link depends on the chosen combination of forward error correction scheme and applicable PHY mode. The optimum PHY mode is primarily dependent on the SINR. External impact factors as for example user terminal velocity and the propagation characteristics of the environment may additionally influence the PHY mode choice. Furthermore, the properties of the technical equipment itself impact the feasible data rate for fixed link conditions. A very substantial parameter in this context is the antenna gain, a hardware-specific characteristic of the wireless devices, which results from the number of physical antenna elements. If we consider fixed and operator-owned access and relaying equipment, one can imagine hardware with more than a simple omni-directional antenna element and thus higher antenna gains. Ultimately, this is also a question of investment costs, i.e. a higher investment will impact the technical quality of the individual components. Exemplarily, assume the antenna gain to be variable. In practice, this can be achieved by antenna equipment with variable sensitivity. Consequently, the equipment enables the relay links to be operated at different physical data rates, although transmission distance, transmission power, and the topology remain fixed. The positioning of single antenna elements and the subsequent transmission angle may provoke a similar change in the available link quality.

Relay Link as Design Parameter Figure 4.4(a) shows, what will happen, if the physically feasible data rate of the relay links is — ceteris paribus — decreased due to some extrinsic impact factor, e.g., equipment quality or positioning. In doing so, the physically feasible data rates for the relay links vary between 12 MBit/s and 54 MBit/s, while link conditions to mobile terminals stay untouched per definition. Figure 4.4(a) illustrates that the overall resource consumption increases dramatically with decreasing relay link capacity, independently on the applied routing strategy. Nevertheless, the best connection strategy leads to even worse results than the other strategies. While the increase in overall resource consumption for large relay link data rates (54 MBit/s) is about 20%, it is around 50% for small relay link data rates (12 MBit/s). The reason for this behavior is obvious: The best connection strategy is totally unaware of the relay links. These links are not included in the optimization process, but only the last hop links to mobile user terminals are considered. This is a very important observation. The worse the rate on the relay link, the more important is a consideration of the relay link topology in the optimization process. This explains the bad performance of the best connection strategy. Any disregard of the path-dependent overall resource consumption can increase the required overall resources by 50%, if the link conditions on the relay link are far from optimum as seen in the case study with $r_{phy}(\vec{e}) = 12$ MBit/s on the relay links. As an extreme case, imagine a wireless infrastructure network with unlimited capacity ($r_{phy}(\vec{e}) \rightarrow \infty$). This is by the way comparable to a replacement of all relay stations by access points. In this case, the relay link does not matter at all for the optimization process, as the required resources tend to zero. A simple last-hop optimization with the best connection strategy is the preferable strategy to apply then. For any intermediate scenario with finite relay link capacities, the ratio in resource consumption between last hop links and relay links is important. Both links should be equally considered in the optimization process, if the relay link capacity is comparable to the last hop link capacity.

Summary There are two major results to point out: On the one hand, the above presented results have shown that a simple last hop optimization according to the best connection strategy can have disastrous consequences on the overall resource consumption. On the other hand, the impact of the wireless relay links becomes less important with increasing relay link capacity.



(a) Relay link quality versus resource consumption. (b) User velocity versus rerouting events.

Figure 4.4.: Resource consumption and rerouting events.

Overlapping Area of Coverage Ranges

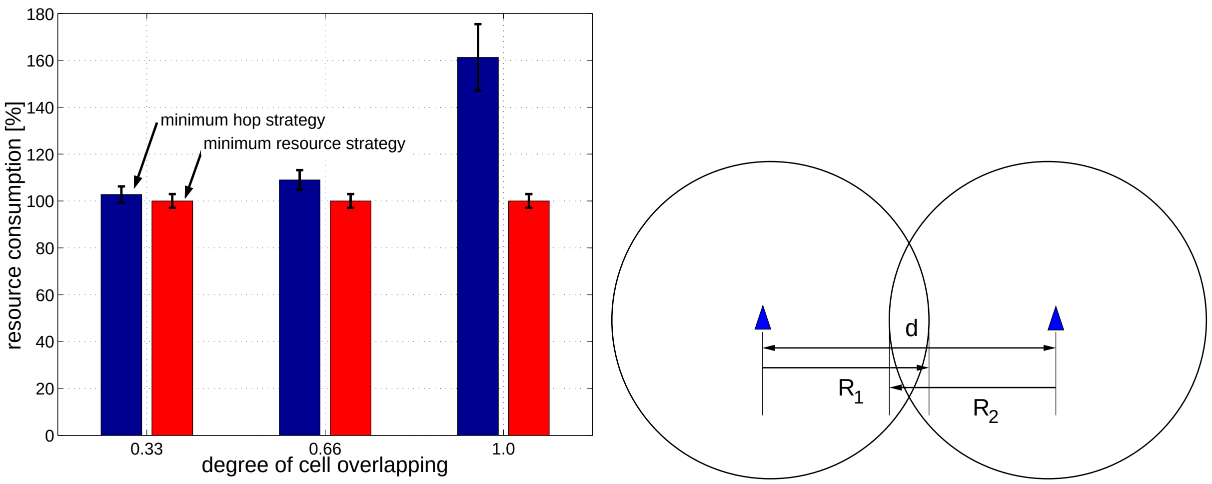
A central aspect in the design of wireless networks is the size of the overlapping area between the transmission ranges of two neighboring routers. Adaptive power assignment at the transmitter or a variable detection threshold sensitivity at the receiver will have direct impact on the maximum transmission range and consequently on the overlapping area between neighboring routers. The following simulation results have been achieved by varying the receiver threshold parameter in a way that the transmission range is ceteris paribus increased. We have omitted the best connection strategy, as this strategy is indifferent to the extension of the coverage ranges.

The coverage overlapping factor ψ is defined as a relationship between the transmission radius and the distance between two routers:

$$\psi = \frac{R_1 + R_2 - d}{d}. \tag{4.20}$$

The corresponding relationship is also illustrated in Figure 4.5(b). The amount of overlapping area defines the region, where a user terminal can choose and handover between different wireless routers. The overlapping factor is therefore an important parameter for handover considerations. It also influences the physical link topology and thus the options and the behavior of routing strategies, as the overall connectivity of the network is increasing with raising the overlapping area.

The effect of an increased overlapping factor between different cells is shown in Figure 4.5(a). While the minimum hop strategy and the minimum resource strategy behave similarly for a small overlapping factor ($\psi = 0.33$), the difference in resource consumption is increasing with raising overlapping degree. Given an overlapping degree of $\psi = 1.0$, the minimum hop strategy leads to an increased resource consumption of about 60% more resources, although the resource consumption for the minimum resource strategy remains unchanged. This difference can be explained



(a) Effects of increasing the overlapping area between neighboring stations.

(b) Illustration of the coverage overlapping factor.

Figure 4.5.: The mutual overlapping coverage area between neighboring routers is an important extrinsic impact factor on the efficiency of the routing strategy.

easily: Provided a small overlapping area, the number of potential routing paths is small such that the chosen routing path is primarily imposed by physical link conditions on the routing strategy. However, with raising overlapping factor, the margin of the routing protocol increases with the number of possible routing paths. While the minimum resource strategy still remains perfectly adjusted to the link conditions, the minimum hop strategy tends to rely on very inefficient long-distance links and thus wastes valuable resources. Recalling the analytical analysis of [Emm05b], a high overlapping factor is essential for networks with small coverage ranges to guarantee a seamless handover between different routers (see Chapter 2). The minimum hop strategy is beyond doubt affiliated with an inefficient handling of resources then and should be avoided from that perspective.

Summary The mutual dependency of the cell topology and the efficiency of the routing strategy is pointed out by the presented results: If the network is planned in a way that the overlapping coverage region is very small, an intelligent routing strategy will have minor impact, as routing will be imposed and overruled by the physical link conditions in most cases. If the network is constructed with huge mutual overlapping areas between the neighboring routers, this constellation will alleviate the handover process. The results imply that a suitable intelligent routing strategy is necessary to save valuable transmission resources then.

Stability of Routing Path

Adapting the routing path perfectly according to the physical link conditions helps to preserve transmission resources. However, one aspect of this permanent reconfiguration should not be omitted: Frequent reroutings may cause additional overhead for signaling, association, and the related exchange of services parameters. Additionally, any reconfiguration is a potential source for errors and failures such that link or route breaks are alleviated. The stability of the routing path is therefore a very critical issue. Two major sources, which cause a link connection to be rerouted, can be determined:

- *mobility*: As the users are considered to be mobile, there will be link breaks due to the physical limitations of the coverage of a wireless router and due to the natural characteristics of the wireless channel. If the mobile terminal is moving from one to another router, the routing protocol is forced to reroute the data connection accordingly, as a connection break-down is unavoidable otherwise. Dynamics due to user mobility cannot be influenced by the choice of the routing protocol, as all routing protocols are subject to these constraints in a similar way.
- *resource allocation*: Nevertheless, a rerouting of a data connection within the network may also be reasonable apart from mobility and physical impacts. Dynamics are also caused by changes of the traffic pattern within the coverage range of a router. Two different situations have to be distinguished again, namely the case, where rerouting is unavoidable, and the case, where rerouting is desirable for optimization purpose. For any system operating below its capacity, rerouting is only desirable, as it might save resources due to the routing process or lead to another goal, for example a higher feasible data rate. This dynamic is a direct consequence of the design and choice of the routing strategy. Some routing strategies result in a less efficient distribution of resources, but also cause less reroutings in return. Other strategies are very resource-efficient, but related to frequent rerouting processes. There is a trade-off between the optimum adaptation to the physical link conditions and the perfect matching of a routing goal on the one hand and stability of the routing path on the other hand.

Every rerouting process does not only mean additional signal overhead caused by the handover procedure itself, but permanent rerouting may also lead to instabilities in the network. The worst case is defined by oscillations, where traffic is shifted back and forth between two routers, like a bouncing ball. Typically, a hysteresis curve is applied to prevent such oscillations. A hysteresis curve for a handover process defines a considerably high threshold level to be surpassed, before the handover is triggered. The immediate handover back becomes thus very unlikely. The numerically evaluated routing strategies do not include any hysteresis curve on purpose. Only thus it is possible to determine the optimum routing path and compare the different strategies directly.

Clearly, it is interesting at this point to evaluate the relationship between the routing strategy and corresponding number of intra-cell handover processes, especially as there is no hysteresis curve included. Figure 4.4(b) illustrates the number of rerouting processes in dependence on the routing strategy. It is shown that the pursuit of the minimum hop strategy requires less reroutings than the minimum resource strategy. The given figure indicates that the difference in rerouting processes depends on the user velocity. While there is practically no difference for the scenario with low user velocity, the minimum resource strategy causes about 50% more reroutings for a corresponding scenario with higher average user velocity. This can be explained by the fact that the minimum hop strategy is less sensitive to link conditions than the minimum resource strategy. Changes of a link condition from one PHY mode to another will not influence the hop count in any respect, but will affect the associated resource allocation. The optimum route adaptation is paid by the price of reroutings for this reason. One possible consequence of this result could be a separation of user groups according to their speed. The optimum resource allocation is only performed for the group of users with low speed, for instance users sitting in a cafe or walking users in a pedestrian zone then. In contrast, a very stable routing configuration, e.g., the minimum hop strategy, could be applied to users with higher mobility, e.g., persons in cars.

Summary As a consequence to the given result, the minimum resource strategy can be pursuit without any impact on the network stability for user terminals with low velocity. In contrast, the minimum resource strategy has to be applied carefully in scenarios with higher average speed, as frequent reroutings are a logical consequence of this strategy.

4.5. Efficient Topology Exploration within a Multi-Hop Cell

4.5.1. Topology Exploration Protocol WOSPF

Design Aspects

We have seen that the establishment of routing paths within a multi-hop cell is a central element for the successful operation of a cellular multi-hop network. To enable this operation, the following open issues with respect to intra-cell routing protocol have to be answered:

- How should an intra-cell routing approach look like, which combines resource efficiency in terms of signaling overhead and optimum route adaptation?
- How can a potential approach be included in the protocol stack of today's wireless communication devices?

We want to investigate these questions and primarily address the issue, whether to use a proactive or reactive routing solution.

Infrastructure and User Subnetwork The preceding studies have illustrated the hybrid structure of the cellular multi-hop network topology. There are static devices in the network, namely the access points and relay stations, and dynamic devices like the mobile terminals. While the static part recalls principles known from fixed

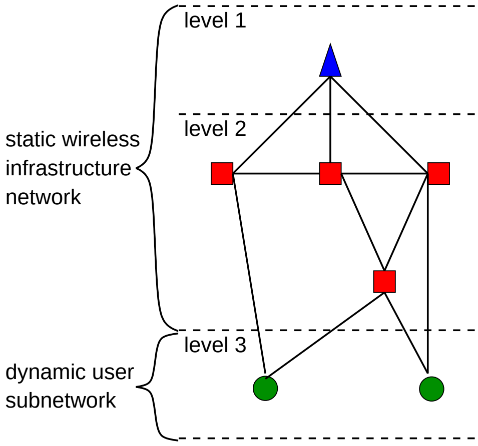


Figure 4.6.: The multi-hop cell can be divided hierarchically into a static infrastructure network and a dynamic user subnetwork.

networks, the dynamic part carries characteristics similar to ad hoc networks. A reasonable protocol solution has to deal with both, the fixed and ad hoc character of the network, at the same time. Considering the cellular multi-hop network as hierarchical network architecture, it can be divided into two parts as shown in Figure 4.6: The upper part includes those network entities, which face almost time invariant transmission conditions, i.e. the links between wireless routers in the fixed infrastructure part of the multi-hop cell. The network formed by these stations is considered to be static at least on a minute-based time scale, as natural fading effects and other dynamic channel characteristics are not accounted for. This part is definitely not subject to any mobility and establishes the upper or infrastructure subnetwork of the wireless multi-hop network. The entities of the lower subnetwork, the mobile user terminals, cannot guarantee any transmission conditions. Hence, the lower subnetwork is also referred to as the dynamic user subnetwork.

Proactive Routing versus Reactive Routing It is possible to apply both, a proactive as well as a reactive intra-cell routing approach. Even hybrid combinations and in-between solutions are possible and also reasonable. Our analysis has shown that a proactive protocol will be the best choice for the operation of a multi-hop cell. Let us name two major reasons that have driven the decision to employ a proactive routing protocol instead of a reactive approach:

- *control aspect:* As the denomination "reactive" implies, an extrinsic action causes the reactive protocol to start its operation. This means being in a tight spot and thus loosing some control. On the contrary, being proactive means acting up to date, even ahead of time. A proactive routing protocol therefore allows overview about the link connectivity and control of the logical connectivity within the network. Measurement data is collected periodically and can be used for the customers benefit, whenever it is desired. This control aspect is a very important part of the philosophy of cellular networks, where a commercial network operator is in charge of the network. The operator wants to provide service guarantees and therefore needs to have all-time control of the network. Employing a fully reactive protocol such as AODV in ad hoc networks or only reactive components as the Hybrid Wireless Mesh Protocol [Bar06] in mesh networks means giving up some of the central control power to the client device itself. This opens the door to potential misuse and restricts the administrative options of the operator. Additionally, if we want to provide an intra-cell routing solution, which is compliant to the existing LTE standard, it is hard to imagine a reactive routing approach to meet these requirements. A proactive routing protocol fits obviously better into the concept and need of a cellular network.

- *resource efficiency:* We have already seen that advanced routing metrics will help to enhance the network's capacity. Intelligent routing metrics make it possible to determine optimized and resource-saving routing paths. These metrics also allow us to find alternative routing paths in case of capacity bottlenecks. The individual optimization of distinct routing paths is therefore embedded into the overall optimization process. It has been shown in Chapter 3 that decentralized approaches can also achieve good load balancing results. The performance has turned out to be appealing in cases, where the intervals between major load changes have been considerably small compared to the reconfiguration intervals of the clustering algorithm. In contrast to the clustering process with bundled user traffic at a relay station, the intra-cell routing process has to consider single user connections. Due to the mobility of single users and their time-dependent traffic demand, the resource requirement may change quickly. The required reaction speed on load changes is consequently by far higher, as the intra-cell routing has to be operated on or even below a second's scale. For that purpose, a decentralized optimization with a reactive scheme is much more difficult.

The proactive routing approach enables a network operation, which is very efficient from the spectrum's perspective. As commonly known, license fees for spectrum allocation are a major impact factor in the operator's business model such that spectral efficiency has always been a major theme in the development of cellular networks. However, the proactive routing approach is associated with inefficiency, which is related to the required signaling overhead. In contrast, the reactive approach is characterized by small signaling overhead, which in turn is paid by a higher price for the resource consumption for data transmissions. A fair comparison between proactive and reactive routing is only feasible considering both, the user data and the signaling overhead. For that purpose, the trade-off between proactive signaling and corresponding routing efficiency will determine the overall benefit. We propose a proactive intra-cell routing protocol in the following. In order to shows its efficiency, we will compare it with the reactive AODV protocol later on.

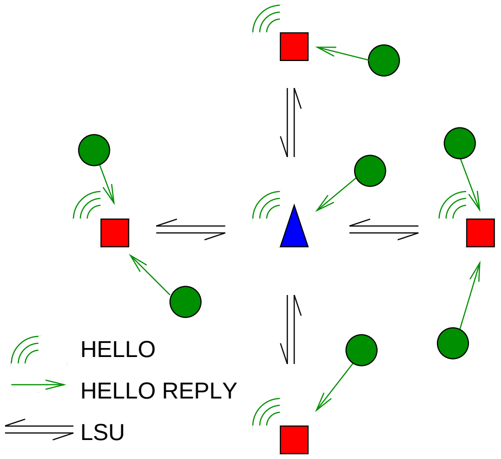


Figure 4.7.: The WOSPF routing protocol uses periodical HELLO and corresponding HELLO reply packets to explore local link conditions. The link state information is globally distributed within the cell by LSU messages.

Protocol Operation

A routing protocol is needed, which mirrors the hybrid position of cellular multi-hop networks somewhere in between fixed and ad hoc networks. It thus seems reasonable that the new protocol has traces of existing routing protocols from the ad hoc as well as the fixed network domain. The following proposal can be considered as a simplified and modified protocol version of the existing OSPF scheme. The new protocol is adjusted to the special demands of cellular multi-hop networks. It is referred to as Wireless Open Shortest Path First (WOSPF), as it is a derivative of OSPF for wireless multi-hop networks. The operation of WOSPF is restricted to the two major packet types of OSPF, namely HELLO and Link State Updates (LSU) packets. While HELLO packets are used to explore local link conditions between neighboring terminals, LSU packets distribute the link state information among all routers within the network. The operation is also illustrated in Figure 4.7.

HELLO Message The HELLO broadcast message signals the presence of a wireless router to surrounding wireless routers and mobile terminals. Any incoming HELLO message is confirmed by the recipient with a corresponding HELLO reply message. The HELLO reply message is sent in unicast mode, as mobile terminals are not enabled to send broadcasts. The packet definition as shown in Figure C.4(a) has been derived from the OSPF RFC 2328. In accordance with the original OSPF specification (see Figure C.4(b)), the HELLO and HELLO reply packet definition includes a 32 bit sender IP address, a gateway or access point IP address, and a 32 bit cell identifier. The gateway IP address and the cell identifier are referred to as designated router and area

ID in the original OSPF standard. The period between two subsequent HELLO packets is defined by a 16 bit beacon interval field. The beacon interval itself is determined by the access point. By receiving HELLO packets from the access point, relay stations are informed about the cell-specific beacon interval. All relay stations are supposed to periodically send out HELLO packets themselves according to the beacon frequency. The number of hops between a wireless router and the access point is signaled in the priority field, whereby the access point is defined to have priority level zero. Each additional hop increases the priority field by one. Each HELLO message includes a given dead interval. This interval determines the period of time, after which a link connection will expire, if no further HELLO reply message is received. Instead of a dead interval, the HELLO reply message contains a connection ID. This field makes it possible to distinguish different applications on the same terminal and thus allows application-specific routing. This means that for example a voice connection could be routed with a minimum hop metric to minimize the delay, while for instance a file download is routed with a minimum resource metric. In practice, one could imagine a mobile terminal to establish several connections to independent routers simultaneously. This is possible, as long as the different routers use orthogonal resources. A time multiplex is thereby the easiest options from the technical perspective. In contrast to the original OSPF packet definition, the 12 byte authentication part is not included. This is due to two reasons: On the one hand, security is out of focus of this thesis, on the other hand the standard-compliant AODV packet does not include corresponding authentication fields either. The header has been adjusted therefore for a fair comparison in terms of signaling overhead. Additionally, a neighbor list is signaled in the original OSPF HELLO specification. This part has been removed as well, as mobile terminals do not require this neighborhood information.

Link State Update Message The access point and relay stations extract the link state information from local HELLO messages and collect it in a corresponding topology table. The topology table consists of a set of topology entries as specified in Figure C.2(a). Each topology entry includes source and target IP addresses of a link connection. Performance indicators as signal strength, SINR, delay, and other measurement parameters are assigned a type of service (TOS) value and can be stored in a TOS metric field. In the following, we want to concentrate on the PHY mode and hop count as two independent performance indicators. This means that an entry is updated, whenever the PHY mode on the distinct link changes. LSU messages are used to incrementally synchronize topology tables among all wireless routers. The incremental update is necessary to reduce the signaling overhead. Each update includes link state information of mobile terminals and their adjacent routers according to the specification in Figure C.5(a). A typical problem known from the OSPF operation in fixed networks is the signaling overhead, which increases polynomially with the size of the network. The overhead results from the flooding of LSU messages through the entire network. We therefore define two possible modes for the exchange of LSU messages:

- *flooding mode*: The LSU messages are flooded through the entire multi-hop cell such that all routers share the same view on the cell topology. The signaling overhead caused by LSU messages increases polynomially in dependence on the number of wireless routers.

- *direct mode*: The direct mode is a measure to reduce the polynomially increasing signaling overhead due to LSU messages. The LSU messages are thereby only sent and forwarded from each relay station to the access point. This approach saves transmission resources, but limits the global topology knowledge to the access point only. Terminal-to-terminal routing within the cell is only feasible with support of the access point in this case. Nevertheless, this mode seems to be an appropriate option, as direct communication between two terminals within the same cell is not a frequent application scenario. We can assume that the large majority of all connections passes through the access point to the backbone network.

4.5.2. Inter-Layer Routing Support by MAC Layer

The previous section has proposed a proactive routing protocol with periodical HELLO messages. The signaling overhead is presumably high in comparison to a reactive protocol like AODV, especially if the system is parametrized in a way that the beacon interval is small. However, a small beacon interval is advantageous, as it enables an accurate tracing of the current topology due to the frequent updates. We therefore want to take a look, whether cooperation with the MAC protocol can help to reduce the signaling overhead. The results will show that a small signaling overhead and a high beacon frequency are not necessarily two contradicting design goals.

Centralized MAC A centralized MAC is a common characteristic of today's cellular networks. The centralized MAC is characterized by the property that transmission resources for multiple wireless user terminals are distributed by a central scheduling entity, e.g., the access point. According to this definition, GSM with TDMA and IEEE 802.16 with OFDMA use centralized MAC protocols. We see the centralized MAC in contrast to a distributed MAC scheme as for instance IEEE 802.11 with its Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) mechanism. Let us recall the fact that cellular multi-hop networks are intended to be an extension of the LTE standard and therefore supposed to be operated by a centralized MAC. The envisioned MAC protocol could be similar to the IEEE 802.16 MAC frame structure as presented in [EMSW02]. A potential TDD-OFDMA frame structure is illustrated in Figure 4.8. This frame structure divides the transmission resource into periodic uplink and downlink intervals. The wireless router broadcasts a preamble at the beginning of each slot. This periodic beacon signal is required by the mobile terminals for physical layer synchronization purpose. The preamble is followed by two OFDM symbols, which represent the Frame Control Header (FCH). The FCH defines size of the downlink map (DL-MAP) and uplink map (UL-MAP). DL-MAP and UL-MAP include the physical layer parameter set for the transmission of uplink bursts and the reception of downlink bursts. The uplink slot also contains a separate ranging subchannel, in which mobile terminals apply for corresponding uplink transmission resources.

These requests are not collision-free, as the access is organized in a contention-based manner according to the slotted ALOHA principle with random back-off timer. Taking a closer look to the IEEE 802.16 standard, two major signaling channels for the exchange of physical layer information can be identified:

- *uplink signaling*: Whenever a user terminal wants to get initial access to a wireless router, the terminal applies for transmission resources in the ranging subchannel with an initial range request (RNG-REQ) message. The inital RNG-REQ message contains the subscriber station MAC address and a standard connection ID 0. The resources are granted by the wireless router in a corresponding range response (RNG-RSP) message. The RNG-RSP includes a set of three terminal-specific 16 bit connection IDs for network management messages. Further connection IDs can be allocated for every application. In dependence on the distinct service class, uplink resources are provided periodically or requested by piggy-backed subheader messages following the regular MAC header of a protocol data unit in the uplink slot. Alternatively, the regular subranging channel or particular so-called unicast, multicast, or broadcast polling channels can be used for further bandwidth requests.
- *downlink signaling*: The corresponding downlink signaling is included in the DL-MAP messages. These messages contain a 48 bit base station ID as well as further physical layer related parameters.

One can imagine two similar methods for the distribution of transmission resources among access point and relay stations within a multi-hop cell: a nested frame structure, where the relay station itself imposes a recursive frame structure within its assigned UL burst (depicted in Figure 2.10), and alternatively a parallelized frame structure, where the access point and the relay stations share resources in a TDM fashion (see Figure A.4). It is assumed in both cases that the distribution of resources and thus the overall frame structure is time-variant and defined by the access point according to the current traffic demand.

Identifying Redundancies A detailed look to the DL-MAP, the ranging subchannel, and especially the piggy-backed subheader add on reveals that the MAC layer information contained in these messages is redundant to the information gathered by the HELLO protocol mechanism in the network layer. We observe that the MAC broadcast frame channel provides equal information to the HELLO message. Similarly, the RNG-REQ message, or subsequent piggy-backed subheaders add ons contain equivalent information to a HELLO reply message. The DL-MAP, RNG-REQ, and piggy-backed requests are necessary and mandatory elements of the OFDMA MAC protocol. Therefore, these messages are independent of the routing protocol and beyond any influence of the network layer.

Inter-Layer Information The available link information in the MAC layer can help to reduce the signaling overhead of the topology exploration protocol on the network

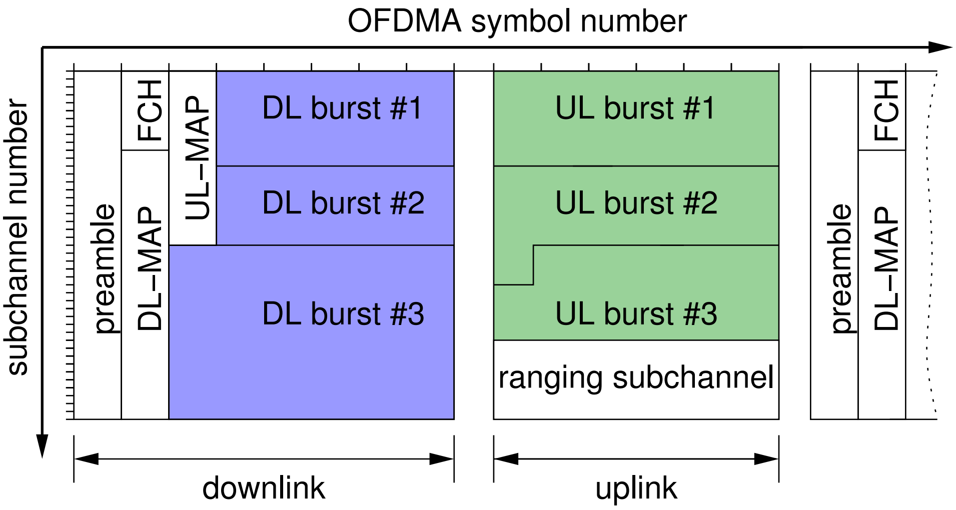


Figure 4.8.: Exemplary TDD-OFDMA frame structure for centralized medium access control.

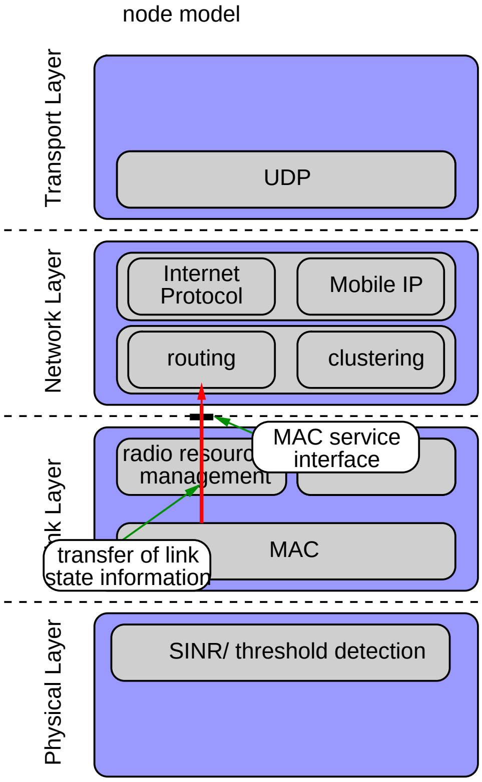


Figure 4.9.: The introduction of a new MAC service interface enables the inter-layer information exchange between MAC and network layer.

layer significantly. Using the Address Resolution Protocol, the router's or the terminal's MAC address can be translated into a corresponding IP address. The individual application-dependent connection ID is available anyway in the MAC frame packets. The source IP address, destination IP address, and the available PHY mode can therefore be passed periodically from the MAC layer to the routing unit in the network layer through a defined inter-layer MAC service interface[2] as shown in Figure 4.9. The corresponding HELLO protocol on the network layer becomes therefore redundant. In doing so, the protocol stack on layer 3 is reduced to LSU messages to be exchanged between wireless routers. In the following, this shortened variant of the WOSPF protocol is referred to as MAC-supported WOSPF. This measure decreases the resulting signaling overhead dramatically, as will be shown in the next section.
Potential drawbacks of the presented solution shall not be omitted at this point: The inter-layer exchange of information contradicts the strict layering principle, according to which each layer provides independent functionalities for other layers. This exchange makes the architecture more complex and may complicate a redesign of the architecture and functional system tests. This fact must be simply accepted from the architectural perspective. Nevertheless, it is important to emphasize that the exchange of information is unidirectional from the lower to the upper layer. Consequently, the proposed operation does not involve cross-layer optimization with a closed-loop bidirectional parameter exchange. Adjustments and modifications in the development process are typically very difficult to handle for cross-layer optimized architectures.

Competitive Approach A comparable routing proposal for cellular multi-hop networks is made in [HYGC07]. This approach includes centralized and distributed routing schemes. In both cases, routing is restricted to the relay-extended infrastructure topology and does not include mobile terminals. [HYGC07] suggests to exchange connectivity and interference information by the IEEE 802.16 proprietary MSH-NCFG message. The approach works for a centralized and distributed routing scheme. Alternatively, the publication explains that REG-RSP messages can be used by the access point to request the positions of relay stations. The location data is used in return to calculate position-based interference patterns and connectivity matrices. The idea to take advantage of OFDMA MAC-specific packet types for the exchange of connectivity information is similar in our proposal and in [HYGC07]. Nevertheless, [HYGC07] is primarily focused on the static wireless infrastructure network. In contrast, our work mainly intends to present an efficient routing scheme, which accounts for the fixed infrastructure network and the dynamic user subnetwork about equally.

Application to IEEE 802.11 The application of the proposal to IEEE 802.11 WLAN is indeed difficult. It is very important to point out that the presented idea does not work together with a decentralized MAC solution. The IEEE 802.11 standard, which is state of the art for today's WLAN networks, foresees two MAC mechanisms, Point

---

[2] A similar MAC service interface is already defined in IEEE 802.15.3 [IEE06b]. Additionally, IEEE 802.21 wants to introduce a corresponding MAC service interface for the purpose of Media-Independent Handover [LSP08].

Coordination Function (PCF) and Distributed Coordination Function (DCF). The DCF is based on the CSMA/CA MAC mechanism to share the wireless medium in a distributed manner. The DCF does not contain any beacon signals, which could be mapped to the information of a HELLO message. The necessary information for the inter-layer exchange is therefore not available on the MAC layer such that the MAC-supported WOSPF is infeasible in this case. Most of today's MAC implementations are based on the DCF and the alternative PCF is often omitted. The PCF is a suitable MAC mechanism for star-shaped topologies with a central coordinating access point. The working principle of the PCF is very similar to the HiperLAN/2 MAC structure, as periodic beaconing and a user feedback channel are included. The additionally proposed inter-layer information exchange is feasible as well in this case. The basic WOSPF protocol is independent on the MAC layer and can be combined with any kind of MAC mechanism.

4.5.3. Simulative Study

Simulation Scenario

The proposed routing protocol has been implemented in a routing unit within the simulation tool as specified in Appendix A. The topology is chosen again according to Figure 4.2, which illustrates a single-cell environment with $|\mathcal{V}_A| = 1$ access point, $|\mathcal{V}_R| = 8$ assigned relay stations, and $|\mathcal{V}_M| = 30$ mobile terminals. The communication radius of the wireless routers is restricted to $R = 120$ m for router-to-terminal connections and $R' = 180$ m for router-to-router connections. The antenna gains of routers and mobile terminals are chosen accordingly. The mobile terminals are uniformly distributed and move according to the random direction mobility model presented in Section A.1. All simulations have been performed for five statistically independent mobility scenarios with a duration of $T = 300$ s each. The presented results are mean values of all simulations. The default values include a beacon frequency of $r_{hello} = 1$ s^{-1} and a maximum user speed of $v_{max} = 10$ m/s for 30 user terminals. The traffic consists of constant bit rate applications with $r = 16$ kbit/s, which are transmitted by UDP clients from the access point to each mobile terminal during the whole simulation time. The presented simulation results include the proposed WOSPF protocol and the RFC-compliant implementation of AODV, AODV-UU of University of Upsala [Wib02], for the purpose of comparison. Due to the physical layer restrictions of the cellular network, broadcast messages from mobile terminals are only delivered to the adjacent wireless routers, any mobile terminal-to-terminal communication is infeasible from the PHY perspective. Only thus is a fair comparison between the routing protocols possible.

Discussion of Results

Mobility Figure 4.10 shows the relationship between signaling overhead and maximum user terminal speed. Two different protocol types are distinguished, namely AODV and WOSPF. WOSPF itself is investigated in four different settings: standard

WOSPF protocol and MAC-supported WOSPF with flooding mode and direct mode each. It can be observed that the number of signaling packets and especially the signaling overhead in terms of bytes per second is increasing with raising average user velocity, which is a very obvious result. Changes in the network topology automatically occur more often in an environment with higher mobility, which results in an increased amount of necessary LSU messages to keep the topology tables incrementally synchronized among each other. Thereby, the proactive topology exploration protocol WOSPF in flooding mode produces by far more signaling overhead than the corresponding reactive protocol AODV. It is also shown that enabling the direct mode decreases the overhead significantly. This is due to the fact that the flooding mode causes any change in the topology table to be flooded by LSU messages among all routers. In contrast, the number of LSU messages is by far smaller in direct mode, as the LSU signaling is restricted to the way back to the access point only. The signaling overhead caused by HELLO messages is very dominant to the little signaling overhead caused by LSU messages. Therefore, the slight increase in LSU-related signaling overhead with raising user speed is almost invisible for WOSPF in direct mode. We can identify dramatical savings in terms of overhead for the MAC-supported WOSPF. This can be explained by the property of WOSPF that, provided a beacon frequency of $r_{hello} = 1\ \mathrm{s}^{-1}$, the number of HELLO and corresponding HELLO reply packets represents the largest portion of the overhead. A very remarkable result is given by the fact that the MAC-supported WOSPF in direct mode even causes smaller overhead than AODV. On the first view, it seems to be quite unbelievable that a proactive protocol can have less overhead than a reactive protocol. Nevertheless, this result is explained by two major facts: The overhead-causing component of the proactive protocol, i.e. the HELLO messages, is not required anymore, as link state information is provided directly by the MAC layer. Additionally, the proactive protocol combines changes in the link state information into bundled LSU messages and signals these messages regularly one-way back to the access point. In contrast, the reactive protocol signals detected link breaks back to the access point, which in return starts a new topology exploration phase. The reactive protocol is therefore lacking the joint proceeding and suffers from this back-and-forth signaling mechanism in case of a link break. The combination of those issues explains the smaller overhead of the MAC-supported WOSPF protocol compared to AODV.

Beacon Interval The beacon interval is typically the Achilles' heel of proactive routing protocols. Generally, a small beacon interval guarantees a high accuracy of the link state information, but also leads to tremendous signaling overhead. Figure 4.11 shows that the MAC-supported WOSPF does not have this disadvantage. The slight increase in signaling overhead for higher beacon frequencies, i.e. for smaller beacon intervals, is caused by additional LSU messages, as we assume the exchange of inter-layer information to be accomplished in periodic intervals according to the beacon frequency. The link tracing becomes more accurate with higher beacon frequency. The number of required LSU messages to update the topology tables is increasing in consequence. However, the impact of this effect is decreasing with raising beacon frequency. As soon as the beacon frequency is high enough to detect all topological changes, the overhead remains constant. The simulated MAC mechanism is able to deliver new inter-layer link state information feedback every 18 ms at maximum, which relates to its frame duration (see Appendix A). Figure 4.11 impressively demonstrates that the use of MAC link state information enables the proactive protocol to operate in an economic manner, as the resulting signaling overhead is tremendous otherwise for accurate beacon frequencies.

Node Number The results in Figure 4.12 indicate the increase in signaling packets with raising node number to ascend much faster for AODV than for WOSPF. The major difference between AODV and WOSPF is the uncoordinated manner, in which the topology information is distributed by AODV, as each node acts independently on other nodes in the reactive protocol. In case of AODV, the route request is distributed around the whole multi-hop cell, each time the distinct last-hop link is broken. This is part of the reactive routing philosophy. Relay stations thereby act as simple forwarding nodes. In contrast, WOSPF assigns more responsibility to the relay stations. Relay stations gather local link state information and distribute updates in bundled packets such that we can observe resulting savings due to this collected forwarding.

Summary In summary, the MAC-supported WOSPF routing protocol can combine a high beacon frequency with small signaling overhead. However, this protocol asks for a centralized MAC scheme like for instance OFDMA. Instead, the general WOSPF protocol is independent on the MAC, but related to considerably high signaling overhead. An interesting observation is given by the aspect that the proactive protocol is not necessarily related to more overhead than a corresponding reactive protocol like AODV, if redundancies with the MAC protocol are used efficiently. The presented topology exploration protocol offers high accuracy and traceability of the current link conditions within a multi-hop cell. In dependence on the chosen mode, i.e. direct or flooding mode, this link state information is available at the access point only or at all wireless router. The given protocol enables advanced routing strategies and traffic engineering techniques. It is thus the basis for the addressing scheme proposed in the next section.

4.6. Data Switching in a Multi-Hop Cell

4.6.1. Label Switching in Cellular Multi-Hop Networks

Design Aspects

A feasible topology exploration scheme to provide permanent knowledge about the current multi-hop cell topology has been proposed in the previous section. Subsequently, we want to address the question, how to use this link state information to efficiently transport data to its final destination. We thereby focus on two issues, the

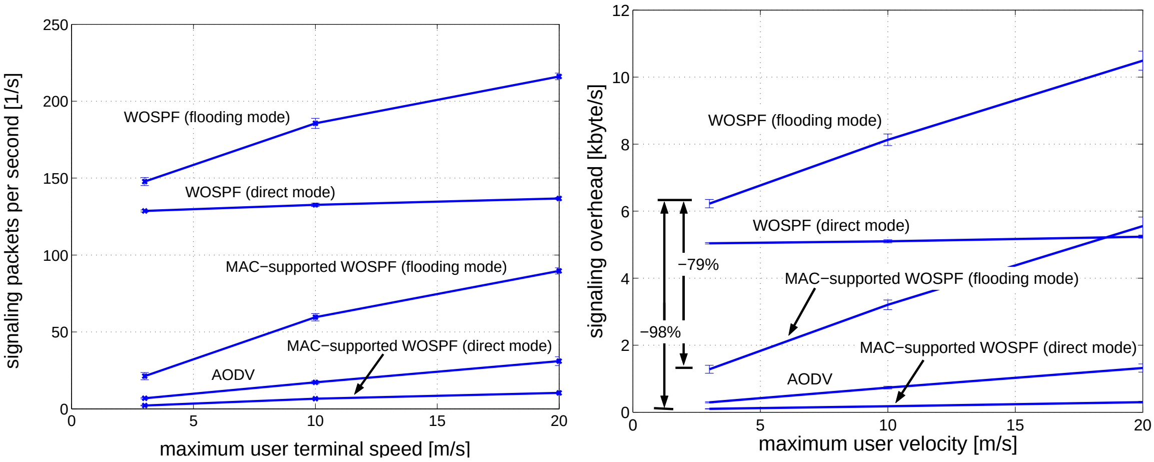


Figure 4.10.: User velocity versus signaling overhead.

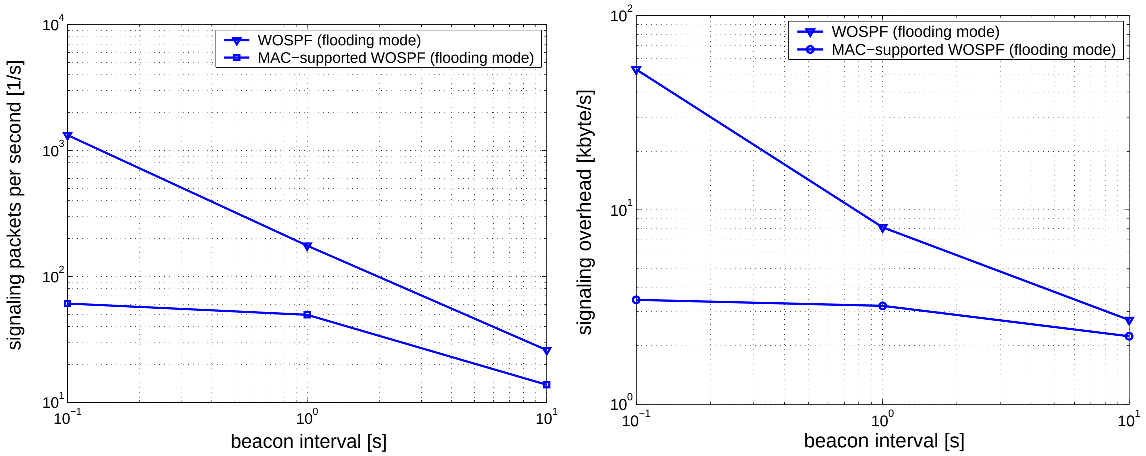


Figure 4.11.: Beacon interval versus signaling overhead.

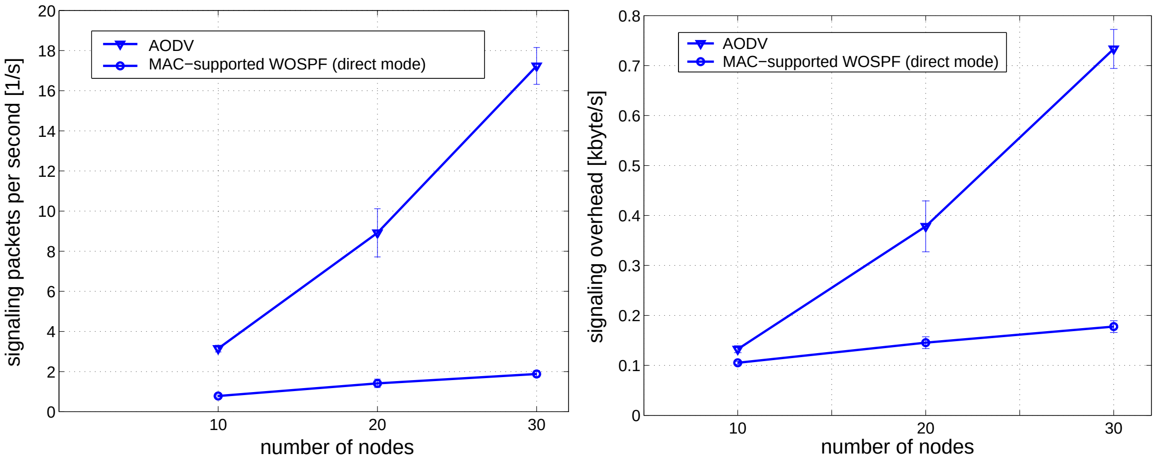


Figure 4.12.: Node number versus signaling overhead.

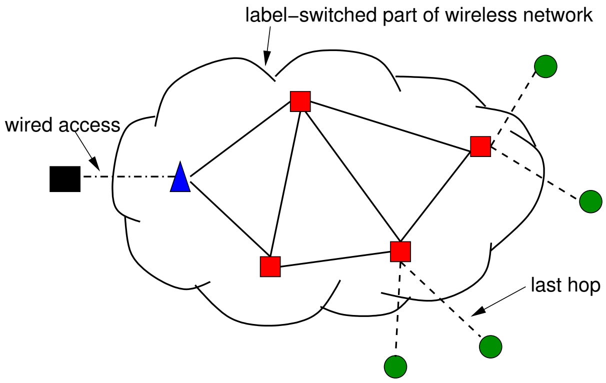


Figure 4.13.: Separation of the multi-hop cell in a static label-switched subnetwork and dynamic user subnetwork.

addressing and the intra-cell handover problem. A solution to the addressing problem has to provide a suitable way to include the transport path information in the data packets. A solution to the handover problem has to guarantee the seamless handover of mobile terminals between the different wireless routers of a multi-hop cell.

It has been shown that the multi-hop cell can be divided into an upper static subnetwork and a lower dynamic network. Due to the static nature of the upper infrastructure network, principles known from the fixed network become applicable. For this part of the network, an adjusted label switching approach will be presented, which is related to MPLS in fixed networks. The static part will therefore be referred to as wireless label-switched network as shown in Figure 4.13. The access point is connected to the wired part of the network by a wired access line, while the mobile terminals are attached to the wireless routers by wireless links. These links are designated as last hop links in Figure 4.13. In analogy to MPLS, this means from the architectural perspective that each access point and relay station can act as ingress and egress router, respectively, for the label-switched network.

Label Switching Approach

Working Principle The choice of the next-hop destination is related to the address of the final destination station for any routing mechanism. In contrast, switching is an addressing mechanism, which assigns path-defining tags to fixed-length data packets. The choice of the next-hop destination depends at any intermediate station on this address tag [KS98]. Prominent examples for such tags are the Virtual Path Identifier and Virtual Channel Identifier in the Asynchronous Transfer Mode (ATM) technology [LB92] or the labels in MPLS. There are different categories of switching mechanisms, cut-through switching and store-and-forward switching. The detection of the packet header is sufficient in cut-through switching [KK80], before the packet is directly forwarded to the next station. Intermediate buffering is therefore not necessary. Store-and-forward switching includes the complete reception of an incoming packet, corresponding compliance checks for its correctness, and intermediate buffering. Any layer 3 switching mechanism therefore belongs to the class of store-and-forward switching

Table 4.2.: The label switching header is included in the data packet in between the layer 3 and layer 2 header.

L2 header	LS header	L3 header	L3 data

mechanisms. A corresponding MPLS protocol with switching functionality on layer 3 is also referred to as packet-based MPLS type in [Law01].

To establish data paths in a multi-hop cell, suppose a switching path to be defined between any pair of wireless routers within the cell. In doing so, a subset of all feasible paths can be chosen and assigned unique labels each. A label is added to each packet header to define a switching path for the packet through the multi-hop cell. This distinct label switching header is thus inserted before the layer 3 header as shown in Table 4.2. In the downlink case, the label is added at the access point and detached at the corresponding relay station, before the packet is finally transmitted from this relay station to the mobile terminal. In the uplink case, the packet is sent from the mobile terminal to the wireless router, where again a label is added. The packet travels through the network along its switching path in accordance to the label. Provided the case that the mobile terminal is directly attached to the access point, it is of course not necessary to add any label.

Switching Path A switching table is available at each wireless router. The switching table provides the relationship between a label and the corresponding IP address of the next hop wireless router. The individual switching table of each wireless router can be interpreted as a local piece of a network-wide jigsaw. Seen from the overall perspective, the switching tables define a system of pre-configured paths through the network. The switching table is assigned by the access point or optionally by another higher-level instance in the wired access network. The switching table can be signaled once initially remaining unchanged during further operation, but could be also signaled and adjusted incrementally, entry by entry, during the operation of the network. A set of signaling messages has been defined in [Reu06] to enable this exchange of switching tables between access point and relay stations. The set of paths does not necessarily contain all possible paths between two terminals in the network. Generally, it is sufficient to define a subset of all possible paths. Our proposal suggests an 8 bit label such that a set of $2^8 = 256$ paths can be addressed within each cell. Thereby, three different types of paths are distinguished:

- *standard*: An arbitrary number of standard paths leading from the access point to any relay station and vice versa can be defined.
- *signaling*: A special path among all feasible paths from each relay station to the access point is chosen to transmit signaling information. This path is especially necessary in the initialization phase. It can be determined by auto-configuration

from knowledge available in the local radio resource management entity or set up by a higher-level instance.

- *default*: A default path can be defined and used for packets without any path information to save transmission resources and overhead. There is one default path from each wireless router to the access point. This path can be established by auto-configuration, but also overruled and reset by the central assignment of the higher-level instance. The advantage of such a default path is that uplink packets could be forwarded without any signaling overhead.

Routing Table The switching table defines paths or path segments through the fixed part of the wireless network. In contrast, the mobile terminal itself is registered in a routing table at the corresponding wireless router. This routing table relates from the routing protocol WOSPF and is therefore part of the topology exploration scheme. Whenever a packet arrives at a router with its label indicating that the path ends, the wireless router will take a look in its routing table to forward the packet directly to the mobile client. If this is impossible for any reason, for instance some unpredictable link break, the packet will be thrown away. Vice verse, each mobile terminal keeps a routing entry with its current default wireless router.

Switching Policy A label is assigned to each data packet. The label consists of an 8 bit path identifier and a 3 bit service class assignment to support prioritized queuing and scheduling at intermediate relay stations. The initial label is assigned at the access point as ingress router with backbone connection. In a similar way as in IEEE 802.16, bandwidth guarantees for different service classes (unsolicited grant services, real-time polling services, non-real-time polling service, best effort) are considered. The assignment of the label is dependent on the switching policy, which could be for example the result of a route optimization with advanced routing metrics as already examined in the previous sections. Concerning the switching policy, two major scenarios are distinguished:

- *terminal-oriented policy:* In this case, the mobile terminal is registered in the routing table of the access point. The routing algorithm identifies a suitable path and adds the corresponding path label to each incoming data packet. We could imagine a corresponding uplink transmission on the reverse path. The label for a path and the resulting reverse path could be related for example by a special bit within the label to characterize uplink and downlink direction. Alternatively, uplink packets can also be guided along the mentioned default path without any label.

- *application-oriented policy:* As already presented, IEEE 802.16 foresees the distribution of connection IDs as a response to an initial ranging request. Different connection IDs are distributed to provide service guarantees for network management, signaling tasks, and applications. The connection IDs are typically subject to different service requirements. A single terminal is therefore registered with different connection IDs at the access point. It is thus possible to switch different applications on different paths. For instance the minimum hop path with small delay could be used for low-data rate voice services, while the high data rate streaming services could be switched on a multi-hop connection with high capacity.

The switching policy allows the transition from a terminal-oriented transport strategy toward an application-oriented or service-oriented data transport strategy.

Compatibility to Cellular Systems Another issue is that the introduced label has a fixed field header length. In contrast, possible source routing approaches typically include a variable hop-dependent number of header bits. Thus, the header length is either variable or characterized by the worst-case reservation of the maximum number of hops. As cellular networks have strong attitude toward efficiency and fixed frame structures, the introduction of label-based signaling seems to be a suitable approach. In addition to that, the switching approach makes it possible to fully rely on existing cellular architecture, especially for the inter-cell handover. Thus, the switching mechanism is a very essential component in the primary idea to develop cellular multi-hop as an extension to existing cellular systems, not as a completely new system.

4.6.2. Comparison between Cellular Multi-Hop Label Switching and MPLS

Figure C.1(a) and C.1(b) compare the standard-compliant MPLS header with the cellular multi-hop label. It is visible that the amount of label-defining bits has been reduced from 20 bits to 8 bits. This due to the fact that an address space of 256 potential paths will be enough for single multi-hop cell. The 3 bits for the service differentiation have been kept. However, the stack bit have been removed due to lack of necessity. Originally, the stack bit has been used to indicate the last header of a sequence of several subsequent MPLS headers. Similarly, the time-to-life bit has been abandoned, as protection against loops due to wrong configuration or slow convergence of the routing algorithm are not necessary either. The proposal keeps several elements, which are typical for MPLS.

Let us first take a look at common characteristics: The use of labels to define path segments through the network is essential to MPLS, the label header of the proposed protocol is derived from the MPLS header. Neither MPLS nor the proposed multi-hop label switching mechanism are intended to provide end-to-end connectivity between a source and a corresponding destination node, but to guide a packet safely through a pre-defined network segment. A packet enters the label switching zone at the ingress router and leaves it at the egress router. Similarly, access point and relay stations act as entry and leaving points for the switched network zone, while the last hop between wireless router and mobile terminal is out of scope. Another common aspect of MPLS

Table 4.3.: Comparison between MPLS and cellular multi-hop label switching.

MPLS	cellular multi-hop label switching
ingress/egress router	access point and all relay stations
label-switch router	relay station
autonomous system	multi-hop cell
connection-dependent label assignment	path-dependent label assignment
on demand label assignment	predefined label assignment
multiple labels per path	only one label per path possible

and the proposed protocol is the lack of ability to work as stand-alone routing protocol independently from other supplementary protocols. While typically the Intermediate System To Intermediate System Protocol (IS-IS) [ISO02] or OSPF are used to collect topology data for MPLS in wired networks, the previously described WOSPF fulfills this task in cellular multi-hop networks.

However, there are also some important differences to MPLS. During the establishment of a MPLS connection, several labels are typically used to reserve a path with exclusive use of transmission resources through the network between ingress and egress router. Each of the labels reserves a path segment on the overall path. In contrast, we only assign a single label to each path. According to MPLS, labels are uniquely assigned in a local environment for a limited time to a distinct data connection. Instead, provided the presented proposal, the label is uniquely and permanently assigned to a path within the multi-hop cell and not to a specific connection. This also means that the task of resource reservation is not managed by the label switching mechanism, but granted by the corresponding routing entity at the access point. Several connections may share the same label, in case they are routed on the same path. The comparison between MPLS and cellular multi-hop label switching is also summarized in Table 4.3. To conclude, the proposed label switching strategy shows several similarities to MPLS, nevertheless it is a proprietary mechanism adjusted to the specific needs of a multi-hop cell.

4.6.3. Intra-Cell Handover

The operation of a multi-hop cell requires a fast and efficient intra-cell handover mechanism. It is assumed that the handover is triggered by the access point as known from cellular networks. The mobile terminal itself only submits measurement reports and follows the instruction of the access point. Relay stations do not initiate any handover, but are strictly dependent on the access point. Let us therefore sketch a principle handover situation. The handover can be characterized by three different phases, as illustrated in the message sequence chart of Figure 4.14:

- *initiation phase*: The topology exploration protocol makes sure that the access point receives regular measurement reports about the connectivity within the multi-hop cell. If the access point decides to handover a mobile terminal from one relay station to another relay station, the access point will trigger the mobile terminal and therefore will send out a corresponding HANDOVER_COMMAND message to the mobile terminal. This message also contains the address of the target relay station.

- *association & duplication phase*: While the mobile terminal is preparing and executing the handover, the access point starts duplicating and multicasting the ongoing data stream to both, the current relay station RS1 and the target relay station RS2 in order to enable a seamless handover.
- *completion phase*: After having successfully completed the handover, the mobile terminal informs the access point by a HANDOVER_COMPLETE message. This is the sign for the access point to stop the transmission of the duplicated data flow to the original relay station RS1. The access point confirms the completion by a corresponding HANDOVER_CONFIRM message.

The successful transmission of the presented messages is assured by a corresponding ARQ protocol. Potential timers preventing failures and causing retransmissions are not depicted in this chart. The handover principle will not change, if the mobile terminal changes its current link from or to the access point itself. The proposed handover mechanism guarantees to the access point full control on the connectivity within the whole cell.

Inter-Cell Handover The presented mechanism is fully compliant to the inter-cell mechanism in LTE as described in [BHKV07] and [RTR07]. Any inter-cell handover can thus be performed as proposed in [BHKV07] for LTE services. In principle, it is also possible to extend the switching mechanism in a way that the access gateway becomes part of the label switching area. This implies that the path mechanism is extended to several multi-hop cells and thus inter-cell handovers can be performed more easily. Nevertheless, this would require architectural changes in the LTE system.

Prototypical Implementation The described protocol has been prototypically implemented on notebooks with Ubuntu Linux 6.06 and IEEE 802.11g WLAN interface, provided by the D-Link DWL-G122 USB WLAN stick with the RT2501 USB open source device driver. The software implementation included the proactive topology exploration within a multi-hop cell, the label-switched addressing, a link quality dependent routing metric, and graphical user interface for the illustration of routing paths [Sch06]. Even if IEEE 802.11g is not comparable with cellular devices from the PHY and MAC layer perspective, some single-user multi-hop field trials have been accomplished successfully and demonstrated prototypically that the intended combination of proactive topology exploration, numerical route optimization, and label-switched addressing is indeed feasible and enables seamless intra-cell handover.

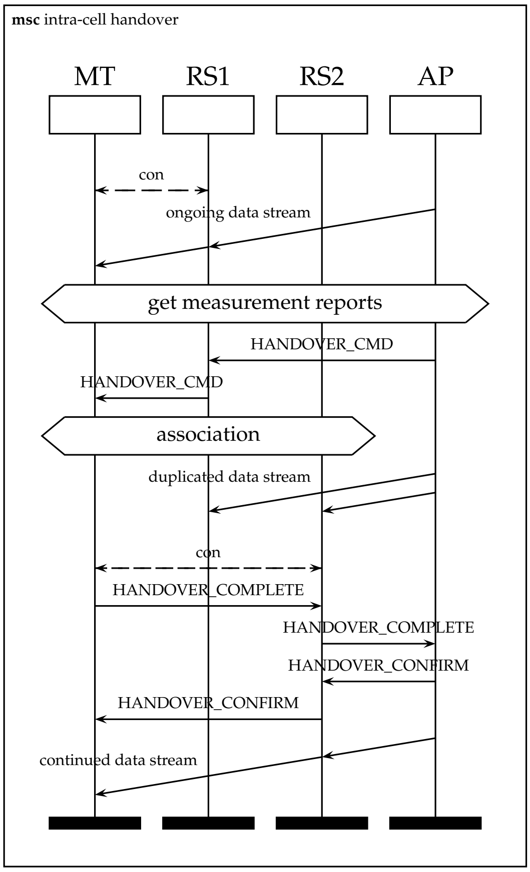


Figure 4.14.: Intra-cell handover between two relay stations of a multi-hop cell.

4.7. Conclusion

The major statement of the first part of this chapter is that a suitable routing metric is unavoidable for the operation of a resource-efficient multi-hop network. An important result is given by the observation that the efficiency of the routing metric is mainly dependent on the physical design of the network. While the impact of the routing protocol is limited in networks with considerably small overlapping area between neighboring routers, the routing metric is a key parameter in networks with large overlapping areas. Furthermore, it has been also illustrated that the efficiency of the relay link is very essential for the overall resource demand within the network. Another interesting aspect is pointed out by the results of the best connection strategy, which may lead to significantly higher resource demand than other routing metrics. The second part of this chapter has illustrated that a proactive protocol is not necessarily related to more overhead than a reactive protocol. However, this fact mainly depends on the design of the MAC protocol. While the proposed inter-layer communication cannot be applied to network architectures with decentralized medium access based on IEEE 802.11 DCF, it is very convenient to cellular-like network architectures with centrally scheduled medium access. The last part of this chapter has shown, how label switching as known from the fixed network domain can be used as efficient addressing mechanism in cellular multi-hop networks. It has been illustrated that the proposed label switching mechanism is able to support application-specific routing. Additionally, the overall architecture enables a fast intra-cell handover and supports the LTE-proprietary inter-cell handover mechanism. To sum up, this chapter has investigated the design and impact of wireless multi-hop routing protocols and comes up with a suitable proposal for intra-cell routing in a relay-extended multi-hop cell.

5. Network Coding in Cellular Multi-Hop Networks

5.1. Introduction

Network coding is a data transport technique, which has been discovered recently. Scientific research has proved that network coding can achieve the multicast capacity of wired and wireless networks. Research has shown that typically this is infeasible with routing such that network coding is spectrally more efficient than routing, as it requires less transmission resources to accomplish the same result.

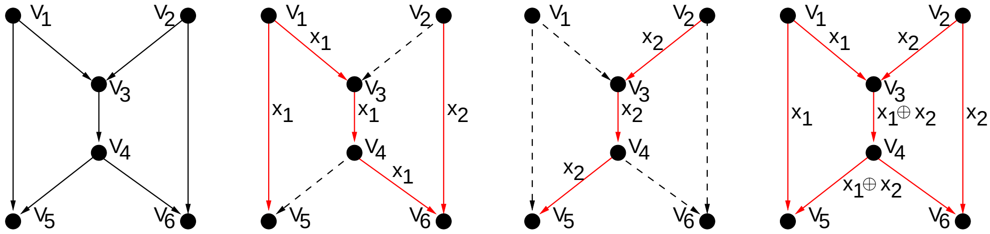
We start our investigation with a short explanation of the graph-theoretical and mathematical models for network coding and take special interest in the difference between network coding and routing. Network coding is known to perform especially well in broadcast scenarios. For that purpose, we want to take a deeper look on the broadcast case in the first part of this section. As the performance of network coding depends on its inherent ability to distribute the information on various data paths, the number of access points in the cellular multi-hop scenario is a key parameter for the efficiency of the network coding approach. We present a result for the expected resource allocation efforts in dependence on the access point density of the cellular multi-hop network. In the second part, we concentrate again on unicast applications. As the general case of arbitrary unicast traffic is evidently hard to handle, we comment on bidirectional relaying and its feasibility in cellular multi-hop networks. Finally, we conclude with a summary of the preceding insights.

5.2. Motivation

Network Coding for Broadcast Applications It has been pointed out in various sections of this thesis that spectral efficiency is a valuable good in cellular networks. A couple of years ago, network coding as a new data transport technique has been brought up by [ACLY00]. Network coding helps to achieve savings in terms of transmission resources. These gains can be primarily observed in multicast or broadcast scenarios. The multicast scenario is characterized by a single node providing data for a set of distinct destination nodes. An extension of the multicast scenario is the broadcast scenario, in which a single node provides data for all other nodes in the network. In the following, we will refer to a distribution of a data from a central node to all access points and relay stations as broadcast. For simplicity reasons, mobile terminals are not accounted for at this point. Potential broadcast applications in cellular networks include user-related services like cell phone television, digital audio broadcasts, the distribution of local news, and advertisement offers. The broadcast mechanism can alternatively be applied to administrative tasks such as the distribution of signaling or network state information to all wireless routers.

Network Coding for Unicast Applications The application of network coding is also interesting for relaying purpose, where a pair of oncoming flows between two nodes is relayed by one or several intermediate stations [WCK05a]. This scenario is known from literature as bidirectional relaying and is a special case of the general unicast scenario. Any joint downlink and uplink connection between an access point and a mobile terminal is a potential use case for bidirectional relaying. Network coding is not restricted to broadcast or bidirectional relaying scenarios, but can be applied to any set of unicast demand flows. A corresponding mathematical problem formulation is provided in [TRL$^+$06] and [RTK06]. The challenge consists in the tremendous computational effort, which makes it impossible to determine a solution for larger scenarios. As apparently a great part of all realistic scenarios within a cellular multi-hop network is covered by the broadcast and the bidirectional relaying case, we will consider these cases only in the following. In doing so, we want to concentrate on the question, whether network coding is a suitable feature for cellular multi-hop networks.

Butterfly Example Let us therefore consider a butterfly-shaped network (see Figure 5.1(a)) as an introductory example into network coding. This example is a popular illustration of the network coding principle and has been brought up by important literature sources as [ACLY00],[KM03], and [LYC03]. Assume a set of communication devices to be represented by 6 nodes as shown in Figure 5.1(a)). The neighboring nodes are connected by graph edges, whose capacities are restricted to the transmission of a single message each. In the following, node V_1 as well as node V_2 want send two corresponding messages x_1 and x_2 to both, node V_5 and node V_6. For simplicity reason, assume the messages x_1 and x_2 to be binary symbols in {0,1}. The situation features a typical multicast example with two sources and two receivers each. It is a possible approach to initiate a message x_1 from node V_1 to be transmitted as shown in Figure 5.1(b). The message travels along two independent transmission paths {(V_1,V_5)} and {(V_1,V_3),(V_3,V_4),(V_4,V_6)}. In addition to that, node V_2 could simultaneously transmit message x_2 on edge (V_2,V_6). A further possibility to also transmit its message to node V_5 is unfortunately infeasible, as the connection over edge (V_3,V_4) is already blocked by the transmission of message x_1. An allocation of new capacities for all edges for instance in a second time frame is imaginable. In doing so, node V_2 could finally send message x_2 over {(V_2,V_3),(V_3,V_4),(V_4,V_5)} to node V_5 (see Figure 5.1(c)). The described approach fulfills all criteria of a typical routing approach: Intermediate nodes simply forward data packets based on the final destination. However, intermediate nodes do not change the content of a data packet in any respect. Accounting for all allocated resources, the requirement sums up to eight transmission resources to realize the proposed routing approach. It can be shown that network coding may arrange the transmission in a more efficient manner. The corresponding illustration is shown in Figure 5.1(d). Node V_1 and node V_2 send out simultaneously their messages on edge



(a) Example network in butterfly topology. (b) Transmission of x_1 blocks x_2. (c) Transmission of x_2 to node 5. (d) Simultaneous transmission of creation of a coded message $x_1 \oplus x_2$.

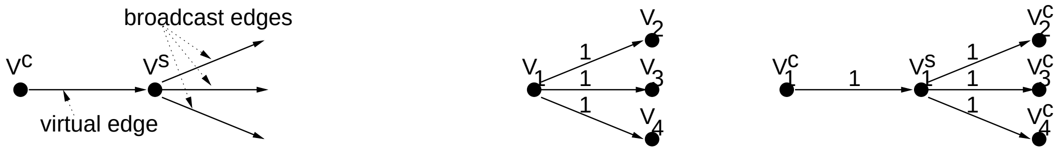
Figure 5.1.: Basic principle of network coding.

(V_1,V_3) and edge (V_2,V_3), respectively. Instead of forwarding the incoming messages separately, node V_3 forwards a combined message $x_1 \oplus x_2$ to node V_4. The exclusive-or operation "$\oplus$" represents the addition of two signals in a simplified mathematical signal space, a Galois field $\mathbf{GF}(2)$. The encoded message is finally sent over edge (V_4,V_5) and (V_4,V_6) to the receiver nodes. Given the fact that node V_5 has already received message x_1 over edge (V_1,V_5), it can reconstruct x_2 using $x_1 \oplus x_2$ and x_1. In completely analog manner, node V_6 reconstructs message x_1. In doing so, network coding allows the simultaneous transmission of messages x_1 and x_2. In comparison to routing, the number of allocated resources is reduced by one to seven resources. This example illustrates that network coding can increase the efficiency of data transmissions in communication networks. Following this idea, we will examine, whether the application of network coding is advantageous in cellular multi-hop scenarios.

5.3. Related Work

5.3.1. Wireless Multicast Graph Model

There is one property of the wireless transmission that becomes especially important for network coding, namely the broadcast ability of the wireless medium. Only minor attention has been paid to this fact, as no more than unicast applications have been considered so far. For this reason, it has been fully sufficient to model the wireless device as a simple graph node. However, the broadcast property is important for any sort of multicast routing and especially for network coding. A graph-theoretical model to describe the multicast property adequately is therefore required. Wu uses in [WCZ$^+$05] a representation of the wireless device by two graph nodes by introducing a virtual transmitter node V^s. The virtual transmitter node is attached to the core node V^c over a virtual edge as shown in Figure 5.2(a). According to [WCZ$^+$05], the virtual edge acts as an artificial bottleneck, which constrains the rate of the outgoing information flow of the transmitter. Let us therefore compare an exemplary transmission in the simple single-node graph model as shown in Figure 5.2(b) and the two-node graph model in Figure 5.2(c). The capacity allocation as given in Figure 5.2(b) is suf-



(a) Graph-theoretical node model.

(b) The assigned capacity constellation in the simple single-node graph model enables simultaneous and independent unicast transmissions from node V_1 to nodes V_2, V_3, and V_4.

(c) The assigned capacity constellation allows a single unicast from node V_1 to nodes V_2, V_3, and V_4 or alternatively a broadcast transmission from node V_1 to the attached nodes.

Figure 5.2.: Graph-theoretical model for unicast and multicast transmission.

ficient for three independent unicast transmissions, as each of the broadcast edges is independent from the other ones. In contrast, the virtual edge of Figure 5.2(c) limits the output capacity of V_1^C. The virtual edge allows the simultaneous broadcast transmission of a single message from a transmitter to all surrounding nodes or alternatively a single unicast transmission of one message from the transmitter to one of the surrounding nodes. A simultaneous transmission of three messages as in the model of Figure 5.2(b) is clearly infeasible. The set of resources on any of the virtual edges and the broadcast edges represents physically a single transmission resource. Although the broadcast transmission in Figure 5.2(c) requires four virtual resource elements, the wireless medium is only blocked for one resource slot. For the investigation of required resources, it is therefore sufficient to only account for the resource elements of the virtual edge. Suppose a network to consist of N wireless devices. In this case, the given graph model results in a set of core nodes $\mathcal{V}^c$ and virtual transmitter nodes $\mathcal{V}^s$ with $|\mathcal{V}^s| = |\mathcal{V}^c| = N$. The graph-theoretical model includes $|\mathcal{V}| = 2 \cdot N$ nodes, whereby the set of overall nodes $\mathcal{V}$ is composed of $\mathcal{V} = \mathcal{V}^s \cup \mathcal{V}^c$.

5.3.2. Linear Network Coding Approach

Let a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, c\}$ define a network topology with nodes $V \in \mathcal{V}$, edges $\vec{e} \in \mathcal{E}$, and corresponding edge capacities $c(\vec{e})$. Symbols of a finite field length are transmitted on the edges. Assume demand functions to determine rate requirements between source and destination nodes. Furthermore, suppose the symbols x_i on the incoming edges of a node to be mapped into symbols y_j on the outgoing edges on a node, $\forall V \in \mathcal{V}$:

$$y_j = \mathrm{f}(x_1, \ldots, x_i) = \sum_{\vec{e}_i \,\in\, \mathcal{E}^{(in)}(V)} \alpha(\vec{e}_i) \cdot x_i(\vec{e}_i). \tag{5.1}$$

The combined set of edge functions and demand functions is referred to as network code [DFZ05]. If the mapping is a linear relationship, the corresponding technique is known as linear network coding. The solution is said to be scalar, if $\alpha(\vec{e}_i)$ is a scalar

variable. Linear network coding has been introduced some years ago in the pioneering work of Ahlswede [ACLY00]. Ahlswede has shown that the multicast capacity of a network can be achieved by linear network coding and that this outperforms routing. [LYC03] has additionally shown that there is a scalar linear solution for any solvable multicast network given a sufficiently large finite field alphabet.
[WCZ$^+$05] provides a mathematical model to describe the network coding problem by a set of linear equations. We will mostly follow the notation of [WCZ$^+$05] and briefly sketch the corresponding approach. For that purpose, assume a multicast session $m \in \{1,\ldots,M\} \in \mathbb{N}^+$ from a source node $V_s^m \in \mathcal{V}$ to a set of destination nodes $V_{d_j}^m \in \mathcal{V}$ with $j \in \{1,\ldots,J\} \in \mathbb{N}^+$. The connection between the source node V_s^m and a destination node $V_{d_j}^m$ is established by a single flow f_j^m and operated at rate r^m. Accordingly, the multicast session m between source and all destination nodes is described by the set of all flows $f_j^m \in \mathcal{F}^m$. We suppose the flow f_j^m to be created by a set of flow elements $f_j^m(\vec{e})$ with $f_j^m(\vec{e}) \geq 0 \; \forall \vec{e} \in \mathcal{E}$.

Envelope Flow The individual flow elements $f_j^m(\vec{e})$ of multicast session m are covered by an envelop flow, which is characterized by a resource requirement $g^m(\vec{e})$,
$\forall m = \{1,\ldots,M\}, \forall j = \{1,\ldots,J\}, \forall \vec{e} \in \mathcal{E}:$

$$0 \leq f_j^m(\vec{e}) \leq g^m(\vec{e}), \tag{5.2}$$

or — described differently — $g^m(\vec{e})$ is the maximum of all flow elements $f_j^m(\vec{e})$,
$\forall m = \{1,\ldots,M\}, \forall \vec{e} \in \mathcal{E}:$

$$g^m(\vec{e}) = \max_{j \in \{1,\ldots,J\}} \{f_j^m(\vec{e})\}. \tag{5.3}$$

Capacity Constraint According to the approach in [WCK05b], the individual resource allocation of each session $g^m(\vec{e})$ adds up to an overall resource allocation $g(\vec{e})$. The overall resource allocation is bounded by the overall capacity $c(\vec{e})$,
$\forall \vec{e} \in \mathcal{E}:$

$$g(\vec{e}) := \sum_{m=1}^{M} g^m(\vec{e}) \leq c(\vec{e}). \tag{5.4}$$

Flow Preservation The multicast flows start at the corresponding source nodes V_s^m. We therefore observe a net flow of rate r^m, which leaves the source nodes,
$\forall m \in \{1,\ldots,M\}, \forall j \in \{1,\ldots,J\}, \forall V_s^m \in \mathcal{V}:$

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V_s^m)} f_j^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V_s^m)} f_j^m(\vec{e}) = r^m. \tag{5.5}$$

A basic physical law also known as law of flow preservation states that the amount flow can only be changed at source node and destination node. For all intermediate nodes, the flow can neither be increased nor be decreased. To keep the flow constant, the equality of incoming and outgoing flows is defined at all intermediate nodes: $\forall m = \{1, \ldots, M\}, \forall j \in \{1, \ldots, J\}, \forall V \in \mathcal{V} \backslash \{V_s^m, V_{d_j}^m\}$:

$$\sum_{\vec{e} \in \mathcal{E}^{(out)}(V)} f_j^m(\vec{e}) - \sum_{\vec{e} \in \mathcal{E}^{(in)}(V)} f_j^m(\vec{e}) = 0. \tag{5.6}$$

Even if the arrival of the flow at the corresponding destination node is not assured by a distinct equation, each flow will arrive at the destination node with rate r^m. The flow enters the network at the source node (see Equation (5.5)). Subsequently, the flow remains constant at all intermediate nodes (see Equation (5.6)). Therefore, the flow will intentionally arrive at the destination node.

The authors of [ACLY00] prove in their publication that the multicast capacity for a linear network corresponds to the minimum of all possible maximum flows between source and each individual destination node. For that purpose, let us consider a source-destination pair within a network graph with fixed capacities. The maximum flow for any source-destination pair is defined as the maximum possible flow between source and destination, whereby no other destination node of the multicast connection is considered. According to the min-cut max-flow theorem [EFS56], the maximum flow between two nodes V_s and V_d of a network graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is determined by its minimum cut $\rho(V_s, V_d)$. We define this cut as a separation of the nodes into two subsets $\mathcal{V}_1 \subset \mathcal{V}$ and $\mathcal{V}_2 \subset \mathcal{V}$ such that $V_s \in \mathcal{V}_1$ and $V_d \in \mathcal{V}_2$ with $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \{\}$.

The capacity of a cut ρ_i is

$$c(\rho_i) := \sum_{\vec{e}_{cut} = (V_1, V_2) : V_1 \in \mathcal{V}_1 \wedge V_2 \in \mathcal{V}_2} c(\vec{e}_{cut}), \tag{5.7}$$

which equals to sum capacity of all graph edges being crossed by the cut. It is known from graph theory that the maximum number of different cuts ρ_i is bounded by

$$\left| \bigcup_i \{\rho_i\} \right| = 2^{|\mathcal{V}|-2}. \tag{5.8}$$

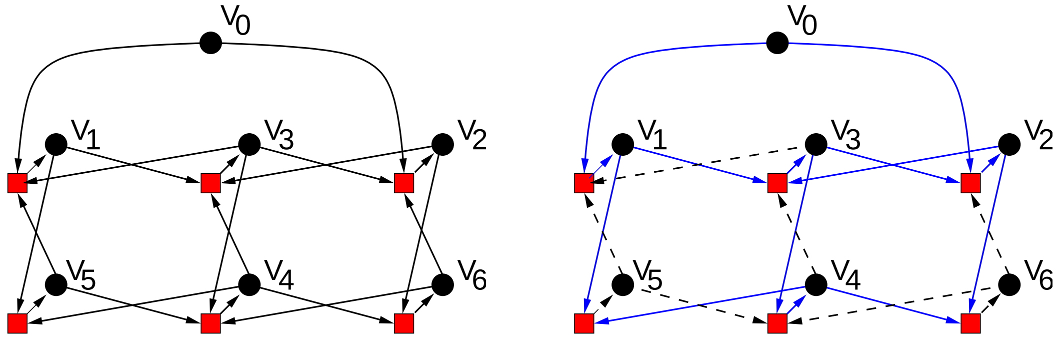
The corresponding minimum cut $\rho(V_s, V_d)$ is therefore defined as

$$\rho(V_s, V_d) := \min_i \rho_i \tag{5.9}$$

and defines the capacity $c(V_s, V_d)$ between source V_s and destination V_d, which is equivalent to the feasible maximum flow. The Ford-Fulkerson algorithm [FF62] is an appropriate algorithm to determine such a maximum flow. This maximum flow procedure is repeated now for all source-destination node pairs of a single multicast connection. The achievement of [ACLY00] is the proof that the multicast capacity of a network is given by the minimum of all maximum flows:

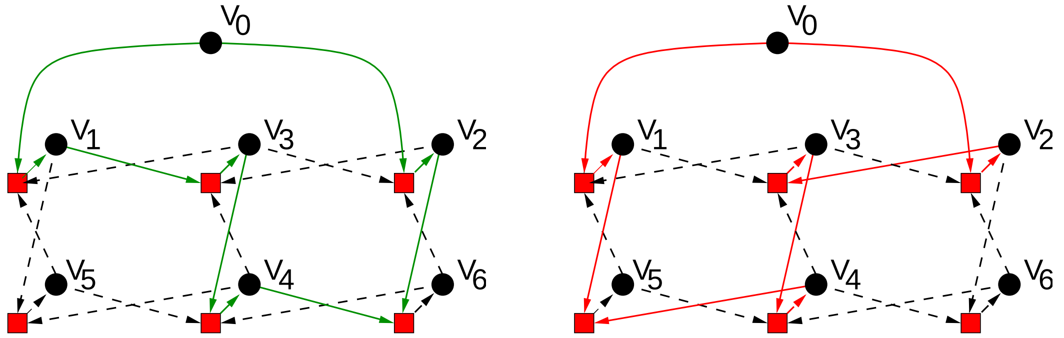
$$c(V_s, (V_{d_1}, \ldots, V_{d_J})) := \min_{j \in \{1, \ldots, J\}} c(V_s, V_{d_j}). \tag{5.10}$$

The same work has also shown that this capacity can be achieved by linear network coding.



(a) Extended scenario with additional source node.

(b) Combined data flow g with $\forall j$: $g \geq f_j$.



(c) Data flow f_1.

(d) Data flow f_2.

Figure 5.3.: Illustration of solution approach with example.

Example Let us illustrate the presented approach with an example and therefore extend the initial butterfly example by the proposed graph-theoretical multicast node model and by an additional source node V_0, which is connected with node V_1 and node V_3 over two edges $\vec{e_1}$ and $\vec{e_2}$ (see Figure 5.3(a)). The available capacity of all edges is constrained by $c = 1$ transmission per time unit. A corresponding multi-cast problem can be formulated with $M = 1$ and rate $r = 2$ transmissions per time unit, whereby node V_0 acts as source and node V_5 and V_6 are destination nodes. The problem formulation corresponds to the previously mentioned example. The solution of the problem with the above given notation leads to two separate subflows f_1 to node V_6 (see Figure 5.3(c)) and f_2 to node V_5 (see Figure 5.3(d)). The envelope flow g, which represents an overall solution to the problem, results from Figure 5.3(b) as common maximum of f_1 and f_2. Simultaneously, g is constrained by the capacity of the individual edges and therefore underlies Equation (5.4). The additionally introduced data source V_0 is necessary to model the overall problem as multicast problem, as only thus network coding according to the multi-cast capacity theorem of [ACLY00] is applicable.

5.3.3. Routing versus Network Coding

We have already seen an approach to describe network coding mathematically. In the following, we want to investigate the relationship between network coding and routing. It will be finally shown that minimum cost routing is a special case of the more general network coding approach. Minimum cost routing can performed in a network graph, where each edge is assigned a constant edge weight or cost. The routing approach, which minimizes of the overall routing costs, is referred to as minimum cost routing then. [WCK05b] proves that the problem of minimum cost routing is equivalent to the problem of minimum energy routing in wireless networks, where each transmission between two nodes is related to a fixed transmission power or — more generally speaking — resource consumption. This also implies that the power requirement has to be independent on interference, i.e. only the SNR but not an SINR can be considered, as otherwise the power requirement is variable. The minimum cost routing problem is equal to the determination of the maximum number of edge-disjoint subgraphs $\mathcal{S} \subseteq \mathcal{G}$, which connect a given source node with the set of destination nodes. This problem is also referred to as Steiner tree packing problem in the literature [JMS03]. Each of the subgraphs is defined by a tree structure and includes a set of paths from the source node to all destination nodes. The number of potential Steiner trees results from the set of all combinatoric possibilities for such paths. Under the assumption that a minimum cost routing shall be determined, the corresponding solution is described by exactly one Steiner tree, the so-called minimum cost Steiner tree as shown by the proof in [WCK05b].

We have already seen that information can be distributed by network coding on an arbitrary number of independent paths between source and destination node. The edge-specific maximum resource allocation $g^m(\vec{e})$ is therefore naturally enveloped by the rate r^m of a multicast:
$\forall m = \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$0 \leq g^m(\vec{e}) \leq r^m. \tag{5.11}$$

A normalization of this equation leads to
$\forall m = \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$0 \leq \frac{g^m(\vec{e})}{r^m} \leq 1, \tag{5.12}$$

whereby we define
$\forall m = \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$\hat{g}^m(\vec{e}) := \frac{g^m(\vec{e})}{r^m}. \tag{5.13}$$

The normalized resource allocation $\hat{g}^m(\vec{e})$ is a continuous variable $\hat{g}^m(\vec{e}) \in \mathbb{R}_0^+$ and can become arbitrarily small, when the number of independent source-destination paths becomes large. We have previously mentioned that a corresponding minimum cost routing solution consists of no more than a single Steiner tree. This is equivalent to a single data path between source and each destination node. In order to prevent the distribution of information on an arbitrary number of paths and to restrict it to a single path, the additional constraint
$\forall m = \{1, \ldots, M\}, \forall \vec{e} \in \mathcal{E}$:

$$\hat{g}^m(\vec{e}) \in \{0, 1\}. \tag{5.14}$$

is sufficient. The resource elements $\hat{g}^m(\vec{e})$ are thus defined as binary values and each resource element $g^m(\vec{e})$ is either not allocated, e.g. zero, or large enough to host a transmission with full data rate r^m, e.g. one. As a consequence of this restriction, any of the edge-disjoint data paths, which could potentially contribute to the solution, achieves rate r^m. Naturally, a minimization of the overall costs will lead to no more than single data path. Any alternative path or path segment will be related to equal or even higher costs, although the rate requirement is already fulfilled by the first path. This will be prevented by the minimization process. Therefore, the additional restriction of the linear network coding approach by Equation (5.14) leads to a tree structure as solution of any minimum cost multicast routing problem, but does not violate or restrict the optimality of this solution. The resource allocation for routing is a subset of the corresponding allocation for network coding:

$$\bigcup_{\vec{e} \in \mathcal{E} \,:\, \hat{g}^m(\vec{e}) \in \{0,1\}} \{g^m(\vec{e})\} \subseteq \bigcup_{\vec{e} \in \mathcal{E} \,:\, \hat{g}^m(\vec{e}) \in \mathbb{R}_0^+} \{g^m(\vec{e})\}. \tag{5.15}$$

The minimum cost routing solution is therefore a special case of the more general network coding solution.

5.4. Minimum Energy Broadcast Transmission

Problem Formulation Imagine a provider, who wants to offer news and advertisement channels to his subscribers and therefore needs to accomplish a broadcast transmission from a central server in the backbone network to all wireless infrastructure routers. The required amount of overall transmission resources for this broadcast is of course essential for a quantification of the spectral efficiency of the transmission processes in the network. A typical property of network coding is the distribution of information on independent data paths from the source node to a distinct wireless router. In contrast, minimum energy routing will rely on a single routing path. The question now is, whether network coding can indeed save valuable transmission resources. This process depends on the number of access points within the network, because the access point density influences the number of independent paths to the relay stations. We want to investigate the performance gains in terms of resource savings of network coding in typical cellular multi-hop environments with variable access point density. The final goal is to make a general statement for networks with a distinct access point density, whether the application of network coding results in significant performance gains or not. Let us therefore consider the following minimization problems of the overall resource allocation

(L1) network coding
minimize overall resource allocation
subject to
network coding constraints
rate assignment

(L2) routing
minimize overall resource allocation
subject to
routing constraints
rate assignment

with a fixed data rate assignment r. The overall resource allocation G is given by the sum of all resources, which are allocated on the corresponding virtual edges of the set of core nodes $\mathcal{V}^c$,
$\mathcal{V}^c = \bigcup_{V^c \in \mathcal{V}} V^c$:

$$G := \frac{1}{|\mathcal{V}^c|} \sum_{\vec{e} \in \mathcal{E}^{(out)}(\mathcal{V}^c)} \sum_{m=1}^{M} g^m(\vec{e}). \tag{5.16}$$

As already mentioned, the physical resource allocation is exclusively given by the sum of resources on the virtual edges, as any resource allocation on the broadcast edges is physically related to a distinct virtual edge due to the broadcast ability of the wireless medium. The simplified linear problem structure for a single broadcast session $M = 1$ is summarized by the following set of equations:

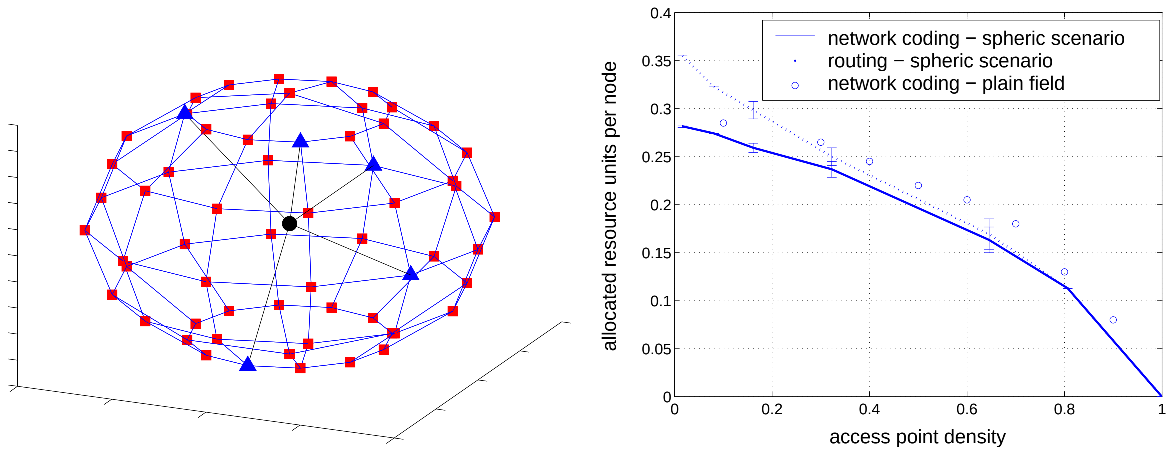
$$\begin{array}{lll}
\textit{minimize} & G := \frac{1}{|\mathcal{V}^c|} \sum_{\vec{e}\in\mathcal{E}^{(out)}(\mathcal{V}^c)} g(\vec{e}) & \\
\textit{subject to} & & \\
& \forall j \in \{1,\ldots,J\}, \forall \vec{e} \in \mathcal{E} : & 0 \leq f_j(\vec{e}) \leq g(\vec{e}) \\
& \forall \vec{e} \in \mathcal{E} : & g(\vec{e}) \leq c(\vec{e}) \\
& \forall j \in \{1,\ldots,J\} : & \sum_{\vec{e}\in\mathcal{E}^{(out)}(V_s)} f_j(\vec{e}) - \sum_{\vec{e}\in\mathcal{E}^{(in)}(V_s)} f_j(\vec{e}) = r \\
& \forall j \in \{1,\ldots,J\}, \forall V \in \mathcal{V}\backslash\{V_s, V_{d_j}\} : & \sum_{\vec{e}\in\mathcal{E}^{(out)}(V)} f_j(\vec{e}) - \sum_{\vec{e}\in\mathcal{E}^{(in)}(V)} f_j(\vec{e}) = 0 \\
& & f_j(\vec{e}), g(\vec{e}) \in \mathbb{R}_0^+ \\
& \text{additional constraint for routing:} & \\
& \forall \vec{e} \in \mathcal{E} : & \hat{g}(\vec{e}) = \frac{g(\vec{e})}{r}, \hat{g}(\vec{e}) \in \{0,1\}.
\end{array}$$

Please take a note on the fact that the presented network coding problem is a LP, while the routing problem represents a MILP. The complexity of the routing problem is therefore much higher than the complexity of the network coding problem. We will analyze both problems by an example scenario in the following.

Simulation Scenario Assume a cellular multi-hop network to be modeled by a randomly meshed network topology with $|\mathcal{V}^c| = 62$ stations, mesh degree $\gamma = 3.7$, and $c(\vec{e}) = 1$ rate unit for all edges. The topology is deployed spherically as shown in Figure 5.4(a). A node in the center of the sphere is acting as source node and connected to a variable, randomly chosen number of access points $\rho \cdot |\mathcal{V}^c|$ among all nodes. Each node has wireless links to some neighbor nodes and therefore may optionally transmit in unicast or broadcast mode to its neighbors according to the given connectivity. For this reason, we consider the wireless multicast graph model as presented in Section 5.3.1 for the subsequent simulations. As the resulting network graph is difficult to sketch, we renounce on a corresponding illustration and refer to the simplified model in Figure 5.4(a). To begin with, a broadcast flow with fixed rate $r = 1$ rate unit is distributed from the central node to the access points and all other relay stations. The connections from the central source node to the access points are supposed to be fiber and therefore not accounted in terms of resources. The given results represent the average values of a series of four simulations. Random topologies with variable access point densities $0 < \rho \leq 1$ have been evaluated in each simulation.

Discussion of Results Figure 5.4(b) reveals that the number of allocated average resources varies between $G = 0$ rate units per node and $G = 0.35$ rate units per node in dependence on the transport scheme and the number of access points. It is very obvious that no additional resources are required in case of all nodes being access points ($\rho = 1$), i.e. $G = 0$ rate units per node for routing and network coding. There is few possibility of distributing information on many different paths by network coding for only few remaining relay nodes ($0.6 \leq \rho \leq 1.0$) such that routing also performs about equally well in this case. However, once the number of access points is decreased, network coding outperforms routing in terms of required resources by about 10% for $\rho = 0.2$. The maximum performance gap equals about 20% for only a single access point, i.e. $\rho \to 0$. We can explain this observation by the ability of network coding to distribute information on various data paths on its way to the destination. Network coding can thus achieve synergy effects, which yield savings in transmission resources. Figure 5.4(b) also shows some sample results of a Manhattan-like plain field network with a regular deployment of 10 x 10 = 100 nodes. It turns out the resource demand per node for the application of network coding in the broadcast case is higher in this case. This can be explained by border limitations, where the gain by combining several flows is typically smaller in the border area of the simulation grid compared to the spheric scenario. Due to complexity reasons — especially for the routing evaluation with integer constraints — it is hard to choose a plain field scenario in a way that those border effects are neglibible. This is the major motivation for the spheric simulation.

Summary The presented results indicate that the use of network coding in wireless multi-hop networks makes only sense in case of a high relay stations to access points ratio. According to these results, we expect performance gains in terms of resource savings of about 10% for an access point density $\rho = 0.2$, i.e. four relay stations per access point on average. Typically, a performance gain around 10% is considered to be significant from the technical perspective and justifies the introduction of new technologies. Therefore, the application of network coding seems to be promising for broadcast applications in cellular multi-hop networks.



(a) Spheric example scenario with 57 relay nodes, 5 access points and one central backbone node.

(b) Average resource allocation per node to enable a rate r=1 rate unit broadcast to all nodes for routing and network coding.

Figure 5.4.: Spheric example scenario and evaluation results.

5.5. Network Coding in Bidirectional Relaying Scenarios

It has been illustrated so far that network coding can achieve valuable resource savings in the broadcast case. A special application of networking coding is the bidirectional relaying scenario, where two stations use an intermediate relay station in between to mutually exchange data. The investigation of this case is especially interesting in cellular multi-hop networks, as a set of a downlink and associated uplink connection between a mobile terminal and a distinct access point fulfills all characteristics of the bidirectional relaying scenario. In the simplest case, there is only one intermediate relay station in between source and destination terminal. This type of link connectivity is described by a model, which is typically referred to as two-way relay channel in the literature. The achievable performance mainly depends on three assumptions, namely the intermediate device acting in half-duplex or full-duplex mode, the connectivity model itself, and the question, whether the packet is decoded at the intermediate node (decode-and-forward) or just amplified (amplified-and-forward). The corresponding capacity theorems for the discrete memoryless relay channel published in [CEG79] are the basis for most performance investigations. The general two-way relaying model as shown in Figure 5.5(a) assumes the two outer stations within mutual transmission range. This constellation enables diversity at the receiver, the exploitation of which is known as cooperative relaying. In contrast, a simplified relaying model assumes both stations outside the mutual transmission range as illustrated in Figure 5.5(b).

Network Coding for Symmetric Relaying Imagine a typical situation in a cellular environment, where the coverage of an access point is extended by a relay station. A

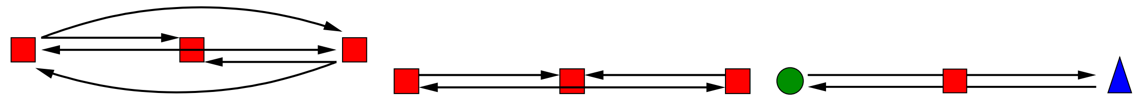

(a) Link connectivity for cooperative relaying. (b) Simplified link connectivity for bidirectional relaying. (c) Example with relay acting as coverage extension for access point - mobile terminal connection.

Figure 5.5.: Connectivity models for simple relaying scenarios.

mobile terminal is connected to the relay station and establishes a connection to the access point. We consider the case of the simple relaying model in Figure 5.5(b), where the direct transmission is infeasible. It is obvious that collision-free routing requires two transmissions per downlink message x_d and further two transmissions per uplink message x_u. The bottleneck in this case is the relay station, which receives both messages and forwards them independently afterwards. Although the scenario includes two unicast data flows, network coding can be applied to achieve performance gains in this case. More generally speaking, this is possible for a symmetric pair of oncoming streams over a chain of relay stations. [WCK05a] explains, how such a scenario can be expressed by a multicast problem formulation. In practice, the intermediate relay station can forward an encoded message $x_u \oplus x_d$ and thus save one transmission. The encoded message can be decoded by the access point and the mobile terminal, as alternatively x_u or x_d are known. A throughput analysis for the amplify-and-forward scheme as well as for the decode-and-forward scheme is given in [PH06].
One of the major problems to this kind of encoded relaying is the property that uplink and downlink traffic loads are typically asymmetric, i.e. the expected downlink traffic is probably higher than the corresponding uplink traffic. Consequently, only a part of the oncoming data flows can be encoded such that the performance gains are restricted to the minimum of both, uplink and downlink traffic. Another even more serious problem are asymmetric channel conditions. Let us therefore consider the example of Figure 5.5(c), where the channel conditions between access point and relay station are presumably much different from the conditions between relay station and mobile terminal. While access point and mobile terminal can adjust their power for the individual transmission to the relay station, the encoded message from the relay station has to be transmitted obviously for both receiver stations with same power level and PHY mode. Due to these implications, the application of network coding for relaying purpose in cellular multi-hop networks is feasible, but even more favorable in networks with about equal link conditions. The infrastructure part of the cellular multi-hop network offers good conditions to this.

Analog Network Coding So far, network coding has been presented as a digital processing technique, where incoming data flows are digitally combined and separated again. However, ongoing research activities for instance in [ZLL06] have shown that it is also possible to combine data flows in an analog manner. This analog network coding or also known as physical layer network coding is a technique, where simultaneously incoming transmission signals are added up physically at the receiver. A spectral analysis for the half-duplex relay channel has been given in [RW05] and later compared with routing in [KMG+07]. [KMG+07] has shown that network coding doubles the achievable throughput for transmissions at a considerably high SNR. Typically this assumption holds in cellular systems with reliable communication links. Another publication in [KGK07] has illustrated, how analog network coding can deal with some lack of synchronization or frequency offsets. As the application of analog network coding does not involve much signaling overhead, this technique seems to be promising as well as applicable in practice as extension of bidirectional relaying.

Cooperative Relaying A performance analysis for the cooperative two-way relay channel model in Figure 5.5(a) can be found in [LJS06]. This work assumes that the incoming packets at the intermediate relay nodes are separately decoded and jointly encoded afterward. Cooperative relaying is characterized by the property that the destination node can also receive the transmission between sender and intermediate node. The encoded packet is transmitted again and the soft-information from these two transmissions is used to make the decoding decision at the final receivers more reliable. The major problem with this kind of technique is the question, how to make this soft information applicable in a real multi-user scenario [DMS08]. Especially when many terminals are involved, the application of cooperative relaying goes on with a loss of degrees of freedom in scheduling. This is due to the requirement that exact knowledge about the scheduling is a prerequisite for the soft-combining of information at the receiver. In other words, the receiver has to know, which transmissions are intended to complement each other. It is not so easy to provide this information, as we cannot assume that the final receiver is able to decode the header information from the first transmission between sender and intermediate station. Though, even if this scheme is very interesting from a theoretical point of view, cooperative relaying is difficult to realize in a practical cellular multi-hop network.

Summary To sum up, the case of bidirectional relaying is an interesting application to cellular multi-hop networks. Due to the joint transmission of uplink and downlink by the relay station, the spectral efficiency of the data transmission is evidently increased compared to the conventional multi-hop transmission scheme. There is huge potential, especially for the technique of analog network coding. Nevertheless, there are also practical hurdles in terms of signaling efforts for the multi-user case and asymmetric link conditions, which may make the joint transmission difficult. Due to its static link conditions, the infrastructure part of the cellular multi-hop network seems to be especially suited for the application of network coding.

5.6. Conclusion

In this chapter, we have addressed network coding as a generalization of the routing concept and especially focused on the potential application of network coding to cellular multi-hop networks. We have identified two interesting fields of application, namely the broadcast case and the case of bidirectional relaying. We were able to show for the broadcast case that network coding indeed yields resource savings for a broadcast transmission from a backbone node to all nodes of the wireless infrastructure network. The amount of savings depends on the access point density of the network. This density shall not exceed $\rho = 0.2$ in order to obtain significant resource savings of more than 10% in the given network with mesh degree $\gamma = 3.7$. We have reached maximum savings of about 20% compared to routing in the example network for the case, in which only a single access point has been deployed. In contrast, routing and network coding perform about equally well for access point densities $\rho \geq 0.8$. It is important to record at this point that the solution of the routing problem is by far more complex than the solution of the corresponding network coding problem. This relates to the issue that the routing problem is defined by a MILP, while network coding is described by a conventional LP. As a direct consequence of the mathematical models, routing cannot outperform network coding, as the solution of the routing problem is a constrained subset of the more general network coding solution. Apart from the broadcast case, bidirectional relaying is especially interesting for the application to cellular multi-hop networks, as each multi-hop connection between an access point and a mobile terminal respresents a bidirectional relaying use case. Focusing on bidirectional relaying, particularly the advanced scheme of analog network coding emerges as a promising idea to additionally increase the spectral efficiency of cellular multi-hop communication. Finally, please take notice of the fact that the subject of network coding is even larger and includes also further research areas such as for instance random network coding, which have not been considered in this work.

6. Summary and Outlook

6.1. Results and Contribution

This thesis includes a comprehensive investigation into the operation of a cellular multi-hop network. In Chapter 2, we have discussed arguments and possible obstacles for the extension of cellular networks by relay stations. In Chapter 3, we have presented Dynamic Cell Clustering as an approach to partition an unstructured wireless infrastructure network into multi-hop cells. In Chapter 4, we have developed a routing protocol to enable intra-cell mobility within a multi-hop cell. Finally, we have examined network coding as a promising data transport technique regarding its application to cellular multi-hop networks in Chapter 5.

Reasons for the Relaying Concept Initially, we have outlined in Chapter 2 that any use of relay stations is primarily economically motivated, as there is no technical justification for the choice of a relay station instead of an access point. Relay stations cannot replace access points adequately, as the relay links back to the access point always consume valuable transmission resources such that a relay station cannot directly compete with an access point. Thus, the relay station remains either a cheap substitute for an access point or a temporary placeholder in the start-up phase of the network. Nevertheless, we have shown that given the strict economic constraints of today's network providers, the reduction of CAPEX and OPEX by the use of relay stations is a very valuable contribution. The deployment of relay stations is therefore a technical approach for conventional cellular operators to compete against distributed low-cost technologies as WLAN or mesh networks in an adequate and cost-efficient manner.

Dynamic Cell Clustering Subsequently, we have presented Dynamic Cell Clustering as a concept to turn a set of access points and relay stations into a cellular network. This strategy follows the idea of a dynamic clustering process, which depends on the network's current load state and thus combines clustering and load balancing. We have explained, how this mechanism helps to cut down CAPEX and OPEX of a network operator. Moreover, we have investigated the clustering process in various example networks and have developed a numerical optimization approach and a greedy heuristic for that purpose. Our results indicate that Dynamic Cell Clustering manages to balance the load efficiently in network topologies with a mesh degree $\gamma \geq 3$. The effect of balancing is also reflected in a significantly lower maximum load per cell. It has been shown that this reduction corresponds to 20%-30% of the original load on average, but can amount to more than 50% in some cases. It has turned out that the greedy heuristic yields good performance results and can also deal with low to moderate network dynamics. Consequently, it is not necessary to include a central instance for the clustering process. Instead, clustering can be accomplished in a self-organizing and distributed fashion. Based upon these promising results, we have developed a protocol stack for the centralized and the decentralized clustering approach. A prototypical simulation result underscores that Dynamic Cell Clustering is also practically feasible in cellular networks.

Intra-Cell Routing In Chapter 4, we have presented existing routing protocols from the fixed and ad hoc network domain. To begin with, we have investigated these protocols regarding their qualification as intra-cell routing approach in a multi-hop cell. Subsequently, we have examined the effects of different routing metrics on the resource demand and especially focused on the question, whether a simple hop count is a sufficient path metric to deal with wireless multi-hop networks. It has been shown that this minimum hop strategy can achieve good performance results, but tends to the formation of bottlenecks. Additionally, we have illustrated that the efficiency of a routing metric primarily depends on topological characteristics of the network. The physical link conditions of the relay links and the overlapping transmission area between neighboring routers are thereby key parameters for choice of an adequate routing metric. In the next step, we have discussed the issue of proactive and reactive routing. We have presented a proactive intra-cell routing approach and explained, why the design of a routing protocol is strongly related to the MAC protocol. An inter-layer exchange of link state information can that way reduce the signaling overhead to a fraction of the original amount. Proactive routing is consequently not necessarily related to higher overhead than reactive routing. In the following, we have complemented the proactive topology exploration protocol with a label switching scheme as suitable addressing mechanism for the data delivery within a multi-hop cell. The given overall protocol stack provides users with intra-cell mobility and is compliant to the current LTE inter-cell handover approach. The proposed protocol represents therefore a suitable approach to handle intra-cell mobility in relay-extended cellular networks.

Network Coding in Cellular Multi-Hop Networks Finally, we have investigated network coding as a generalization of the routing concept and examined its applicability to cellular multi-hop networks in Chapter 5. Our results indicate that network coding yields resource savings in the broadcast case. Savings of about 10% or higher are feasible for an access point density $\rho \leq 0.2$. This corresponds to a required average amount of at least four relay stations per access point and thus is a realisitic value. Performance gains of more than 10% are considered to be significant from the technical perspective such that the application of network coding for broadcast scenarios in cellular multi-hop networks is reasonable. Although network coding is feasible in networks with larger access point densities, we have shown that its application remains practically without impact on the performance compared to routing then. Finally, we have considered the case of bidirectional relaying and discussed its relevance for cellular multi-hop networks.

6.2. Outlook

Paradigm of Self-Organization In future work, we could imagine an extension of the principles of self-organization in the Dynamic Cell Clustering process. Especially the direct interaction between the clustering unit and the radio resource management unit should be investigated in detail to combine resource allocation and load balancing in cellular multi-hop networks. As also mentioned in the thesis, cell border areas are a special challenge for any radio resource management. Therefore, it is additionally necessary to examine the interaction between the intra-cell routing and the radio resource management unit to extend the self-organization skills of the radio resource management. A step into this direction is accomplished by Jan Ellenbeck in [EHB08].

Spectral Efficiency of the Relay Concept As pointed out within the thesis, the efficiency of the relay links is very important for the overall success of the relay-based network. The special characteristic of the relay link implies that this link is typically subject to worse propagation conditions than above roof top microwave connections. However, better channel conditions than the links to regular mobile terminals are definitely required, as otherwise the relaying advantage is absolutely questionable. Obviously, this is a borderline scenario with chances for gains, but also with risks of failure. It would be therefore very helpful and also important to examine a single two-hop or multi-hop link connection with realistic and sophisticated ray-tracing propagation models. This thesis can only give advice, how typical link conditions should behave in order to get a benefit from the relay topology. A detailed channel analysis can assess beyond that, whether the conditions of a realistic link scenario fulfill the outlined requirements. We therefore propose an investigation of the three main constellations, namely the relay station positioned above roof top, the relay mounted on the traffic light, and relay positioned randomly on street level.

Optimized Link Configuration Electronic beamforming can help to improve the link conditions within the wireless infrastructure network. The definition of scanning intervals and a coherent concept for the estimation of the angle of arrival are necessary requirements. A subsequent, dynamic connectivity optimization of a randomly meshed infrastructure topology is one interesting objective of research. The optimization could be accomplished in a similar framework as presented by Christian Hartmann in [HKV08].

Hierarchical Handover The integration of GTP in the simulator could help to get accurate values for the handover delay. It is a question of special interest to investigate the real-time handover of a relay station including its attached mobile terminals from one base station to another base station. The handover problem is also related to security issues, which typically arise with the association and authentication process of a terminal at a router. So-called hash chains could be applied to cipher step-by-step the end-to-end links within the relay-extended multi-hop cell, as talks with Stephan Eichler have shown. The hash chain mechanism is able to accelerate the intra-cell, but potentially also the inter-cell handover process. It is definitively worth looking on the security requirements of such a system and finally developing a link-level simulator, which performs a real-time intra-cell handover and real-time hierarchical inter-cell handover including the required security features.

Summary To make a concluding overall statement, cellular multi-hop networks appear as a very promising network architecture to establish high-performance communication networks in densely populated areas. Relay stations potentially represent the missing jigsaw piece to unify the architecture of mesh networks and cellular networks for the convergence into future high data rate networks. As illustrated in the initial Figure 2.1, the relay technology is already an important component in the 4G convergence process, where different approaches complement each other in order to finally provide an overall set of services. The future development of the LTE standard will primarily influence research on relay stations and determine, whether the cellular relay technology will evolve from theoretical concepts into practical deployments.

A. Simulation Model

A.1. Environment Model

Mobility Model The set of mobile terminals $\mathcal{V}_M$ is subject to mobility according to the random direction mobility model developed in [RMSM01]. Research in [PLBV05] has shown that the stationarity of the node and velocity distribution is a major requirement to any mobility model, if statistically correct simulation results shall be accomplished. The random direction mobility model foresees each node to choose a random movement direction from $\theta \in [0, 2\pi]$ and a corresponding speed $v \in [0, v_{max}]$. Whenever a node touches the simulation boundary, a new speed and direction are chosen. If the new direction leads outside the simulation area, the process of determining a direction is repeated, until a valid direction is found. The example in Figure A.1 illustrates a potential movement of a user terminal. The terminal starts its movement in direction θ_1 at position X_1 and travels with speed v_1. Once the node reaches the simulation boundary featured by the building wall at position X_2, the node choses a new direction θ_2 and speed v_2, until it arrives at position X_3. At position X_3, direction and speed are updated anew, before the journey is continued. [Bet01] has shown that such a model results in a uniform spatial node distribution. The simulation lasts for an overall duration T. An additional travel duration $t \in [0; t_{max}]$ with $t_{max} \ll T$ is introduced to update a user's speed from time to time. We have typically chosen $t_{max} = 0.1 \cdot T$ for the simulations. The described mobility model results in a uniform node and velocity distribution and is thus compliant with the demands to a perfect simulation according to [PLBV05].

Channel Model An adjusted version of the free space propagation model [Fri46] is taken into account for the calculation of the path loss between neighboring stations. The channel model does not include any slow or fast fading, multi-path propagation, or any other physical layer-related effects. The need for this simple model is founded in the dramatic processing overhead, which more detailed channel models cause in system level simulations. According to this simple transmission model, the received signal power $P_r(d)$ is dependent on the transmission power P_s and the distance d between sender and transmitter:

$$P_r(d) = P_s \cdot a \cdot \left(\frac{c}{4\pi f}\right)^2 \frac{1}{d^\delta}, \tag{A.1}$$

whereby the carrier frequency $f = 5.3$ GHz and the speed of light $c = 3 \cdot 10^8$ m/s are considered. A path loss coefficient $\delta = 2$ is used for LOS transmissions. Additionally, we assume that for all transmissions in nLOS $\delta = 4$ is sufficient to model the propagation characteristics adequately. The differentiation between LOS and nLOS con-

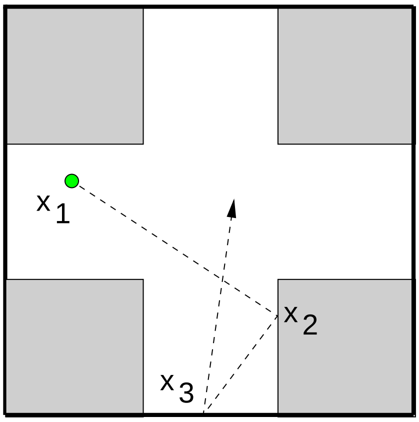


Figure A.1.: Illustration of a potential movement defined by the random direction mobility model.

nections allows the provision for buildings and obstacles, which accounts for a more realistic simulation environment. The antenna gain a is a multiplier, which depends on the antenna characteristics of the sender and receiver device.

Traffic Model A constant bit rate traffic generator acts as source of defined data flows. The traffic model is based on a given rate requirement r and a predefined fixed packet length l_p per data flow. The ratio of rate r and packet length l_p describes the required average packet rate r_p:

$$r_p := \frac{r}{l_p}. \tag{A.2}$$

The inverse of the packet rate represents the average time interval between two subsequent data packets:

$$T_p := \frac{1}{r_p}. \tag{A.3}$$

A single data packet is transmitted periodically per time frame T_p, whereby its exact transmission time is uniformly distributed within the interval. The transmission time $t(p_k)$ of the k^{th} packet p_k is therefore calculated by

$$t(p_k) := k \cdot T_p + t_k \tag{A.4}$$

with the uniformly distributed random variable $t_k \in [0, T_p]$. The average number of packets across the overall simulation time therefore follows the intended data rate requirement.

Obstacle Model We assume a city center with streets and buildings as suitable environment for the application of cellular multi-hop networks. We model this environment by a regular Manhattan grid topology as suggested in [ETS98]. The scenario implies that mobile terminals cannot enter the buildings such that the simulation area is bounded by the building walls and the border of the squared simulation grid.

A.2. Device Model

The wireless devices are realized according to the OSI reference model. The functionality of the single layers in the simulation model is implemented as follows:

Physical Layer The basic task of the physical layer model includes the transmission and reception of data packets over the wireless channel. The correct reception of a packet presumes two major prerequisites: First of all, the received signal power P_r needs to outnumber a given minimum threshold power P_{thres}. Secondly, the SNR has to reach a minimum level of $\gamma_{thres} = 1.95$ dB. The SNR of a received packet is calculated as

$$\gamma = 10 \log \left(\frac{P_r}{P_n} \right). \tag{A.5}$$

The data packets are transmitted with individual data rates, which are adapted to the instantaneous link conditions. An optimum link adaptation process without feedback is accomplished for that purpose. The available SNR is mapped to an achievable physical data rate r_{phy} by a corresponding link adaptation function:

$$r_{phy} = \begin{cases} 54\,\text{MBit/s} & \text{for} \quad \gamma \geq 25\ \text{dB} \\ 36\,\text{MBit/s} & \text{for} \quad 25\ \text{dB} > \gamma \geq 19\ \text{dB} \\ 27\,\text{MBit/s} & \text{for} \quad 19\ \text{dB} > \gamma \geq 14\ \text{dB} \\ 18\,\text{MBit/s} & \text{for} \quad 14\ \text{dB} > \gamma \geq 11\ \text{dB} \\ 12\,\text{MBit/s} & \text{for} \quad 11\ \text{dB} > \gamma \geq 8\ \text{dB} \\ 9\,\text{MBit/s} & \text{for} \quad 8\ \text{dB} > \gamma \geq 4\ \text{dB} \\ 6\,\text{MBit/s} & \text{for} \quad 4\ \text{dB} > \gamma \geq 2\ \text{dB} \\ 0\,\text{MBit/s} & \text{else.} \end{cases} \tag{A.6}$$

By definition, broadcast packets are always transmitted in the lowest PHY mode, i.e. at a rate of 6 MBit/s. As shown in Figure A.2, the given link adaptation function results from a stepwise approximation of the envelope of a HiperLAN/2 link adaptation curve. This graph has been presented in [KJSMT01] and assumes LOS channel conditions.

Link Layer ZERO-MAC, a simplified MAC module, has been developed in order to perform system level simulations. Zero-MAC emulates a centralized MAC and replaces the MAC at access points and relay stations by transmission queues for that purpose. The MAC layer at mobile terminals is fully transparent, i.e. the packets simply pass the MAC layer without further processing apart from a mandatory check of the layer 2 header information. The queuing system provides one separate queue

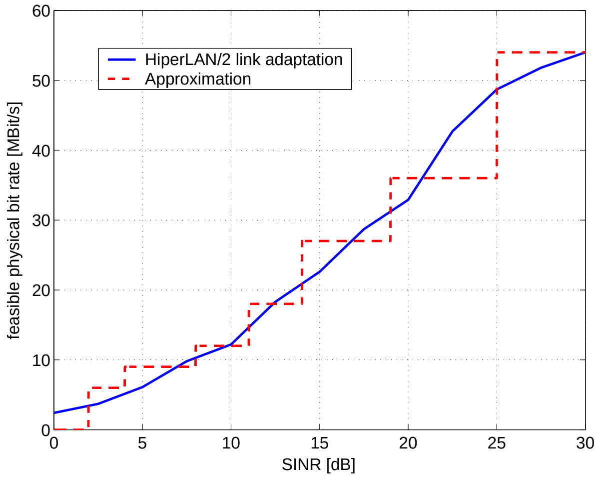


Figure A.2.: Stepwise approximation of the envelope of the HiperLAN/2 link adaptation curve.

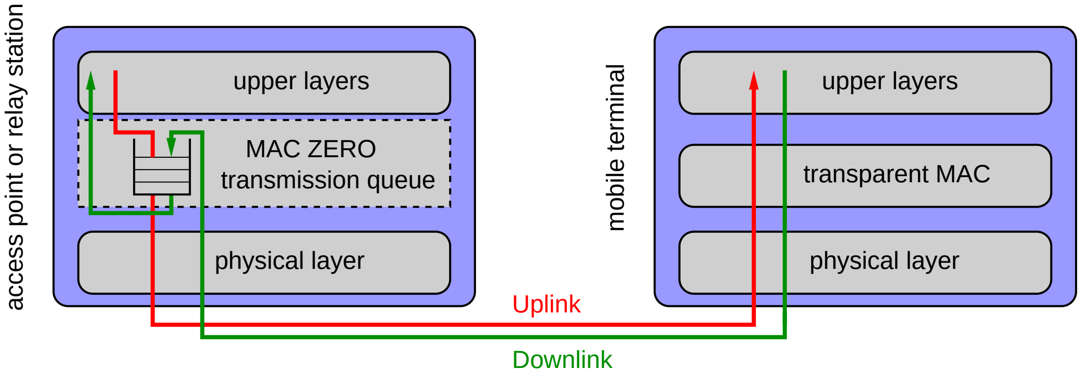


Figure A.3.: Transmission queues are attached to access points and relay stations. Each packet passes exactly one transmission queue, which models the finite transmission capacity of the wireless medium and its optimum assignment.

to each wireless router. In doing so, any queue represents nothing else but an exclusively assigned transmission resource or channel. ZERO-MAC acts optimum in a way that the common transmission resource is shared without collisions. Incoming packets in the uplink or downlink direction are subsequently enqueued in a fixed number of buffer slots. Packet losses may occur, once the buffer is overloaded. After being queued, the packets await the further processing. The queue is characterized by a dynamic service time function, which subsequently processes and transmits each packet within a specific service interval. This duration is given by the quotient of packet length and the available physical data rate between sender and receiver. Consequently, the process is very similar to a TDMA system, in which users are scheduled sequentially request by request.

We distinguish between downlink and uplink transmission as shown in Figure A.3. In the downlink case, a packet arriving from upper layers is enqueued in the transmis-

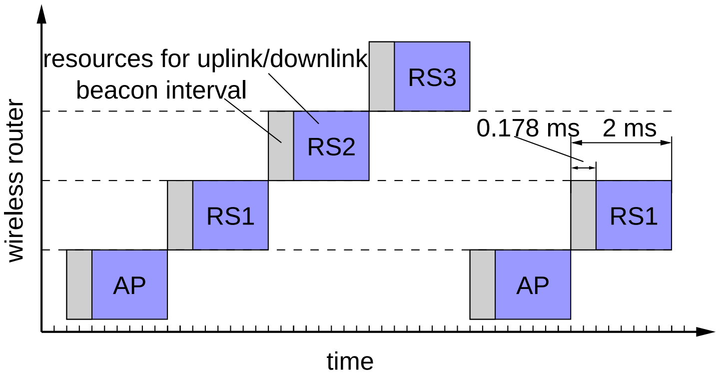


Figure A.4.: The radio resource management unit assigns transmission resources exclusively in a periodic superframe structure.

sion queue of the distinct access point or relay station. Once it is ready for transmission, it is sent down to the PHY layer and transmitted immediately over the wireless channel as sketched in Figure A.3. At the receiver, the packet is only checked for the correct header address information and handed on to the higher layers without any specific processing. In the uplink case, the packet moves down the layers of the mobile terminal, before being propagated on the wireless channel. Subsequently, the packet arrives at the distinct access point or relay station. Again, the uplink packet is enqueued at the MAC layer on the receiver side and will be received from higher layers not until passing the transmission queue. The uplink case is also shown in Figure A.3. If two routers want to communicate mutually, the packet is always placed in the transmitter's queue, i.e. the transmitter allocates resources for the transmission process. The packet consequently passes the MAC layer at the receiver node without entering the MAC queue then.

A radio resource management unit allows the global allocation of resources and assures the synchronization of all queues within the network. In doing so, the radio resource management unit can block distinct transmission queues periodically and impose a fixed frame structure on each router. Suppose the transmission medium to be shared among all routers of a multi-hop cell in such a way that a distinct router has exclusive medium access available at a certain instant of time. A corresponding resource sharing could look like the medium access realization sketched in Figure A.4, in which the transmission queues of access point and relay stations share the medium equally. For that purpose, we assume a 2 ms frame duration and a periodic superframe structure of $(|\mathcal{V}_A| + |\mathcal{V}_R|) \cdot 2$ ms. Within the 2 ms frame duration, a signaling overhead period of 178 μs is virtually reserved, which corresponds to the typical administrative signaling overhead in the HiperLAN/2 MAC specification. In consequence, each router may use the medium for uplink and downlink traffic during 2 ms - 0.178 ms = 1.288 ms per superframe interval. Apart from the MAC module and the radio resource management entity, the ARP unit is embedded in the link layer to resolve IP addresses into MAC addresses.

Network Layer and Application Layer The network layer includes a joint routing and clustering unit to enable the data transport in a cellular multi-hop network. Additionally, the standard IPv4 protocol stack and the corresponding Mobile IPv4 mobility framework are embedded. On top, a UDP client is established, which is connected to the constant bit rate traffic generator.

B. Parameter Setup — Overview

This section summarizes the simulation parameter setup:

γ	mesh degree
η	percentage of load change
ρ	access point density
τ	traffic load
N_{sim}	number of simulations
P_n	noise power
P_r^{thres}	threshold received power
P_s	transmission power
T	simulation duration
$\mathcal{V}$	number of nodes
$\mathcal{V}_A$	number of access points
$\mathcal{V}_R$	number of relay stations
$\mathcal{V}_M$	number of mobile terminals
Δt	approximation interval
a^{mob}	antenna gain for mobile links
a^{stat}	antenna gain for relay links
f_{load}	frequency of load change
l_b	grid block length
l_p	packet length
l_s	grid street width
r	data rate
d_{max}	maximum transmission range
r_{hello}	beacon frequency
v_{max}	maximum user velocity

Table B.1.: Chapter 3.4. Dynamic Cell Clustering (graph model).

Figure	N_{sim}	$\mid\mathcal{V}\mid$	$\mid\mathcal{V}_A\mid$	$\mid\mathcal{V}_R\mid$	γ	ρ	f_{load}	η
3.8(a)	300	28	6	22	3	0.2	0.0	0.0
3.8(b)	300	28	6	22	3	0.2	0.0	0.0
3.9(a)	300	360	72	288	2-8	0.2	0.0	0.0
3.9(b)	300	360	72	288	2-8	0.2	0.0	0.0
3.10(a)	300	360	36-360	0-324	4	0.1-1.0	0.0	0.0
3.10(b)	1	120-720	24-144	96-576	4	0.2	0.0	0.0
3.11(a)	10	360	72	288	4	0.2	0.0-1.0	0.05
3.11(b)	10	360	72	288	4	0.2	0.0-1.0	1.0

Table B.2.: Chapter 3.5. Dynamic Cell Clustering (ns-2 simulation).

Figure	N_{sim}	$\mid\mathcal{V}\mid$	$\mid\mathcal{V}_A\mid$	$\mid\mathcal{V}_R\mid$	T	l_b	l_s
3.14	1	12	4	8	200s	120m	60m
τ	l_p	P_s	P_n	a^{mob}	a^{stat}	d_{max}	P_r^{thres}
100kbyte/s	512byte	0.281W	$1{\cdot}10^{-10}$W	1	16^2	90m	$7{\cdot}10^{-10}$W

Table B.3.: Chapter 4.4. routing metrics (general parameter setup).

N_{sim}	$\mid\mathcal{V}\mid$	$\mid\mathcal{V}_A\mid$	$\mid\mathcal{V}_R\mid$	$\mid\mathcal{V}_M\mid$	T	Δt	l_b	l_s	v_{max}	r
5	39	1	8	30	250s	10ms	120m	60m	10m/s	16kbit/s

Table B.4.: Chapter 4.4. routing metrics (individual parameter setup).

Figure	P_s	P_n	a^{mob}	a^{stat}	d_{max}	P_r^{thres}
4.3(a)	0.281W	$1 \cdot 10^{-10}$W	1	16^2	120m	$4 \cdot 10^{-10}$W
4.3(b)	0.281W	$1 \cdot 10^{-10}$W	1	16^2	120m	$4 \cdot 10^{-10}$W
4.4(a)	0.281W	$1 \cdot 10^{-10}$W	1	$\{2^2;4^2;16^2\}$	120m	$4 \cdot 10^{-10}$W
4.5(a)	0.281W	$1 \cdot 10^{-10}$W	1	16^2	120m-180 m	$\{1.76;2.5;4.0\}{\cdot}10^{-10}$W
4.4(b)	0.281W	$1 \cdot 10^{-10}$W	1	16^2	120m	$4.0 \cdot 10^{-10}$W

Table B.5.: Chapter 4.5. topology exploration (general parameter setup).

P_s	P_n	a^{mob}	a^{stat}	P_r^{thres}
0.281W	$1 \cdot 10^{-10}$W	1	16^2	$4 \cdot 10^{-10}$W
T	l_b	l_s	r	l_p
300s	120m	60m	16kbit/s	512byte

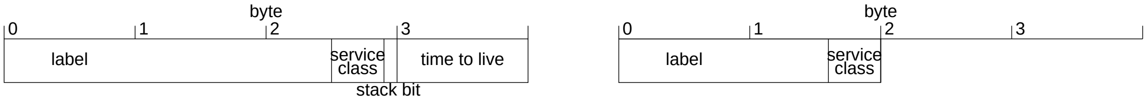
Table B.6.: Chapter 4.5. topology exploration (individual parameter setup).

Figure	N_{sim}	$\mid\mathcal{V}\mid$	$\mid\mathcal{V}_A\mid$	$\mid\mathcal{V}_R\mid$	$\mid\mathcal{V}_M\mid$	v_{max}	r_{hello}
4.10	5	39	1	8	30	{3;10;20}m/s	1.0s^{-1}
4.11	5	39	1	8	30	10m/s	{0.1;1.0;10.0}s^{-1}
4.12	5	{19;29;39}	1	8	{10,20,30}	10m/s	1.0s^{-1}

Table B.7.: Chapter 5.4. network coding.

Figure	N_{sim}	$\mid\mathcal{V}\mid$	$\mid\mathcal{V}_A\mid$	$\mid\mathcal{V}_R\mid$	γ	r
5.4(b)	4	62	1-62	0-61	3.7	1

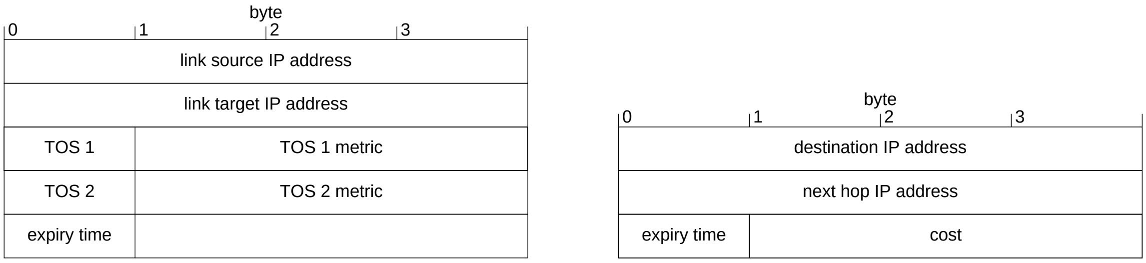
C. Packet Definitions



(a) MPLS packet header format.

(b) Cellular multi-hop LS header format.

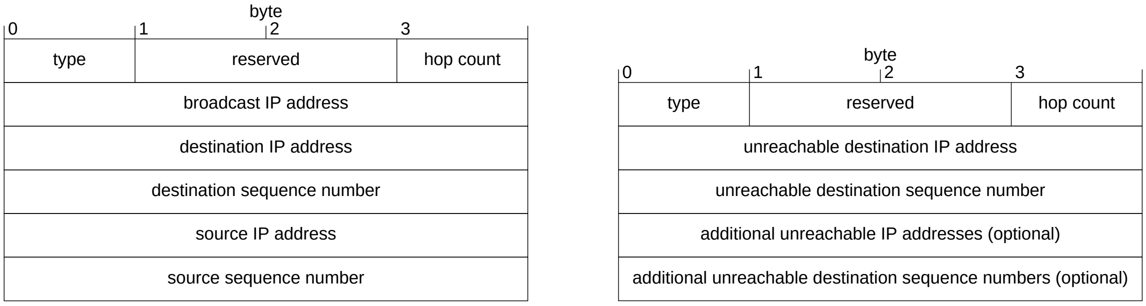
Figure C.1.: MPLS header and cellular multi-hop label switching header.



(a) Topology table entry.

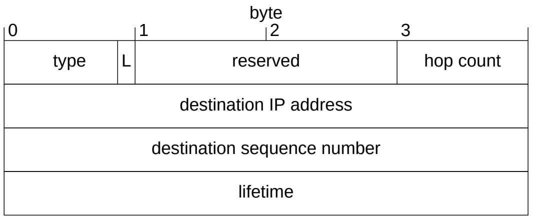
(b) Routing table entry.

Figure C.2.: Topology table and routing table entry.



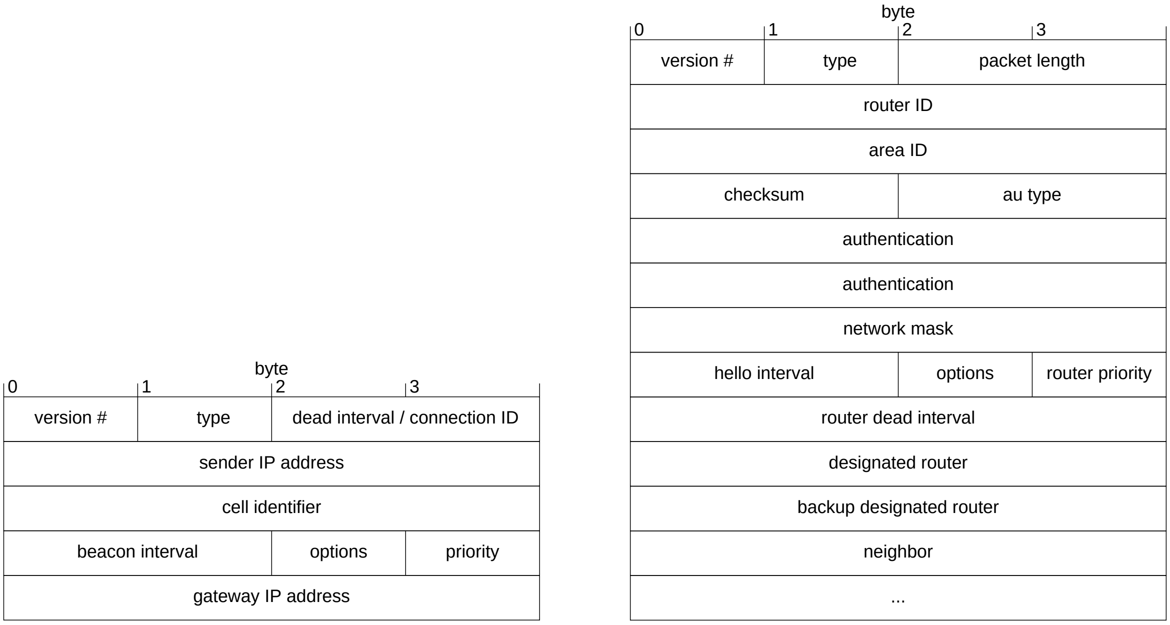
(a) AODV packet format: route request.

(b) AODV packet format: route error.



(c) AODV packet format: route reply.

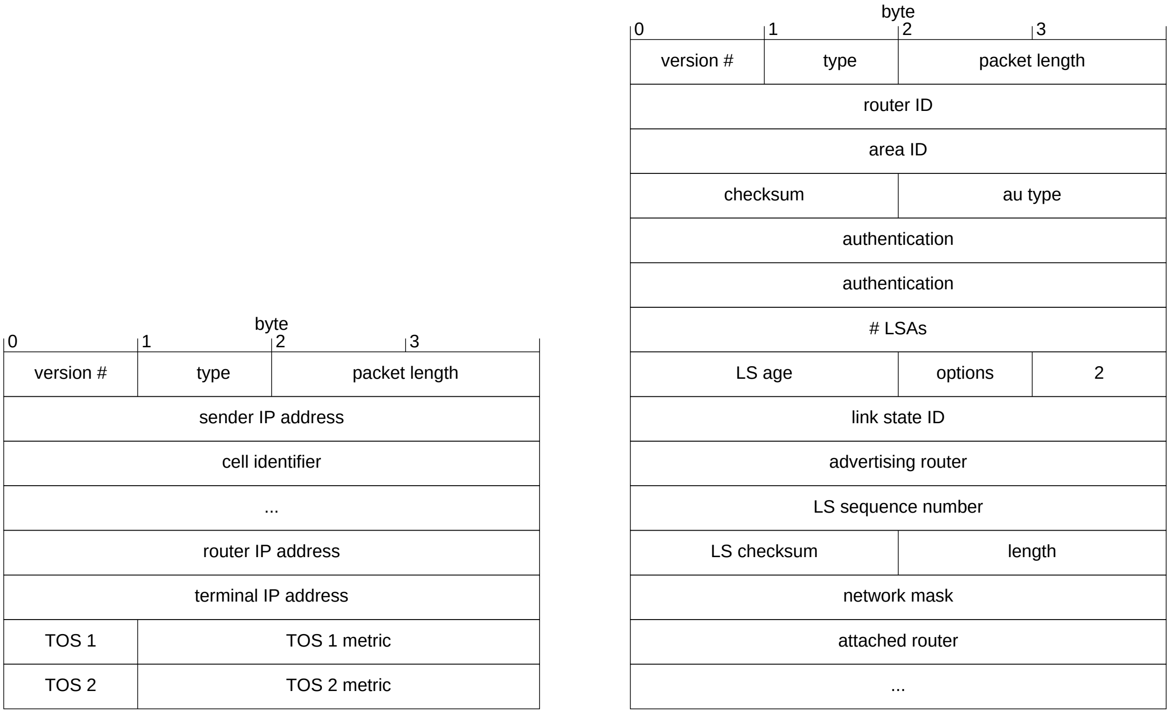
Figure C.3.: AODV packet format.



(a) WOSPF packet format: HELLO.

(b) OSPF packet format: HELLO.

Figure C.4.: WOSPF and OSPF HELLO packet format.



(a) WOSPF packet format: link state update.

(b) OSPF packet format: link state update.

Figure C.5.: WOSPF and OSPF link state update packet format.

D. Abbreviations

AGW Access Gateway

AODV Ad hoc On demand Distance Vector routing

AP access point

ARP Address Resolution Protocol

ARQ Automatic Repeat Request

ATM Asynchronous Transfer Mode

BGP Border Gateway Protocol

BS base station

CAPEX Capital Expenditure

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance

CTS clear-to-send

DCA Dynamic Channel Allocation

DCC Dynamic Cell Clustering

DCF Distributed Coordination Function

DMAC Distributed Mobility-Adaptive Clustering

DL-MAP downlink map

DSDV Destination-Sequenced Distance Vector Routing

DSL Digital Subscriber Line

DSR Dynamic Source Routing

ENB Enhanced NodeB

FCA Fixed Channel Allocation

FCH Frame Control Header

GGSN Gateway GPRS Support Node

GPRS General Packet Radio Service

GSM Global System for Mobile Communications

GTP GPRS Tunneling Protocol

HARQ Hybrid ARQ

HiperLAN High Performance Radio Local Area Network

HSDPA High Speed Downlink Packet Access

HSUPA High Speed Uplink Packet Access

HWMP Hybrid Wireless Mesh Protocol

ILP integer linear problem

IS-IS Intermediate System To Intermediate System Protocol

IP Internet Protocol

LOS line of sight

LP linear optimization problem

LSU Link State Updates

LTE Long-Term Evolution

MAC Medium Access Control

MILP mixed integer linear problem

MPLS Multi-Path Label Switching

MT mobile terminal

nLOS non line of sight

OFDMA Orthogonal Frequency Division Multiple Access

OLSR Optimized Link State Routing

OPEX Operational Expenditures

OSPF Open Shortest Path First

PCF Point Coordination Function

RNC Radio Network Controller

RERR route error

RNG-REQ range request

RNG-RSP range response

RREP route reply

RREQ route request

RS relay station

RSVP Resource Reservation Protocol

RSVP-TE Resource Reservation Protocol - Traffic Engineering

RTS ready-to-send

SC-FDMA Single Carrier Frequency Division Multiple Access

SGSN Serving GPRS Support Node

SINR signal to interference and noise ratio

TC topology control

TCL Tool Command Language

TDMA Time Division Multiple Access

TOS type of service

UDP User Datagram Protocol

UL-MAP uplink map

UMTS Universal Mobile Telecommunications System

VRP Vehicle Routing Problem

WIMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

WOSPF Wireless Open Shortest Path First

Bibliography

[3GP07] 3GPP. Technical specification 3rd Generation Partnership Project; technical specification group core network and terminals; General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (release 8). In *3GPP TS 29.060 V8.2.0*, December 2007.

[3GP08a] 3GPP. Long Term Evolution of the 3GPP radio technology. In *http://www.3GPP.org*, 2008.

[3GP08b] 3GPP. Universal Mobile Telecommunications System (UMTS); High Speed Downlink Packet Access (HSDPA); overall description; stage 2. In *ETSI TS 125 308 v7.5.0*, January 2008.

[3GP08c] 3GPP. Universal Mobile Telecommunications System (UMTS); High Speed Packet Access (HSPA) evolution; Frequency Division Duplex (FDD). In *ETSI TR 125 999 v7.0.1*, January 2008.

[ABG$^+$01] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. RFC 3209: RSVP-TE: Extensions to RSVP for LSP tunnels. In *Request for Comments*, Internet Engineering Task Force, December 2001.

[ACLY00] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. In *IEEE Transactions on Information Theory*, volume 46, issue 4, pages 1204–1216, 2000.

[ACM07] Ramón Agüero, Johnny Choque, and Luis Munoz. On the relay-based coverage extension for non-conventional multi-hop wireless networks. In *Proc. of IEEE International Conference on Communications 2007*, pages 5531–5536, Glasgow, Scotland, June 2007.

[ADF$^+$01] L. Andersson, P. Doolan, N. Feldman, A. Fredette, and B. Thomas. RFC 3036: LDP specification. In *Request for Comments*, Internet Engineering Task Force, January 2001.

[AMA$^+$99] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus. RFC 2702: Requirements for traffic engineering over MPLS. In *Request for Comments*, Internet Engineering Task Force, September 1999.

[Bah06] Michael Bahr. Proposed routing for IEEE 802.11s WLAN mesh networks. In *Proc. of Workshop on Wireless internet*, volume 220, issue 5, pages 1002 – 1007, Boston, USA, August 2006.

[Bah07] Michael Bahr. Update on the hybrid wireless mesh protocol of IEEE 802.11s. In *Proc. of IEEE Conference on Mobile Adhoc and Sensor Systems*, pages 1–6, Pisa, Italy, October 2007.

[Bar06] Ulrich Barth. 3GPP Long-Term Evolution / System Architecture Evolution Overview. In *Presentation, Alcatel*, September 2006.

[Bas99] Stefano Basagni. Distributed clustering for ad hoc networks. In *Proc. of Symposium on Parallel Architectures, Algorithms, and Networks*, pages 310–315, Fremantle, Australia, June 1999.

[Bet01] Christian Bettstetter. Mobility modeling in wireless networks: categorization, smooth movement, and border effects. In *ACM SIGMOBILE Mobile Computing and Communications Review*, volume 5, issue 3, pages 55–66, New York, USA, July 2001.

[Bet04a] Christian Bettstetter. The cluster density of a distributed clustering algorithm in ad hoc networks. In *Proc. of IEEE Conference on Communications*, Paris, France, June 2004.

[Bet04b] Christian Bettstetter. Mobility modeling, connectivity, and adaptive clustering in ad hoc networks. In *Herbert Utz Verlag*, September 2004.

[BF03] Christian Bettstetter and Bastian Friedrich. Time and message complexities of the generalized distributed mobility-adaptive clustering (GDMAC) algorithm in wireless multihop networks. In *Proc. of IEEE Vehicular Technology Conference*, pages 176–180, Jeju, Korea, April 2003.

[BHKV07] Lajos Bajzik, Peter Horvath, Laszlo Korossy, and Csaba Vulkan. Impact of intra-LTE handover with forwarding on the user connections. In *Proc. of IST Mobile and Wireless Communications Summit*, pages 1–5, Budapest, Hungary, July 2007.

[BL05] B.A. Bakaimis and T. Lestable. Connectivity investigation of mobile relays for next generation wireless systems. In *Proc. of IEEE Vehicular Technology Conference*, volume 4, pages 2192 – 2195, Stockholm, Sweden, May 2005.

[BL07] Xingkai Bao and Jing Li. Efficient message relaying for wireless user cooperation: Decode-Amplify-Forward (DAF) and Hybrid DAF and coded-cooperation. In *IEEE Transactions On Wireless Communications*, volume 6, issue 11, pages 3975–3984, November 2007.

[BMM07] Leila Boukhalfa, Pascale Minet, and Serge Midonnet. QoS support in a MANET based on OLSR and CBQ. In *Proc. of International Conference on Networking*, Sainte-Luce, Martinique, April 2007.

[Bro04] Gabriel Brown. What price a 3G base station? *Unstrung Insider*, April 2004.

[Bun99] Scott C. Bundy. Antenna downtilt effects on CDMA cell-site capacity. In *Proc. of IEEE Radio and Wireless Conference*, volume 2, pages 99–102, Denver, USA, August 1999.

[BZB+97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. RFC 2205: Resource ReSerVation Protocols (RSVP) – version 1 functional specification. In *Request for Comments*, Internet Engineering Task Force, September 1997.

[CEG79] Thomas M. Cover and Abbas A. El Gamal. Capacity theorems for the relay channel. In *IEEE Transactions on Information Theory*, volume 25, issue 5, pages 572–584, September 1979.

[Cis06] Cisco Systems Inc. Border gateway protocol. In *Internetworking Technologies Handbook*, 2006.

[CJ03] T. Clausen and P. Jacquet. RFC 3626: Optimized link state routing protocol (OLSR). In *Request for Comments*, Internet Engineering Task Force, October 2003.

[Com99a] IEEE LAN/MAN Standards Committee. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. In *IEEE Std 802.11-1999*, 1999.

[Com99b] IEEE LAN/MAN Standards Committee. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 2.5 GHz band. In *IEEE Std 802.11b-1999*, September 1999.

[Com99c] IEEE LAN/MAN Standards Committee. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz band. In *IEEE Std 802.11a-1999*, November 1999.

[Com03] IEEE LAN/MAN Standards Committee. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: further higher data rate extension in the 2.4 GHz band. In *IEEE Std 802.11g-1999*, 2003.

[Com04] IEEE LAN/MAN Standards Committee. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 7: Medium Access Control (MAC) Quality of Service (QoS) enhancements. In *IEEE Std 802.11i*, September 2004.

[CYCWB01] Wan Choi, Bong Youl Cho, and Tae Won Ban. Automatic On-Off switching repeater for DS/CDMA reverse link capacity improvement. In *IEEE Communications Letters*, volume 5, issue 4, pages 138–141, April 2001.

[DFZ05] Randall Dougherty, Christopher Freiling, and Kenneth Zeger. Insufficiency of linear coding in network information flow. In *IEEE Transactions on Information Theory*, volume 51, issue 8, pages 2745–2759, August 2005.

[Die97] Reinhard Diestel. Graph theory. In *Second Edition, Springer*, volume 4, March 1997.

[Dij59] Edsger W. Dijkstra. A note on two problems in connexion with graphs. In *Numerische Mathematik 1*, pages 269–271, 1959.

[DMS08] Mischa Dohler, Sidi-Mohammed Meddour, Djamal-Eddine amd Senouci, and Ahmed Saadani. Cooperation in 4G - hype or ripe? In *IEEE Technology and Society Magazine*, volume 27, issue 1, pages 13–17, 2008.

[EEH$^+$07] Jörg Eberspächer, Stephan Eichler, Christian Hartmann, Silke Meister, Robert Nagel, Robert Vilzmann, and Hans-Martin Zimmermann. **Wireless Multi-hop Networks: Classification, Paradigms and Constraints**. In *LKN-TR-4*, Institute of Communication Networks, Technische Universität München, October 2007.

[EFK$^+$06] Hannes Ekström, Anders Furskär, Jonas Karlsson, Michael Meyer, Stefan Parkvall, Johan Torsner, and Mattias Wahlqvist. Technical solutions for the 3G Long-Term Evolution. In *IEEE Communications Magazine*, pages 38–45, March 2006.

[EFS56] P. Elias, A. Feinstein, and C. E. Shannon. Note on maximum flow through a network. In *IRE Transactions on Information Theory*, volume 2, pages 117–119, 1956.

[EHB08] Jan Ellenbeck, Christian Hartmann, and Lars Berlemann. Decentralized inter-cell interference coordination by autonomous spectral reuse decisions. In *Proc. of European Wireless*, Prague, Czech Republic, June 2008.

[EKM07] Michelle Effros, Ralf Koetter, and Muriel Médard. Breaking network logjams. In *Scientific American, Inc.*, pages 78–85, May 2007.

[Emm05a] Marc Emmelmann. Influence of velocity on the handover delay associated with a radio-signal-measurement-based handover decision. In *Proc of. IEEE Vehicular Technology Conference*, volume 4, pages 2282 – 2286, Los Angeles, USA, September 2005.

[Emm05b] Marc Emmelmann. Influence of velocity on the handover delay associated with a radio-signal-measurement-based handover decision. In *Technical Report TKN-05-003*, Berlin, Germany, April 2005.

[EMSW02] C. Eklund, R.B. Marks, K.L. Stanwood, and S. Wang. IEEE standard 802.16: a technical overview of the WirelessMAN air interface for broadband wireless access. In *IEEE Communications Magazine*, volume 40, issue 6, pages 98 – 107, June 2002.

[ESW07] Michael Einhaus, Marc Schinnenburg, and Bernhard Walke. Distributed and adjacent subchannels in cellular OFDMA systems. In *Proc. of International Symposium on Wireless Personal Multimedia Communications*, page 5, Jaipur, India, December 2007.

[ETS96] ETSI. HIPERLAN type 1, functional specifications;. In *ETS 300-652*, June 1996.

[ETS98] ETSI. Universal Mobile Telecommunications System (UMTS); selection procedures for the choice of radio transmission technologies of the UMTS. In *ETSI TR 101 112 v3.2.0*, April 1998.

[ETS00] ETSI. HIPERLAN type 2; functional specification; Data Link Control (DLC) layer; part 1: basic data transport functions. In *ETSI TS 101 761-1*, April 2000.

[ETS02] ETSI. HIPERLAN type 2; Data Link Control (DLC) layer; part 2: Radio Link Control (RLC) sublayer. In *ETSI TS 101 761-2*, January 2002.

[EVB01] Jörg Eberspächer, Hans-Jörg Vögel, and Christian Bettstetter. *GSM - Switching, Services, and Protocols*. John Wiley & Sons, 2nd Edition, March 2001.

[Fal00] K. Fall. ns notes and documentation. *The VINT Project*, February 2000.

[FAMZ03] Frank Fitzek, Diego Angelini, Gianluca Mazzini, and Michele Zorzi. Design and performance of an enhanced IEEE 802.11 MAC protocol for multihop coverage extension. In *IEEE Wireless Communications*, pages 30–39, Dublin, Ireland, December 2003.

[Fea07] Seth Fearey. Joint venture: Silicon valley network announces test cities for wireless silicon valley initiative. In *Joint Venture: Silicon Valley Network, http://www.jointventure.org*, February 2007.

[FF62] Lester R. Ford and Delbert R. Fulkerson. Flows in networks. In *Princeton University Press*, 1962.

[FL02] H. Fattah and C. Leung. An overview of scheduling algorithms in wireless multimedia networks. In *IEEE Wireless Communications*, volume 9, issue 5, pages 76–83, October 2002.

[FNO99] Anders Furuskär, Jonas Näslund, and Håkan Olofsson. Edge-enhanced data rates for GSM and TDMA/136 evolution. In *Ericsson Review*, volume 1, Munich, January 1999.

[For56] Lester R. Ford. Network flow theory. In *P-923*, The RAND Corporation, Santa Monica, USA, August 1956.

[Fri46] Harald T. Friis. A note on a simple transmission formula. In *IRE*, volume 34, pages 254–256, 1946.

[FRT02] Bernard Fortz, Jennifer Rexford, and Mikkel Thorup. Traffic engineering with traditional IP routing protocols. In *IEEE Communication Magazine*, October 2002.

[FT00] Bernard Fortz and Mikkel Thorup. Internet traffic engineering by optimizing OSPF weights. In *Proc of. IEEE Infocom*, Tel-Aviv, Israel, March 2000.

[GBL03] Ingo Gruber, Georg Bandouch, and Hui Li. Ad hoc routing for cellular coverage extension. In *Proc. of IEEE Vehicular Technology Conference 2003*, volume 3, pages 1816–1820, Jeju, South Korea, April 2003.

[GL04a] Ingo Gruber and Hui Li. Cellular-ad hoc network interoperation for coverage extension. In *Proc. of IEEE 6th CAS Symp. on Emerging Technologies: Mobile and Wireless Communication*, pages 513–516, Shanghai, China, May 2004.

[GL04b] Ingo Gruber and Hui Li. A novel ad hoc routing algorithm for cellular coverage extension. In *Proc. of 11th International Telecommunications Network Strategy and Planning Symposium NETWORKS 2004*, pages 163–168, Vienna, Austria, June 2004.

[GLDS07] Larissa Globa, Andriy Luntovskyy, Alexander Dyadenko, and Alexander Schill. Performance-cost-ratio estimation by WIMAX/802.16 networks deployment. In *Proc. of Crimean Conference on Microwave & Telecommunication Technology*, pages 325–327, Sevastopol, Ukraine, September 2007.

[Gru07] Claus Gruber. C++ library graph. In *http://www.layonics.com*, 2007.

[Har02] Christian Hartmann. On capacity and blocking of multi service CDMA systems with smart antennas. In *Proc. of IEEE Vehicular Technology Conference*, volume 1, pages 72– 76, Vancouver, Canada, September 2002.

[HCC05] Jingyi He, Jiancong Chen, and S.-H. Chan. Extending WLAN coverage using infrastructureless access points. In *Proc. of IEEE Workshop on High Performance Switching and Routing*, pages 162–166, Hong Kong, China, May 2005.

[HHKV01] P.-H. Hsiao, A. Hwang, H.T. Kung, and D. Vlah. Load-balancing routing for wireless access networks. In *Proc. of IEEE Infocom*, pages 986–995, Anchorage, USA, February 2001.

[HKV08] Christian Hartmann, Moritz Kiese, and Robert Vilzmann. Optimizing the node degree in wireless multihop networks with single-lobe beamforming. In *Proc. of INFORMS Telecommunications Conference*, Maryland, USA, March 2008.

[HMZ$^+$07] Guido R. Hiertz, Sebastian Max, Rui Zhao, Dee Denteneer, and Lars Berlemann. Principles of IEEE 802.11s. In *Proc. of International Conference on Computer Communications and Networks*, pages 1002 – 1007, Honolulu, Hawaii, August 2007.

[Hör07] Katharina Hörnstein. Die neue Masche: Funknetze dehnen sich aus. In *CISCOnnect online, CISCO Deutschland*, March 2007.

[HYGC07] Jianhua He, Kun Yang, Ken Guild, and Hsiao-Hwa Chen. Application of IEEE 802.16 mesh networks as the backhaul of multihop cellular networks. In *IEEE Communications Magazine*, volume 45, issue 9, pages 82–90, September 2007.

[IEE04] IEEE. IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems. In *IEEE 802.16-2004*, 2004.

[IEE05a] IEEE. Approved Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. In *IEEE P802.16e/D12*, 2005.

[IEE05b] IEEE. IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11 Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements. In *IEEE 802.11e-2005*, 2005.

[IEE06a] IEEE. Draft Amendment to Standard for Information Technology - Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment ESS Mesh Networking. In *IEEE P802.11s/D0.01*, March 2006.

[IEE06b] IEEE. IEEE Standard for Information technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 1: MAC Sublayer. In *IEEE 802.15.3*, 2006.

[ILO07] ILOG Inc. ILOG CPLEX. In *http://www.ilog.com*, 2007.

[ISO02] ISO. Intermediate system to intermediate system routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473). In *ISO/IEC 10589:2002, Second Edition*, 2002.

[JCH84] R. Jain, D. Chiu, and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. In *DEC Research Report TR-301*, 1984.

[JCL$^+$01] P. Jacquet, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot. Optimized link state routing protocol for ad hoc networks. In *Proc. of IEEE INMIC*, Lahore, Pakistan, December 2001.

[JLMV01] P. Jacquet, A. Laouiti, P. Minet, and L. Viennot. Performance analysis of olsr multipoint relay flooding in two ad hoc wireless network models. Technical Report 4260, INRIA, September 2001.

[JMB01] David B. Johnson, David A. Maltz, and Josh Broch. DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In *Ad Hoc Networking, Addison-Wesley*, volume 5, pages 139–172, December 2001.

[JMS03] K. Jain, M. Mahdian, and M. Salavatipour. Packing steiner trees. In *Proc. of ACMSIAM Symposium on Discrete Algorithms*, 2003.

[KGG05] Gerhard Kramer, Michael Gastpar, and Piyush Gupta. Cooperative strategies and capacity theorems for relay networks. In *IEEE Transactions on Information Theory*, volume 51, issue 9, pages 3037– 3063, September 2005.

[KGK07] Sachin Katti, Shyamnath Gollakota, and Dina Katabi. Embracing wireless interference: Analog network coding. In *Proc. of ACM SIGCOMM 2007*, pages 257–269, Kyoto, Japan, August 2007.

[KJSMT01] Jamshid Khun-Jush, Peter Schramm, Göran Malmgren, and Johan Torsner. HiperLAN2: broadband wireless communications at 5 GHz. In *IEEE Communications Magazine*, volume 40, issue 6, pages 130–136, 2001.

[KK80] Parviz Kermani and Leonard Kleinrock. Issues and trends in router design. In *IEEE Transactions on Computers*, volume 29, issue 12, pages 1052–1060, December 1980.

[KKY03] D. Katz, K. Kompella, and D. Yeung. RFC 3630: Traffic engineering (TE) extensions to OSPF version 2. In *Request for Comments*, Internet Engineering Task Force, September 2003.

[KM03] Ralf Kötter and Muriel Médard. An algebraic approach to network coding. In *IEEE/ACM Transactions on Networking*, volume 11, issue 5, pages 782 – 795, October 2003.

[KMG$^+$07] Sachin Katti, Ivana Maric, Andrea Goldsmith, Dina Katabi, and Muriel Médard. Joint relaying and network coding in wireless networks. In *Proc. of IEEE International Symposium on Information Theory*, pages 1101–1105, Nice, France, June 2007.

[KN96] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey. In *IEEE Personal Communications*, volume 3, issue 3, pages 10–31, June 1996.

[KS98] S. Keshav and Rosen Sharma. Issues and trends in router design. In *IEEE Communications Magazine*, volume 36, issue 5, pages 144–151, May 1998.

[KT51] Harold W. Kuhn and Albert W. Tucker. Nonlinear programming. In *Berkeley Symposium, University of California Press*, pages 481–492, Berkeley, USA, 1951.

[Law01] J. Lawrence. Designing multiprotocol label switching networks. In *IEEE Communications Magazine*, pages 135–141, Berlin, Germany, July 2001.

[LB92] Jean-Yves Le Boudec. The Asynchronous Transfer Mode: a tutorial. In *Computer Networks and ISDN Systems*, volume 24, issue 4, pages 279–309, May 1992.

[LD60] A. H. Land and A. G. Doig. An automatic method of solving discrete programming problems. In *Econometrica*, volume 28, pages 497–520, 1960.

[LHWS96] Peter Kok Keong Loh, Wen Jing Hsu, Cai Wentong, and Nadarajah Sriskanthan. How network topology affects dynamic load balancing. In *IEEE Parallel & Distributed Technology: Systems & Technology*, volume 4, issue 3, pages 25–35, September 1996.

[LJS06] Peter Larsson, Niklas Johansson, and Kai-Erik Sunell. Coded bidirectional relaying. In *Proc. of IEEE Vehicular Technology Conference*, pages 851–855, Melbourne, Australia, May 2006.

[LNBL05] P. Lahdekorpi, J. Niemelii, J. Borkowski, and J. Lempiainen. WCDMA network performance in variable repeater hotspot traffic cases. In *Proc. of IEE International Conference on 3G and Beyond*, London, UK, November 2005.

[Lot05] Matthias Lott. ARQ for multi-hop networks. In *Proc. of IEEE Vehicular Technology Conference*, volume 3, pages 1708 – 1712, Dallas, USA, September 2005.

[LQLG07] Bo Li, Yang Qin, Chor Ping Low, and Choon Lim Gwee. A survey on mobile WiMAX. In *IEEE Communications Magazine*, volume 45, issue 12, pages 70–75, December 2007.

[LSP08] G. Lampropoulos, A.K. Salkintzis, and N. Passas. Media-independent handover for seamless service provision in heterogeneous networks. In *IEEE Communications Magazine*, volume 46, issue 1, pages 64 – 71, January 2008.

[LYC03] Shuo-Yen Robert Li, Raymond W. Yeung, and Ning Cai. Linear network coding. In *IEEE Transactions on Information Theory*, volume 49, issue 2, pages 371–381, 2003.

[Moy98] J. Moy. RFC 2328: OSPF version 2. In *Request for Comments*, Internet Engineering Task Force, April 1998.

[NH85] Lionel M. Ni and Kai Hwang. Optimal load balancing in a multiple processory system with many job classes. In *IEEE Transactions On Software Engineering*, volume 11, pages 491–496, May 1985.

[Pab04] Ralf Pabst et al. Relay-based deployment concepts for wireless and mobile broadband radio. In *IEEE Communications Magazine*, pages 80–89, New York, US, September 2004.

[PB94] C. E. Perkins and P. Bhagwat. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. In *Proc. of ACM SIGCOMM*, pages 234–244, London, UK, August 1994.

[PBRD03] C. Perkins, E. Belding-Royer, and S. Das. RFC 3561: Ad hoc on-demand distance vector (AODV) routing. In *Request for Comments*, Internet Engineering Task Force, July 2003.

[PH06] Petar Popovski and Yomo Hiroyuki. Bi-directional amplification of throughput in a wireless multi-hop network. In *Proc. of IEEE Vehicular Technology Conference*, volume 2, pages 588–593, Melbourne, Australia, May 2006.

[PLBV05] Santashil Palchaudhuri, Jean-Yves Le Boudec, and Milan Vojnovic. Perfect simulations for random trip mobility models. In *Proc. of Annual Symposium on Simulation*, pages 72 – 79, San Diego, USA, April 2005.

[PMB08] G. Paschos, P. Mannersalo, and T. Bohnert. Cell capacity for IEEE 802.16 coverage extension. In *Proc. of IEEE Consumer Communications and Networking Conference*, pages 933–937, Las Vegas, USA, January 2008.

[Poo07] I. Poole. Elements tutorial - what exactly is... LTE? In *Communications Engineer*, volume 5, issue 3, pages 46 – 47, June 2007.

[PR99] C. E. Perkins and E. M. Royer. Ad-hoc on demand distance vector routing. In *Proc of. IEEE Workshop on Mobile Computing Systems and Applications*, pages 90–100, New Orleans, USA, February 1999.

[PRDM00] C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina. Performance comparison of two on-demand routing protocols for ad hoc networks. In *Proc of. IEEE Infocom*, volume 1, pages 3–12, Tel-Aviv, Israel, March 2000.

[PRDM01] C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina. Performance comparison of two on-demand routing protocols for ad hoc networks. In *IEEE Personal Communications*, volume 8, issue 1, pages 16–28, February 2001.

[RE04] M. Rahman and P. Ernstrom. Repeaters for hotspot capacity in DS-CDMA networks. In *IEEE Transactions on Vehicular Technology*, volume 53, issue 3, pages 626–633, May 2004.

[Reu06] Hans-Jürgen Reumerman et al. **WIGWAM System Concept Version 3.0 Part 4: MAC Layer**. In *BMBF Project WIGWAM*, Dresden, March 2006.

[RL95] Y. Rekhter and T. Li. RFC 1771: A border gateway protocol 4 (BGP-4). In *Request for Comments*, Internet Engineering Task Force, March 1995.

[RMA03] Siddharth Ray, Muriel Médard, and Jinana Abounadi. Random coding in noise-free multiple access networks over finite fields. In *Proc. of IEEE Globecom*, volume 4, pages 1898–1902, San Francisco, USA, December 2003.

[RMSM01] E. Royer, P. M. Melliar-Smith, and L. Moser. An analysis of the optimum node density for ad hoc mobile networks. In *Proc. of IEEE International Conference on Communications*, volume 2, issue 3, pages 857–861, Helsinki, Finland, June 2001.

[RRE06] B. Roy, C. Roy, and M. Einhaus. A case study on frequency reuse in OFDMA systems using hierarchical radio resource management. In *Proc. of Asian-Pacific Microwave Conference 2006*, pages 851–856, Yokohama, Japan, December 2006.

[RSW06] Pathma Rathinavelu, Gottfried Schapeler, and Andreas Weber. UMTS coverage and capacity enhancement using repeaters and remote RF heads. In *Proc. of IEEE International Conference on Advanced Information Networking and Applications*, Vienna, Austria, April 2006.

[RTK06] N. Ratnakar, D. Traskov, and R. Kötter. Approaches to network coding for multiple unicasts. In *Proc. of International Zurich Seminar on Communications*, pages 70 – 73, Zurich, Switzerland, July 2006.

[RTR07] Andras Racz, Andras Temesvary, and Norbert Reider. Handover performance in 3GPP long term evolution (LTE) systems. In *Proc. of IST Mobile and Wireless Communications Summit*, Budapest, Hungary, July 2007.

[RVC01] E. Rosen, A. Viswanathan, and R. Callon. RFC 3031: Multiprotocol label switching architectures. In *Request for Comments*, Internet Engineering Task Force, January 2001.

[RW05] Boris Rankov and Armin Wittneben. Spectral efficient signaling for half-duplex relay channels. In *Proc. of Asilomar Conference on Signals, Systems, and Computers*, pages 1066–1071, Pacific Grove, USA, November 2005.

[SBR04] Kimaya Sanzgiri and Elizabeth M. Belding-Royer. Leveraging mobility to improve quality of service in mobile networks. In *Proc. of International Conference on Mobile and Ubiquitous Systems: Networking and Services*, pages 128 – 137, Boston, USA, August 2004.

[Sch06] Simon Schubert. Development of a demonstrator for cellular multi-hop networks. In *Diploma Thesis, Technische Universität München*, December 2006.

[Sch08] Bert Schuiling. LTE operator business case. In *Presentation, Nokia Siemens Networks GmbH*, Munich, January 2008.

[Sei05] Alexander Seitz. Design and implementation of a scenario generator for simulations with ns-2. In *Bachelor Thesis, Technische Universität München*, September 2005.

[SLKS03] Hyukmin Son, Sanghoon Lee, Soo-Chang Kim, and Yeon-Seung Shin Shin. Soft load balancing over heterogeneous wireless networks. In *IEEE Transactions on Vehicular Technology*, 2003.

[SVY06] I. Siomina, P. Varbrand, and D. Yuan. Automated optimization of service coverage and base station antenna configuration in UMTS networks. In *IEEE Wireless Communications*, volume 13, issue 6, pages 16–25, December 2006.

[SWPI03] Daniel C. Schlutz, Bernhard Walke, Ralf Pabst, and Tim Irnich. Fixed and planned relay based radio network deployment concepts. In *Proc. of 10th Wireless World Research Forum*, New York, USA, October 2003.

[TRL$^+$06] Danail Traskov, Niranjan Ratnakar, Desmond S. Lun, Ralf Koetter, and Muriel Médard. Network coding for multiple unicasts: An approach based on linear optimization. In *Proc. of International Zurich Seminar on Communications*, pages 1758–1762, Zurich, Switzerland, July 2006.

[TV02] Paolo Toth and Daniele Vigo. Models, relaxations and exact approaches for the capacitated vehicle routing problem. In *Discrete Applied Mathematics*, volume 123, pages 487–512, 2002.

[WCK05a] Yunnan Wu, Philip A. Chou, and Sun-Yuan Kung. Information exchange in wireless networks with network coding and physical-layer broadcast.

In *Proc. of Annual Conference On Information Sciences and Systems*, Baltimore, USA, March 2005.

[WCK05b] Yunnan Wu, Philip A. Chou, and Sun-Yuan Kung. Minimum-energy multicast in mobile ad hoc networks using network coding. In *IEEE Transactions on Communications*, volume 53, issue 11, pages 1906–1918, November 2005.

[WCZ$^+$05] Yunnan Wu, Philip A. Chou, Qian Zhang, Kamal Jain, Wenwu Zhu, and Sun-Yuan Kung. Network planning in wireless ad hoc networks: a cross-layer approach. In *IEEE Journal on Selected Areas in Communications*, volume 23, issue 1, pages 136–150, January 2005.

[Wib02] Björn Wiberg. Porting AODV-UU implementation to ns-2 and enabling trace-based simulation. In *Thesis*, Uppsala, Sweden, April 2002.

[WIM04] WIMAX Forum. Business case models for fixed broadband wireless access based on WiMAX technology and the 802.16 standard. In *http://www.wimaxforum.org*, October 2004.

[WMLC05] H. Wiemann, M. Meyer, R. Ludwig, and Pae O Chang. A novel multihop ARQ concept. In *Proc. of IEEE Vehicular Technology Conference*, volume 5, pages 3097 – 3101, Stockholm, Sweden, May 2005.

[WN99] Laurence A. Wolsey and George L. Nemhauser. Integer and combinatorial optimization. In *Wiley-Interscience Series in Discrete Mathematics and Optimization*, 1999.

[WQDT01] Hongyi Wu, Chunming Qiao, S. De, and O. Tonguz. Integrated cellular and ad hoc relaying systems: iCAR. In *IEEE Journal on Selected Areas in Communications*, volume 19, issue 10, pages 2105 – 2115, October 2001.

[WXCA07] Junfang Wang, Bin Xie, Kan Cai, and D.P. Agrawal. Efficient mesh router placement in wireless mesh networks. In *Proc. of IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems*, volume 4, pages 1–9, Pisa, Italy, October 2007.

[WYT$^+$05] R. Wakikawa, T. Yokota, K. Tasaka, H. Horiuchi, K. Uehara, and J. Murai. Experimentation of networked vehicle with multihomed mobile router. In *Proc. of IEEE Vehicular Technology Conference*, volume 1, pages 334 – 338, Dallas, USA, September 2005.

[Xia05] Yang Xiao. IEEE 802.11n: enhancements for higher throughput in wireless LANs. In *IEEE Wireless Communications*, volume 12, issue 6, pages 82 – 91, December 2005.

[YT04] E. Yanmaz and O.K. Tonguz. Dynamic load balancing and sharing performance of integrated wireless networks. In *IEEE Journal on Selected Areas in Communications*, volume 22, issue 5, pages 862–872, June 2004.

[YTM+02] E. Yanmaz, O.K. Tonguz, S. Mishra, Hongyi Wu, and Chunming Qiao. Efficient dynamic load balancing algorithms using iCAR systems: a generalized framework. In *Proc. of IEEE Vehicular Technology Conference*, volume 1, pages 586 – 590, Vancouver, Canada, September 2002.

[ZFH06] Hans-Martin Zimmermann, Florian Franz, and Rüdiger Halfmann. **Adaptive Routing In Wireless Multi-Hop Networks**. In *Proc of. Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, September 2006.

[ZG05] Hans-Martin Zimmermann and Ingo Gruber. **A Voronoi-based Mobility Model for Urban Environments**. In *Proc of. European Wireless*, Nicosia, Cyprus, April 2005.

[ZHW06] Hans-Martin Zimmermann, Rüdiger Halfmann, and Martin Weckerle. **Wireless OSPF in Cellular Multi-Hop Networks**. In *Proc of. Asian-Pacific Conference on Communications*, Busan, Korea, August 2006.

[Zim06] Hans-Martin Zimmermann. **UMTS - mobile Breitbandkommunikation für heute und morgen**. In *Technik in Bayern*, 2006.

[Zöl03] Stefan Zöls. Visualisierungstool für die Simulation von Ad Hoc Netzen. In *Diplomarbeit, Technische Universität München*, October 2003.

[ZL04] Hans-Martin Zimmermann and Hui Li. **Deployment Strategies for Transmitters in a Manhattan-like Environments**. In *Proc of. IEEE Vehicular Technology Conference*, volume 7, pages 5120 – 5124, Los Angeles, USA, September 2004.

[ZL08] Hans-Martin Zimmermann and Ying-Chang Liang. **Physical Layer Network Coding For Unicast Applications**. In *Proc of. IEEE Vehicular Technology Conference*, pages 2291–2295, Singapore, May 2008.

[ZLA05] Hans-Martin Zimmermann, Matthias Lott, and Angel Arnaudov. **Uplink Propagation Characteristics in 5 GHz Networks**. In *Proc of. IEEE Vehicular Technology Conference*, volume 2, pages 1342–1346, Dallas, USA, September 2005.

[ZLE05] Hans-Martin Zimmermann, Matthias Lott, and Jörg Eberspächer. **Frequency Reuse for Cellular Multi-Hop Networks**. In *Proc of. IEEE Vehicular Technology Conference*, volume 4, pages 2297 – 2301, Dallas, USA, September 2005.

[ZLL06] Shengli Zhang, Soung C. Liew, and Patrick P. Lam. Physical-layer network coding. In *Proc. of ACM MobiCom 2006*, pages 358–365, Los Angeles, USA, September 2006.

[ZSH06] Hans-Martin Zimmermann, Alexander Seitz, and Rüdiger Halfmann. **Dynamic Cell Clustering in Cellular Multi-Hop Networks**. In *Proc of. International Conference on Communications Systems*, Singapore, November 2006.