Technische Universität München
Lehrstuhl für Kommunikationsnetze

Path Lifetimes and Fair Medium Access in Wireless Multihop Networks

Ingo Gruber

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

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr.-Ing. Josef S. Kindersberger
Prüfer der Dissertation:	1.	Univ.-Prof. Dr.-Ing. Jörg Eberspächer
	2.	Univ.-Prof. Dr. rer. nat. habil. Carmelita Görg, Universität Bremen

Die Dissertation wurde am 25.05.2005 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 08.11.2005 angenommen.

Necessity is the Mother of Invention

Dutta-Roy, 2001
IEEE Communications Magazine

Abstract

This work deals with certain performance aspects of wireless ad hoc networks. Statistical lifetimes of single- and multipath routes are theoretically and simulative evaluated. Simulations based on realistic propagation models characterize the performance of ad hoc networks within urban environments. A novel extension based on the IEEE WLAN 802.11 protocol guarantees fair medium access among participating nodes. For that purpose, mobile nodes and the central access point exchange information about the optimal access rates. This strategy achieves almost complete fairness in flat as well as in urban environments.

Keywords: Ad hoc networking, wireless multihop, path lifetimes, geometric model, single path, multipath routing, diversity overhead, urban environments, propagation model, Walfisch-Ikegami, mobility model, performance measurement, evaluation, simulation, access point, fairness, medium access, queuing, protocol development.

Zusammenfassung

Diese Arbeit befasst sich mit verschiedenen Performanzaspekten von drahtlosen ad hoc Netzen. Statistische Lebensdauern von Einfach- und Mehrfachpfaden werden theoretisch und simulativ untersucht. Zur Charakterisierung der Leistungsfähigkeit von ad hoc Netzen in urbanen Umgebungen werden Simulationen auf Basis eines realistischen physikalischen Ausbreitungsmodells durchgeführt. Zur Gewährleistung eines fairen Medienzugriffs wird ein neues Verfahren entwickelt, das auf dem WLAN-Protokoll IEEE 802.11 beruht. Dazu tauschen die mobilen Knoten mit dem zentralen Access Point Informationen über die Zugriffsraten aus. In flachen ebenso wie in urbanen Umgebungen lässt sich damit nahezu vollständige Fairness erzielen.

Preface

Over the last couple of years, research interest in ad hoc networks considerably increased and it became one of the hot topics in communication networks. An ever increasing number of conferences and journal topics solely covering this issue confirms this trend. Therewith, it becomes difficult to identify yet unexplored ad hoc research areas. However, overcoming this challenge leads to new and unexpected insights. Therefore, I hope that this thesis will improve the understanding about theoretical and practical characteristics of ad hoc networks. It contributes in three main topics of ad hoc networks: lifetimes of arbitrary routes are of interest for the dimensioning of networks, the performance of networks in urban environments gives some insight into the network behavior under realistic conditions, and the provisioning of fairness within distributed networks will increase user satisfaction and contentment.

While working at the Lehrstuhl für Kommunikationsnetze I had the opportunity to team up with great people. First of all I want to thank my advisor Prof. Dr.-Ing. Jörg Eberspächer. He convinced me in working at the Technische Universität München and he continuously supported me. I am also very pleased that Prof. Dr. rer. nat. habil. Carmelita Görg is the second examiner of this thesis. Together with Stefan Aust of her team, we had an excellent cooperation. Both played an important role during our research project. I always appreciated the fruitful discussions and the cooperative teamwork.

Further on, I thank all my current and former colleagues. In particular, I will never forget the team work and beneficial discussions with one of my best friends and room mate Rüdiger Schollmeier. Besides, I would like to mention the continuous provisioned infrastructural support of Dr. Maier and Thomas ‘Willy’ Kurzhals. To collaborate with Stephan Eichler, Hans-Martin Zimmermann, Dr. Christian Hartmann, Peter Hinterseer, Stefan Zöls and Ingo Bauermann was a great pleasure. The discussions with Matthias Scheffel, Robert Prinz and Claus Gruber about fixed network communication issues were always very motivating. I also had the possibility to work with researchers outside the university. In particular, the cooperation with Dr. Wolfgang Kellerer from NTT Docomo and Dr. Hui Li from Siemens was excellent. Another thanks goes to my graduate students. Especially the work with Christian Mathiesen, Oliver Knauf, Andreas Bäßler, Florian Niethammer and Georg Bandouch considerably helped me in writing this thesis.

Last but not least, a very special thanks goes to my girlfriend Barbara, my beloved family and my friends. They always support me.

Munich, May, 2005 Ingo Gruber

Contents

1 Introduction **1**

2 Ad Hoc Networks: Principles and Challenges **6**

2.1 Principles and applications 6
2.2 Challenges 9
2.3 Enabling technologies 12
2.4 Routing protocols 14
2.5 Simulation environment 19
2.6 Open research issues 22

3 Path Lifetimes in Mobile Ad Hoc Networks **29**

3.1 Applied routing strategies and algorithms 30
3.2 Related work 35
3.3 Path lifetime analysis for two-hop connections 39
3.4 Path lifetime evaluation for arbitrary multi-hop connections 47
3.5 Simulative evaluation of multi-hop and multi-path scenarios 59
3.6 Routing overhead analysis 70
3.7 Summary 85

4 Evaluation of Ad Hoc Networks in Urban Environments **87**

4.1 Introduction 87
4.2 Existing ns-2 propagation models 88
4.3 Related work 92
4.4 Existing path loss prediction models 94
4.5 WIM model and urban mobility model 98
4.6 Simulations 104
4.7 Conclusion 111

5 MAC Layer Extension Providing Fair Throughput **112**

5.1 Introduction 113
5.2 Shortcomings of the conventional WLAN 802.11 standard 119
5.3 Related Work 125

5.4 Enhancements achieving fair throughput distributions 128
5.5 The fair-MAC protocol extension 133
5.6 Evaluation 137
5.7 Summary 147

6 Conclusion 149

A Abbreviations 151

B Terminology 153

C Necessary Simulation Duration and Repetitions 154

C.1 Simulation environment 154
C.2 Sufficient simulation times 155
C.3 Necessary numbers of simulations 157

1 Introduction

In 1982, IBM introduced the first personal computer (PC) with an 8 MHz processor and 512 Kbytes of RAM. It achieved just a few thousand floating point operations per second (Kilo-Flops). Even though, the PC could only fulfill some basic tasks, it was a milestone. It initially allowed everyone to own a computer. The performance of computers changed dramatically within the following 20 years. Common desktop computers have 3 GHz processor frequency and their computational power is rated in Giga-Flops, already approaching Tera-Flop regions. In 1965, Intel CEO Gordon Moore published his observation that the number of transistors and the computational power of common processors doubles every 12 to 18 months. This statement became famous as Moore's Law [1]. It holds since 30 years and researchers expect that it will continue at least through the end of this decade. Consequently, mobile phones or PDAs already contain more computational power than standard PCs ten years ago.

Another trend is the ubiquitous interconnection of PCs, notebooks and other electronic devices. This created and still creates new possibilities of usage and applications. Electronic mail, Internet, Voice-over-IP (VoIP), or remote maintenance are just a few applications which would be impossible otherwise. The miniaturization of devices together with these IP-based applications paved the way for the emergence of wireless interconnection possibilities. Wireless standards allow mobile computing without the need to be physically connected to a network. However, these standards for IP networks only support nomadic computing [2] scenarios, in which the position of participants is mostly static throughout a session, and handovers between different access points (AP) occur seldom. They particularly do not support the continuous mobility or the temporary connectivity of participants.

Therefore, research is focusing on a different approach. Protocols do not integrate single users into an existing infrastructure, but a group of participants spontaneously form a self-organized network. As the mobility of users is not limited in any way, the topology continuously changes. Such a network is created on an ad hoc basis, and only for the purpose to interconnect all participating nodes. Therefore, these networks are referred to as mobile ad hoc networks (MANET). Due to the possibility to create and organize a network without central management entities, ad hoc networking is characterized as the art of networking without a network [3].

MANETs have several characteristics, which make them unique in comparison to other existing networks. In the following, the most important ones are briefly highlighted:

- The main purpose of the network is the provisioning of a decentralized platform for the sake of information exchange between all participating nodes.
- All participants are free to move independently from each other, without the necessity to inform others about their movement. For this reason, the network does not guarantee the availability of certain services in any way.

- Participating devices are interconnected via wireless links. Links are either uni- or bidirectional. In case of bidirectional links, all participating nodes must support them.
- Devices can be of any size and may be portable. Due to their small form factors, their reduced weight, and their portability, they are referred to as nodes. As battery lifetime is a scarce resource for mobile devices, the radio transmission power is limited and hence the maximal radio range is limited as well.
- Networks contain several nodes. The size of the network is not limited, only scalability issues may prevent an infinite number of participants. The diameter of networks may be magnitudes larger than the maximal radio transmission range.
- Due to the limited radio range and the size of the network, paths between distant nodes consist of several node-to-node links of adjacent nodes. Participating nodes must act as routers or relays in order to forward data for other connections. This cooperating characteristic is the major attribute of MANETs.
- They are without infrastructure. Consequently they do not rely on any fixed network entities to set up network connections or to maintain provisioned services.
- Consequently, the network is self-organizing in order to allow the distributed maintenance and cooperation between all participating nodes.

Figure 1 depicts a schematic example topology for common ad hoc network setups. It illustrates most of the above stated network characteristics. The source node *S* initiates a connection to destination node *D*. As its radio transmission range is limited and node *D* is too distant, the connection consists of a multi-hop path, utilizing node *A* and *B* as relays. Other nodes are participating within the network, but they are not used as relays for this particular path. Additionally, nodes may be out of the proximity of any other node, thus, they become physically disconnected and isolated. In case node *B* moves out of the transmission range of its relaying neighbors, the path between *S* and *D* breaks without any previous indication. After the notification of the permanent path failure, the source node must establish a different path, e.g. utilizing node *C* instead, in order to continue to communicate with node *D*.

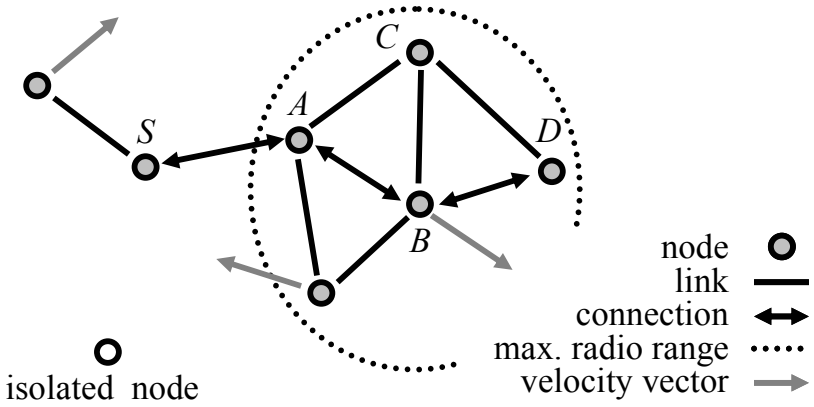

Figure 1: Schematic example topology of a mobile wireless ad hoc network.

Since the first appearance of wireless ad hoc networks as DARPA packet radio networks in the 1970s [4], they became an interesting research object in computer industry. During the last couple of years, tremendous improvements were made in the research of

MANETs. The advance was driven by standardization bodies like the Internet Engineering Task Force (IETF) and the Institute of Electrical and Electronics Engineers (IEEE), but also from an ever increasing interest of the industry in the unique characteristics of MANETs. Research puts emphasis on the behavior of ad hoc networks within different scenarios as well as the optimization of communication between distant nodes. In order to achieve this optimization, several network parameters, like node connectivity and maximal path lengths were identified. Scalability limits of MANETs, with respect to the number of nodes, the network load, as well as the dynamics of topology changes, are key issues. Additionally, the investigation of performance degrading effects is essential. The impact of short path lifetimes, high overheads, as well as unfairness among participating nodes limit the general usability of ad hoc networks. In order to obtain reasonable results, which match real world observations, sufficient simulation models are necessary. These research issues are the foundation to establish a plain understanding of ad hoc networks and to explain its inherent attributes, characteristics, and effects.

The thesis contributes to three performance degrading effects, which limit the usability of ad hoc networks. However, the path lifetime investigations are mostly independent from subsequent evaluations of the other two examined challenges:

- *Impact of different routing strategies on the path lifetimes*
  Several previous publications identified the path lifetime of routes as a key parameter to improve the overall network performance with respect to throughput, overhead and packet loss. Certainly, the length of path lifetimes is a statistical parameter, but besides the coincidence of the route choice, it clearly depends on certain network parameters. In order to allow statements concerning the ability of routing strategies to improve the average path lifetime, theoretical as well as simulative evaluations were carried out.

- *Impact of urban environments on the performance of ad hoc routing protocols*
  Existing research attempts mainly focus on flat environments, to evaluate the performance of ad hoc networks. However, the impact of different urban environments is not yet considered sufficiently. Therefore, a novel approach includes a physical wave propagation model to simulate ground plans with arbitrary building profiles. Results show new insights into ad hoc network behavior under various conditions. It clearly points out that the cooperation of nodes within city scenarios significantly differs from the interworking abilities within flat environments.

- *Link layer protocol extension to efficiently support ad hoc networks with APs*
  Common MANET approaches focus on the performance of truly distributed networks, where sources and destinations are uniformly spread within the network. However, the presence of gateways allowing the utilization of Internet based services, changes the distribution of the network load. The novel protocol extension is particularly able to achieve almost perfect fair throughputs among the existing network sources, and therefore allows a superior performance in comparison to existing approaches.

The remainder of this thesis is structured as follows:

Chapter 2 provides a description of the basic principles of ad hoc networks and the key features making it a unique approach. It includes the challenges of the novel network paradigm in comparison to existing network approaches. The chapter describes the impact of several ISO/OSI layers on the performance of MANETs and classifies some of the respective and most important methods into an overall system view. Additionally, it contains a short introduction into the network simulator ns-2 simulation suite and important simulation parameters. The chapter wraps up with several prevailing and discussed research topics.

Chapter 3 focuses on the impact of path lifetimes on the performance of ad hoc routing strategies. The literature suggests that the path lifetime is the key parameter for network performance, as it is directly related to the throughput and to the packet loss. The initial theoretical analysis examines a two hop ad hoc route. The subsequent analysis extends the scenario towards paths with an arbitrary number of mobile nodes and disjoint multipath setups. For both scenarios, a mathematical analysis of the resulting path lifetime distributions is presented. Simulations of three different multipath routing strategies verify the analytical results. The evaluation comprises various network setups and illustrates the superior behavior of non-disjoint ad hoc routing strategies. The following section analytically determines the diversity overhead for all three multipath routing strategies in comparison to the shortest-path routing. The chapter closes with a summary, concerning the improved path lifetimes in comparison to the additional overhead.

Chapter 4 investigates the performance of MANETs in urban environments. It contains a survey and validation of the currently existing ns-2 wave propagation models, as the ns-2 program suite is the utilized simulation tool. As none of the previous models is suitable as basis for simulation of urban scenarios, the following section presents numerous existing propagation models for non-flat environments. The Walfisch-Ikegami model proves itself as most appropriate and therefore it is described in detail. Although it shows some shortcomings as well, it achieves sufficient prediction accuracy with reasonable computational effort. The validation of existing mobility models shows their insufficiency to reproduce reasonable urban mobility traces. Therefore, a new mobility model for urban environments is introduced. Simulations were carried out, using the generic Manhattan grid, the original Munich city center, the common free-space, and a sparsely deployed scenario environment. Different building deployments lead to different ad hoc performance characteristics. They illustrate that the ground plan heavily affects the performance of MANETs.

The following chapter 5 examines the behavior of MANETs in the presence of APs. The non-uniform load distributions around APs force networks to behave significantly different to networks with uniformly distributed load. It initially emphasizes the shortcomings of the existing MAC protocols. The queuing algorithm and the per-node-fair medium access scheme are identified as unfavorable for MANETs containing central APs. The subsequent section discusses several known approaches and protocols. A validation pinpoints their inability to cope with frequently altering network topologies. In order to avoid both identified shortcomings, the following section presents circumventions which achieve fair throughputs, while they simultaneously do not degrade the overall network performance. The corresponding MAC protocol extension combines all necessary characteristics, while it does not increase the additional induced

packet overhead. Simulations in flat environments confirm the superior performance of the novel MAC extension in comparison to existing approaches. Subsequent simulations verify that the extension is even able to cope with urban ad hoc scenarios as presented in chapter 4. The conclusion summarizes the findings and gives an outlook of possible improvements.

Finally, chapter 6 summarizes the contributions and provides some possible future work areas for further improvements.

Appendix A contains a list of abbreviations, while appendix B introduces a terminology with necessary phrases. Appendix C briefly illustrates the impact of simulation time on the simulation results and describes its handling throughout this thesis.

Parts of the results and concepts covered in this thesis were previously published in [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17] and [18] and have been accepted for publication [19]. Additionally, some protocols and algorithms have been published as patents [20], [21] or been filed as patent applications [22], [23], [24]. Beyond these, several new and yet unpublished results are presented in this work.

2 Ad Hoc Networks: Principles and Challenges

2.1 Principles and applications

As mentioned in chapter 1, MANETs do not require central entities to setup operative communication platforms. Particularly, they do not rely on access points or base stations. The form factors of nodes may vary significantly, but they are usually at least portable like notebooks or even have packet compatible sizes like mobile phones and PDAs. All devices are wirelessly interconnected and generally move independently from other nodes. Due to the lack of any fundamental management control to organize the radio resource management, nodes are self-responsible for accessing the medium. To achieve this, all nodes must utilize a distributed medium access control (MAC), allowing to solve concurrent medium access from adjacent neighbors. Due to the limited battery power, the maximal radio transmission range of ad hoc nodes is magnitudes smaller compared to the coverage of cellular devices. Because of the size of networks in comparison to the short radio range, nodes are usually unable to reach communication partners within one hop. Therefore, sources set up connections over several intermediate nodes to exchange data with distant destinations. Nodes participating within a MANET commit themselves to relay packets for connections from other sources without expecting compensation.

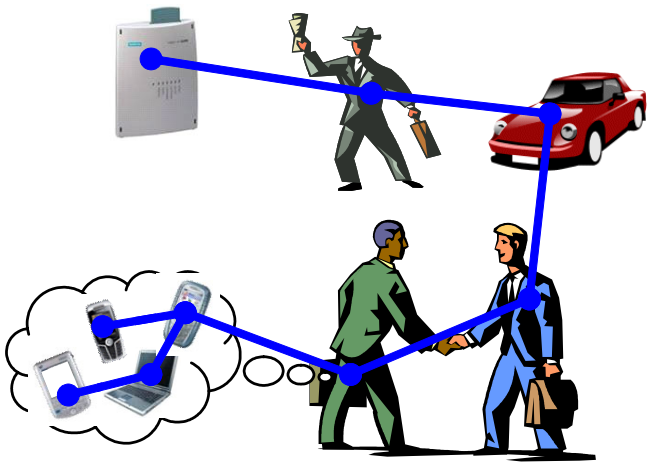

Figure 2: Possible application scenario for mobile ad hoc networks.

Obviously, the utilization of unstructured and distributed MANETs includes several advantages as well as disadvantages. The most important disadvantage is that participants do not obtain a service guarantee. The uncontrollable characteristics of MANETS prevent the provisioning of hard guarantees for certain network availabilities. The lack of service is as common as an uninterrupted service experience. Even if a network is available in general, the particular participation cannot be taken for granted.

The unpredictability of the network in general and its participating nodes in particular impedes predictions about future network conditions. The decentralized approach generates more overhead than centralized networks. Consequently the ratio between exchanged data and overhead is worse. Due to the necessity of multi-hop routes and wireless transmissions, the objectionable monitoring of information exchange is simple to achieve. Special security precautions are necessary to prevent eavesdropping and detect packet tempering or dropping during the relaying. Additionally, a trust basis between communication partners is hard to set up without an initially trustful entity.

One of the most important advantages is the possibility of instant information spreading among participating nodes. The absence of any central entity allows the spontaneous network setup. Participants only require their communication device, no additional expenses are necessary. In case the network utilizes an unlicensed frequency band like the 2.4 GHz industrial-scientific-medical (ISM) band, the creation of communication platforms at no charge is possible. The distributed approach of MANETs prevents single point of failures. From an end-to-end perspective, the network generally allows numerous resilient paths through the network. The failure of single nodes does not affect the operational basis of the complete network. Only numerous node failures, particularly if located in close proximity, impair the self-contained operation of the network. As presented, a secure ad hoc communication platform is difficult to achieve, but the distributed manner of ad hoc networks also has some advantages with respect to security issues. The eavesdropping of information with simple wiretapping of central infrastructure entities or the installation of security back doors is not feasible. The participation as part of the network is unconditionally necessary, with the likelihood of detection. Comparable to cellular networks, ad hoc networks allow the spatial reuse of frequency bands. But in contrast to their cellular counterparts, the distributed manner of MANETs allows a self-controlled distribution of bandwidth. However, the most important advantage is the possibility to spontaneously set up a network, without the need to configure it.

In order to complete the general consideration about ad hoc networks, possible topologies and application scenarios will follow. Two fundamentally different topologies exist to form MANETs. The first topology is defined as pure ad hoc network. It perfectly follows the ad hoc network paradigm of networking without a network. Besides differences in form factors or computational power, all nodes are equivalent. Statistically, sources and sinks of information exchanges are uniformly distributed among the network participants. Hence the network load is as well perfectly uniformly distributed. Whereas infrastructure based MANETs always cooperate with some kind of existing communication entity. Consequently, the underlying network topology is not uniformly distributed anymore. It varies significantly and depends on the considered scenario. The cooperation with existing network infrastructure requires different approaches and solutions in comparison to pure MANETs.

Although only two fundamental network structures exist, the possibilities of ad hoc networks allow several different application scenarios with, to some extent, strongly differing network parameters. Possible applications comprise pure networks, cellular extension, interworking with APs, military and disaster recovery applications as well as scenarios within vehicular environments. In the following, all of them are briefly described.

Within pure MANETs, all nodes simply participate for the sake of communication among each other. Possible scenarios occur in meetings or at conferences. The number of participating nodes may range from just a few to several hundreds while the mobility of nodes is very limited, and certainly does not exceed pedestrian speed.

Further on, MANETs can be used to extend the coverage of existing cellular network infrastructures [18]. Figure 3 illustrates an example. In existing cellular network setups, mobile devices terminate connections. In contrast to that, the mobile station within an extension scenario acts as relay for connections to nodes within the ad hoc domain. Nevertheless, the cellular base station remains the gateway to the wired domain. The number of participating nodes in the ad hoc domain is limited. The few mobile devices which are simultaneously located within the cellular and the ad hoc domain must relay all the traffic to and from the base station. Additionally, devices require at least two air interfaces to participate in both domains.

A similar scenario arises, when MANETs are situated around an AP. Chapter 5 discusses this particular application and challenges in detail. Not only the direct neighbors of the AP participate, but all nodes within the network are able to utilize the Internet gateway service of the AP. Obviously, the network topology is similar to the previously mentioned cellular extension. Therefore, the requirements are related in both cases as well. However, in contrast to the support of cellular extensions, nodes require only a single air interface as long as the AP is a common part of the network. Possible applications are at airports, train stations or in city centers.

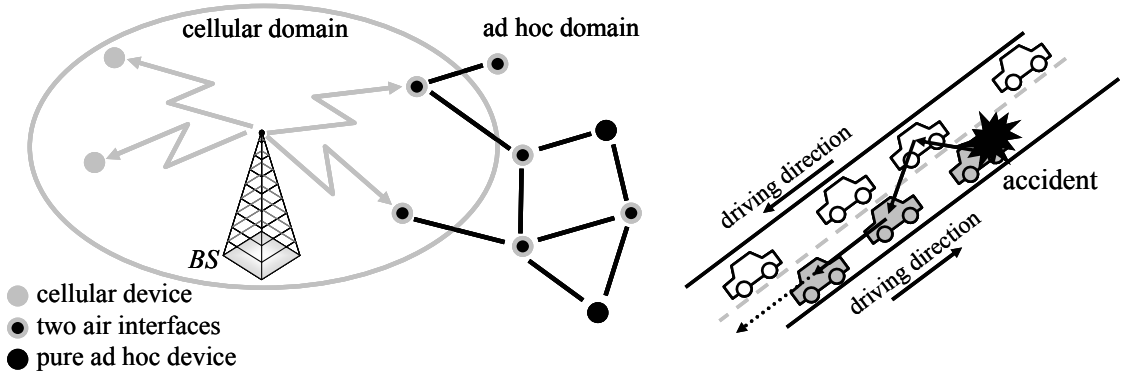

Figure 3: Cellular coverage extension and vehicular ad hoc scenario.

MANETs are also very valuable for military information exchange purposes. Their ability to setup communication during out-of-area engagements without the necessity to deploy fixed infrastructure is an important advantage over previous solutions. Possible applications may contain pure ad hoc scenarios but as well cooperation scenarios with fixed wireless infrastructure. However the potential of eavesdropping attempts by hostile forces requires the usage of high security communication protocols. The number of potential participants can range from just a few to even thousands, while all kinds of node velocities must be supported.

The advantageous usage of MANETs for civilian recovery operations is based on the assumption that any fixed infrastructure would be unavailable after the occurrence of disasters. However, the requirements are not as strict as for military operations. The need for a completely secure communication is unnecessary, and the number of participating

nodes is as well limited, but can also comprise hundreds of nodes. However, a simple usability must be guaranteed to ensure instant and uncomplicated information exchange.

Research also discusses the vehicular environment as possible scenario. Ad hoc network enabled cars could exchange information to prevent or bypass traffic-jams. On highways, spontaneously created MANETs can inform subsequent cars about an accident (see Figure 3) and force them to slow down. Additionally, it can guide the emergency ambulance to the place of accident. Within another possible vehicular scenario, MANETs assist drivers to detect possibly dangerous driving situations. Cars exchange information among each other about their current driving status, and therewith are jointly able to prevent accidents. Due to the possible high velocities of cars, networks have to be robust against link breaks and rapidly changing network conditions. The transmitted amount of data can be just a few bytes to inform neighboring nodes about accidents or it comprises continuous flow of several MBit/s to entirely exchange driving status information. However, the number of participating nodes is limited, because driver assistance information only has to be forwarded several hops around the originator.

As another application, peer-to-peer (P2P) overlay networks relying on MANETs are possible. They could greatly benefit from the distributed characteristic of ad hoc networks. The sharing of information among all participating nodes is one of the key characteristics of both networks [15]. The same idea leads to the conviction that location based services (LBS) as an application on top of MANETs outperform their counterparts based on cellular infrastructure. The limited spatial expansion of ad hoc networks easily allows the determination of the closest point of interest as well as the guidance towards it.

As a last application, wireless sensor networks are introduced. The MIT Technology Review [25] has rated them as one possible technology that will change the world. Although they are quite similar to MANETs, their main purpose is the gathering of information and not the provision of communication possibilities between distant nodes. Sensor nodes are usually specialized devices to detect changing environmental parameters, like temperature, pressure, or vibrations. Additional applications cover home automation, the industrial sector (undersea exploration, power plants), and medical monitoring (vital body functions). These new applications consequently lead to novel network setups. In order to allow area-wide monitoring, nodes utilize wireless multi-hop connections towards a central processing device to forward their gathered information. Depending on the particular scenario, sensor networks could be significantly larger than MANETs, while individual nodes are usually static. Furthermore, the novel network setup and power constraints prevent the usage of existing MANET communication protocols and necessitate the development of new approaches.

2.2 Challenges

After the introduction of the principles and applications of ad hoc networks, the following section focuses on the network inherent challenges. Obviously, the paradigm of unrestricted and uncontrolled information exchange lets emerge novel challenges yet. In the following, the most urgent challenges are listed:

- Wireless network protocols:
  The exchange of status information about the currently utilized links and paths further limits the available wireless bit rate. Ad hoc network protocols always have to find a reasonable trade off between the continuous exchange of accurate topology information and the maximization of the achievable data throughput. In small networks with rarely changing topologies, the performance optimization towards maximal throughput is favorable. While in dynamic networks the exchange of status information is essential to allow at least some data communication between distant nodes.

- Performance:
  The wireless multi-hop capability of ad hoc networks is novel. Path length and network conditions certainly affect the performance of ad hoc networks. Frequent topology changes due to node mobility cause numerous path breaks. Multipath routing algorithms can improve the network reliability by discovering additional backup routes. They also achieve a superior efficiency, because they minimize the necessity to perform overhead-prone route creations.

- Scalability:
  Ad hoc networks certainly face a scalability challenge [26]. Within cellular networks, only the link between mobile and base station is wireless. It is independent from the network size and the number of participating devices. Besides, ad hoc connections contain numerous wireless hops. Their average number increases proportional to the square of the participating nodes within the network [27]. In order to allow large networks, scalability is one of the most urgent challenges [28].

- Distributed networking:
  Additional challenges arise from the distributed manner of ad hoc networks. Fairness among participating ad hoc nodes is a major challenge. The equal allocation of available throughputs as well as a comparable distribution of network management tasks is difficult to achieve. The absence of a central entity does not allow the coordinated distribution of network information. The usage of a call admission control (CAC) to limit the maximal conveyed number of connections is difficult [29]. Furthermore, the collaborative limitation of the maximal acceptable individual node throughput circumventing network overloading is difficult to achieve.

- Mobility support:
  Another challenge is the support of high node mobility. As already raised, the support of vehicular environments [30] allows the creation of yet impossible applications. Obviously, the dynamics of topology changes increases with the average node velocity, if the average radio transmission range is kept constant. The average lifetime of links decreases and consequently that of routes as well. In order to allow reasonable networking within these highly dynamic networks, the selection of promising paths or static nodes is essential. The augmentation of routing protocols with additional information from the global positioning system (GPS) or from in-car sources like the speed indicator seems promising. However, these information sources are not at hand generally and therefore universal solutions are yet unavailable.

- Addressing:
  Routing protocols and higher layer applications require unique node IDs to address each participating node. While the assignment of fixed addresses is possible within closed user scenarios, it is unfavorable in scenarios with varying participants. Therewith, protocols must detect duplicate addresses and solve the ambiguity in a distributed manner [31]. The other challenge occurs, if two previously separated networks with different address domains merge. The protocol has to establish a common address space to ensure node accessibility from the Internet. Consequently it has to transfer the old addresses into the new space, while maintaining the uniqueness of individual node IDs. Obviously, this task requires notable overhead to perform, but in general is inevitable.

- Power constraints:
  Although power-saving is often rated as indispensable for sensor networks, this issue is as well crucial for mobile ad hoc networks. Optimal power efficiency requires multi-hop connections to reduce the transmission power. However, multi-hop forwarding generates as well an increasing amount of protocol overhead, which worsens the power efficiency [32]. However, there is always a trade-off necessary between multi-hop forwarding and low overheads. Initially, power-saving is a true multi-layer task, as all layers can contribute independently. All layers must keep their additionally induced protocol overhead as small as possible. The efficient transmission of information is in the responsibility of physical layer protocols. MAC protocols minimize the number of necessary transmissions, while routing algorithms have to discover short and durable paths.

- Security:
  The organization of cellular networks guarantees secure Authentication, Authorization, and Accounting (AAA). Each network forms a closed user group with a commonly used infrastructure, which ensures the trust basis. Therewith, the network provider inherits the task of identifying each participant. To ensure privacy of individual participants against third persons, solely the unique International Mobile Subscriber Identity (IMSI) [33] of mobile phones is used to initially authorize the user at the system. Afterwards the temporary IMSI (TIMSI) is utilized for any further communication. The eavesdropping of a TIMSI does not allow the identification of the original IMSI. However, the absence of a common provider makes the organization of a distributed trust basis in MANETs difficult [34]. Authentication of identities always requires a higher computation effort to proves credentials and certificates as well as an increased data overhead to transmit them. Additionally, to prevent the eavesdropping of information between distant participants, the wireless data communication must be entirely encrypted [35, 36].

- Inter-domain handover
  The currently used cellular networks utilize only a single wireless link between node and base station, the remaining route of a connection is entirely arranged within the fixed network domain. Consequently, terminal mobility only affects this single hop. The fixed network interconnection of adjacent base stations allows them to seamlessly handover moving mobile devices, before ongoing calls are affected or even interrupted. With only a single wireless hop between a node and an access point, the common topology of WLAN supported networks around access points is similar. However, they do not yet support seamless handovers between different access points [37].

2.3 Enabling technologies

The facilitation of the above described applications requires the utilization of particular underlying network technologies. These different network types are usually separated with respect to their intended purpose. The primarily considered wireless networks are the wireless personal area networks (WPAN), the wireless local area networks (WLAN), and the wireless metropolitan area networks (WMAN). As their notations suggest, they are utilized for short range, medium range as well as long range radio transmissions.

For most network types, an IEEE as well as an ETSI standardization board exist. The IEEE developments are pooled within the 802.×× working groups, while the Hiper××× standards comprise the ETSI efforts. These standards mainly describe the physical (PHY) as well as the data link control (DLC) layer. Their key technology characteristics are listed in Table 1.

Table 1: Key characteristics of distributed wireless networks.

	short range	medium range	long range
Standard	802.15	802.11, HiperLAN	802.16, HiperMAN
Status	deployed	widely deployed	standardization
Mobility	pedestrian	medium	fixed, pedestrian
Radio range	~10 m	~100 m	50 km
Network capacity	1…500 MBit/s	54 MBit/s	70 MBit/s
Spectrum	1…5 GHz	2.4 and 5 GHz	2…11 GHz, 10…66 GHz
Bandwidth	20..500 MHz	20..25 MHz	1.5…28 MHz

The long range IEEE 802.16 [38] and HiperMAN standard [39] were originally developed as a wireless point-to-multi-point (PMP) alternative to the digital subscriber line (DSL) [40]. IEEE 802.16 and HiperMAN are formed in close cooperation and therefore interoperate seamlessly. The promotion efforts are concentrated within the WiMAX (Worldwide Interoperability of Microwave Access) forum [41]. The 802.16 standard has a theoretical radio range of about 50 km and a shared network capacity of 70 MBit/s. First standard compliant products are announced for 2006. While it was originally specified for fixed broadband access, the working group 802.16e (WG-e) extends it to support mobile clients and seamless roaming between service areas. The latest group WG-f targets on meshed multi-hop networks. It supports the relaying of data and therewith elegantly expands the service coverage.

For WLAN networks, the IEEE develops the 802.11 [42] standards suite, while the ETSI proposes HiperLAN [43]. They utilize the unlicensed industrial, scientific, medical (ISM) bands at 2.4 and 5 GHz. However, widely deployed are only products supporting WLAN 802.11x. The prior 802.11b standard supports 11 MBit/s, while the latest extensions 802.11a/g and HiperLAN/2 offer up to 54 MBit/s. They have been originally developed as extension of fixed LAN networks. Their general usage is as “hotspot” protocol to provide wireless Internet access at airports, campuses, and train stations. Consequently they are optimized to support existing infrastructure. While the HiperLAN standardization favored a centralistic approach, the 802.11 protocols use a distributed medium access. Section 5.1 introduces the 802.11 protocol suite more in detail.

Although both standard suites support direct node-to-node communication without the presence of an access point, they do not achieve sufficient efficiency for reasonable multi-hop connections. Xu and Saadawi [44] illustrated some significant performance shortcomings. Therefore, the focus of the recently established WG-s [45] is to overcome these multi-hop inefficiencies at least for small networks with semi-mobile participants. In contrast, the attempt of the WG-n is to increase the bit rate of 802.11 even further. Although, the 802.11n standard is not yet accepted, proposed bit-rates are greater than 100 MBit/s. Table 2 shows the most important existing and emerging 802.×× high-speed standards and drafts.

To raise the bit-rates, the discussion focuses on the support of smart antennas. Multiple input, multiple output (MIMO) [46] devices use several antennas simultaneously during transmissions [47]. Obviously, an advanced signal processing on the receiver side must reconstruct the original data out of the linearly superposing transmissions. Siemens AG recently illustrated [48] the capability of MIMO by transmitting 1 GBit/s over a 100 MHz channel. The other possible application for smart antennas is spatial division multiple access (SDMA) [49] as basic physical access scheme. Transmitters and receivers create certain antenna beam forms and increase the antenna gain for the designated direction, while at the same time minimize the impact from other transmitters. Therewith, receivers separate the signals from multiple users which are separated in space.

Table 2: Existing and emerging 802.×× high-speed standards.

Standard	802.11a	802.11g	802.11n	802.15.3a
Access scheme	OFDM	OFDM	OFDM,MIMO	UWB
Status	products	deployed	draft	draft
Mobility	stationary	medium	medium	stationary
Radio range	~50 m	~100 m	~100 m	~10 m
Network capacity	54 MBit/s	54 MBit/s	> 100 MBit/s	500 MBit/s
Spectrum	5 GHz	2.4 GHz	2.4 GHz	3…11 GHz
Bandwidth	20 MHz	25 MHz	25 MHz	500 MHz

For short range networks, the most prominent protocol is Bluetooth [50]. As IEEE 802.15.1, it was the first standard within the 802.15 PAN standardization. It was originally developed as cable replacement technology. With relatively low bit rates of 1 MBit/s, it is able to wirelessly connect headsets, printers, computer mice, and keyboards. As some of the proposed applications have certain QoS requirements Bluetooth supports the isochronous data exchange. In contrast to previous protocols, Bluetooth uses a master-slave technique for node-to-node communication. While it supports up to eight nodes, the differentiation between masters and slaves leads to difficulties to setup multi-hop connections. Therefore, Bluetooth as basic technology is inappropriate to support truly distributed ad hoc networks. The ZigBee alliance [51] promotes the 802.15.4 protocol of the same name. It uses a star topology to wirelessly connect up to 255 nodes. Its main purpose is in-house automation, and therefore its devices are designed to consume minimal energy and support low bit rate (<250 kBit/s) services.

Two further standards within the WLAN 802.15 suite mainly focus on high-speed data communication. The 802.15.3 draft [52] utilizes the existing unlicensed ISM bands and

achieves data rates of 50 MBit/s. Additionally, the 802.15.3a proposal [53] targets on the usage of ultra wide band (UWB). The UWB technique spreads transmission impulses so that they require significant bandwidth. Standard drafts propose up to 500 MHz channels. As the combined transmitted energy remains constant, the spreading decreases the transmission energy per Hertz below the common noise level. Consequently, concurrent radio transmissions above the noise level are not affected and UWB seamlessly coexists with other radio transmissions techniques. In order to reconstruct the original information, UWB receivers use correlation methods. The 802.15.3a drafts consider data rates of up to 500 MBit/s and radio ranges of about 10 m. Furthermore, Intel Corp. announced UWB as favored technique for its wireless USB (universal serial bus) standard [54].

2.4 Routing protocols

While the MAC layer protocol controls the node-to-node packet transmissions, the establishment of favorable paths through the network is within the responsibility of ad hoc routing protocols. Numerous protocols have been proposed to set up wireless connections between distant ad hoc nodes. Most of them are particularly developed for individual scenarios and applications, and therefore have special advantages and disadvantages. The literature [55, 56] divides the routing protocols based on their general routing strategy. Most protocols utilize proactive or reactive routing strategies. However, there also exists hybrid and cluster-based algorithms using both methods. Figure 4 notes some of the most important protocols, but does not give a complete comprehensive survey. Independent from the respective routing strategy, protocols must cope with topology changes due to unpredictable terminal mobility. The limited maximal radio range of the air interface intensifies frequent link breaks whenever nodes leave each others proximity. The consequence is a continuously changing network topology.

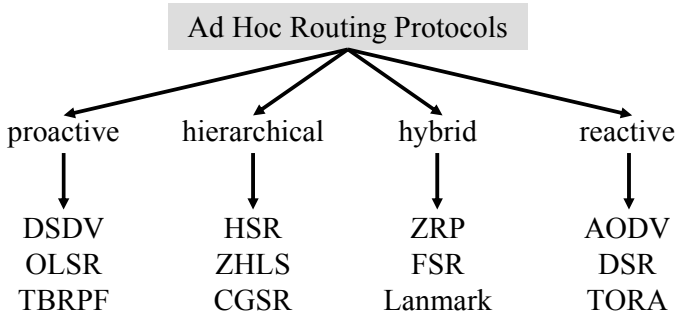

Figure 4: Classification of various ad hoc routing protocols.

2.4.1 Proactive routing protocols

Proactive routing protocols maintain a full system view. Nodes discovering a modification of a particular link status forward this information to every participating network node. Every node stores one or more tables to cache routing information. Therefore, nodes are able to set up connections immediately after the upper layer requests them. In case of high mobility and frequent link changes, proactive algorithms require high signaling overhead to preserve the network.

In the following, three proactive protocols are briefly described. The Destination Sequenced Distance Vector Routing protocol (DSDV) was the first proposed proactive routing scheme [57]. It is based on the classical Distributed Bellman Ford (DBF) algorithm [58-60], but guarantees loop-freedom. Broadcasted route updates contain the address of the destination, the number of necessary hops to reach the destination, the sequence number of the route information regarding the destination, and a new sequence number unique to the broadcast. Sequence numbers guarantee that nodes are able to distinguish between new and aged information. Receivers update their own routing tables and, if necessary, broadcast the tables. As every node maintains all necessary information of the entire network topology, route setups can be processed very fast with the locally stored information.

The 'Topology Broadcast based on Reverse Path Forwarding' (TBRPF) [61] is one of two proactive RFCs [62] standardized by the IETF. It is a link-state protocol and its main characteristic is the attempt to significantly reduce the number of necessary forwards of route update messages. In order to achieve this, each node maintains a minimum spanning tree of the network with itself as root. Figure 5 presents an example. In case of topology changes, nodes update their own tree information and spread the information among all participating nodes. Other nodes accept route update packets only in case the packet utilized the shortest path from the originator to itself, otherwise they discard the packet. Therewith, the protocol allows an early cancellation of unnecessary route maintenance packets and minimizes the induced overhead.

The other IETF standardized proactive routing protocol is the 'Optimized Link-State Routing' (OLSR) protocol [63, 64]. As the name implies, it is as well a link state protocol. Instead of broadcasting topology control messages to all neighbors, nodes forward the information only to those nodes, determined as essential to achieve complete information spreading. All other nodes are able to overhear the information exchange between these multipoint relay (MPR) stations. With the help of periodically broadcasted hello-messages, nodes keep track about the network topology within their two-hop neighborhood and about the necessary MPR stations. With the help of the topology control messages, all nodes are able to determine a reduced network frame of MPRs, necessary to completely connect the network. Consequently, the network only maintains the MPR framework and therewith OLSR reduces the routing overhead.

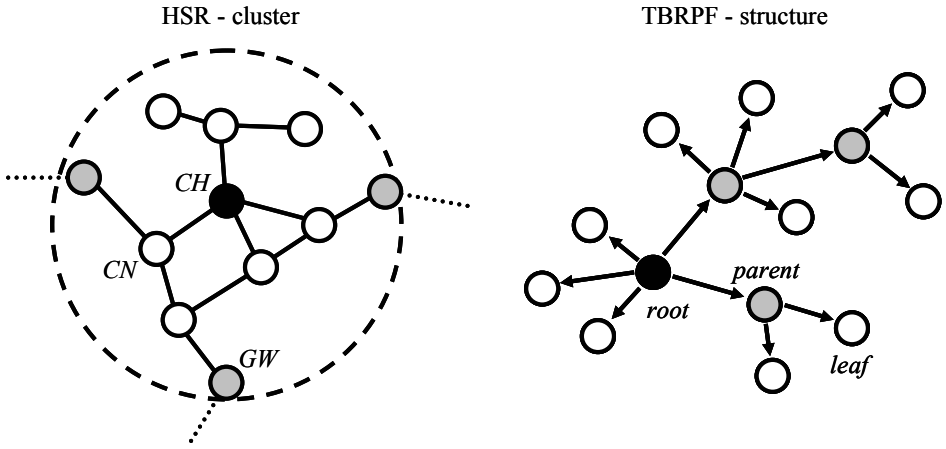

Figure 5: Hierarchical State Routing cluster and TBRPF structure.

2.4.2 Hierarchical routing protocols

The major advantage of hierarchical routing algorithms is the significant reduction of routing table storage and processing overhead. The Cluster-head Gateway Switch Routing (CGSR) [65] and the Zone-based Hierarchical Link State Routing Protocol (ZHLS) [66] use the DSDV routing algorithm described in the previous section as basis. In the following, the operational building blocks of hierarchical routing protocols are explained at the example of the Hierarchical State Routing (HSR) [67]. HSR is a physical clustering algorithm, as it relies on geographical and neighboring relationship. In hierarchical networks, nodes group themselves into clusters. Clusters contain a single cluster head *CH*, gateways *GW*, and multiple internal cluster nodes *CN*. Figure 5 presents an arbitrary network section of a cluster. Cluster heads manage the group formation and afterwards maintain their cells. HSR allows first-level cluster heads to form clusters of higher levels. Cluster head election is not specified within the algorithm, but can be based on the highest MAC layer address. Within clusters, each node monitors the link states to its neighbors. Cluster heads collect this information and forward them to neighboring cluster heads and to their higher level cluster head. As nodes move, clusters may split, merge and change cluster membership. While moving, cluster nodes send location updates to new and previous *CHs* to inform them about the altered cluster membership. Cluster heads announce changes to the highest level cluster from which the alternating membership is visible. HSR introduces the hierarchical ID (HID) to address nodes on different levels. Therefore, gateways have multiple HIDs as they are addressable by different clusters. This logical partitioning of the network allows the maintenance of large networks, because network maintenance messages do not have to propagate through the whole network, but only to certain cluster heads. However, mobility and location updates management is the main drawback, as it again causes additional overhead. Therefore, clustering is most suitable in low mobility networks while in highly dynamic environments the additionally induced overhead of HSR exceeds the overhead of flat link state algorithms.

2.4.3 Hybrid routing protocols

The Zone Routing Protocol (ZRP) [68] is often referred to as hybrid ad hoc routing protocol. It combines proactive and reactive elements. The ZRP maintains routing information for a local zone, and establishes routes on demand for destinations beyond this local neighborhood. It limits the scope of the local zone by defining a maximal number of hops for the local zone. Using ZRP with a maximal hop count of zero for its local neighborhood creates a reactive routing algorithm, and using it with hop count greater than the hop diameter of the network creates a pure proactive routing algorithm. The routing algorithm used in the local zone can be based on every table-driven routing algorithm. To determine routes beyond its local zone, the source forwards the request to its border nodes, which define the maximal extension of the proactively cached local zone. Upon reception, border nodes check if the destination is contained in their local zones. If so, the node sends a route reply on the reverse path back to the source. Otherwise they add their own addresses to the route request packet and again forward the packet to their border nodes. The main advantage of the ZRP is a reduced number of required route request (RREQ) messages and further on the possibility to establish new routes without the necessity to completely flood the network. The main disadvantage is the increased complexity of the routing algorithm and the determination of the optimal size of the local zone.

Other hybrid ad hoc routing algorithms initially invented to achieve scalability have as well shortcomings. The Broadcast Based Routing Protocol (BCBR) [13, 20] requires an additional broadcast channel. In contrast to that, the Fisheye State Routing Protocol (FSR) [69] only maintains an accurate routing table for nodes within close proximity, while it keeps only approximate information about more distant nodes. It is based on DSDV and therewith suffers from fast changing topologies as well. Inaccurate routing information for distant nodes also cause longer paths. The Global State Routing (GSR) protocol [70] is a subset of FSR. LANMAR [71] is based on FSR but particularly supports group mobility. Hence it only improves network performance when nodes move within groups.

2.4.4 Reactive routing protocols

In contrast to proactive routing, reactive protocols only maintain necessary topology information to keep connections alive. They do not maintain a complete overview over the network. Hence they are unable to set up new connections without delay, but save a great amount of signaling packets. For the creation of new routes towards destinations, they flood the network to search for an appropriate path. However routes are only available after completing the route request – route reply cycle. The most common reactive algorithms are the Ad hoc On-Demand Distance Vector Routing (AODV) protocol [72] and the Dynamic Source Routing (DSR) protocol [73]. AODV is already an RFC [74], while DSR still has draft status [75]. As third, often referenced reactive protocol, the Temporally-Ordered Routing Protocol (TORA) [76] exists. An IETF draft originally described the TORA approach to create a directed acyclic graph for routing purposes. However, it is not yet further developed and the draft expired.

Despite the initial route setup delay, reactive routing algorithms outperform their proactive counterparts under most conditions. Previous publications [73, 77] depict that AODV and DSR are favorable particularly for frequently changing network topologies. Additionally, they illustrate the inability of TORA to achieve a comparable network performance.

Das et al. presents in [78] that AODV outperforms DSR within challenging scenarios. For increasing network loads, network sizes and maximal node velocities, AODV is able to achieve higher throughputs. Whereas DSR is more suitable for less challenging conditions and for situations, in which low routing overhead is essential. Both routing protocols are considered within this thesis and therefore, the following section introduces them more in detail.

Ad hoc On-Demand Distance Vector Routing (AODV)

The AODV protocol utilizes pure shortest-path routing. It combines the use of DSDV destination sequence numbers with an on demand route discovery mechanism. Therefore, it does not maintain a full topology overview over the network. In case source *S* requires a path to destination *D*, it initially checks its cache, whether it contains a fresh route to *D*. Otherwise it creates a route request (RREQ) and broadcasts it to all its neighbors. Figure 6 depicts an example scenario, and illustrates the route establishment mechanism between source *S* and destination *D*. After the RREQ reception, nodes store the source node ID together with the ID of their upstream neighbor. Afterwards, they check if they are the searched destination, and if not, again forward the message to all neighbors. In case nodes receive copies of the same RREQ, they discard them. Figure 6 illustrates discarded messages as crossed arrows. The RREQ traverses through the

network until the RREQ reaches *D* or a node with information about a valid route to *D*. Upon reception, they create a route reply (RREP) message, and send it back to the nodes from which they initially received the RREQ. Nodes use the cached predecessor information and forward the RREP on the reverse path back to *S*. This time, every node saves the ID of the downstream neighbor from which it received the RREP. Therewith, relay nodes know the IDs of their downstream as well as their upstream neighbors. The path is established after *S* received the RREP. Nodes, not part of the established route, discard information about this particular RREQ after a timer expired. The distributed recording of path information allows the source to transmit data packets without knowing the absolute path towards *D*. This mechanism minimizes routing overhead, as data packets do not contain full path information. However, it requires extra memory in all intermediate nodes to store the routing table entries. In case of permanent errors of downstream links, the detecting nodes generate an RERR packet with *S* as destination. All upstream intermediate nodes delete the respective routes, and *S* creates a new RREQ to establish another path to *D*.

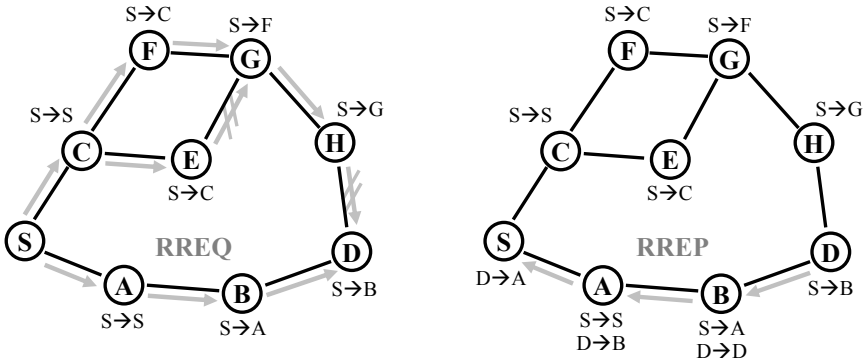

Figure 6: AODV routing scheme.

Dynamic Source Routing (DSR)

Comparable to AODV, DSR requires a route discovery and a route maintenance mechanism. As the name implies, it is based on the concept of source routing. A source *S* initiates the route discovery process, whenever it has to send data to a destination *D*. In case the routing cache of *S* does not contain a valid route to *D*, *S* broadcasts an RREQ to its neighbors. The RREQ only contains its own ID as initial route information. Figure 7 depicts an example network and illustrates all forwarded routing information. In case a receiving node is not the destination, it adds its own address to the packet and forwards it. Additionally, nodes discard all subsequently received copies of the same RREQ. For example, node *G* in Figure 7 discards the RREQ packets received from node *E*. In case an intermediate node knows a valid route to the destination, it generates an RREP message back to *S*. The RREP contains a combination of the source route received with the RREQ and the cached route to *D*. Otherwise *D* receives the RREQ and gets knowledge about the full path back to *S*. Therewith, it creates an RREP packet back to *S*. In any case, the RREP contains the full path information from *S* to *D*. After the reception of the RREP packet, *S* also has the entire necessary route information towards *D*. In contrast to AODV, DSR includes the complete route information in every generated data packet. Consequently, intermediate nodes do not cache path information and therewith minimize memory consumption. Route error packets (RERR) fulfill the task of route maintenance. Upstream nodes create them, whenever the downstream link shows

permanent failures. Every node relaying RERR packets, deletes the broken hop and all routes utilizing this particular link. The source node deletes the stale route as well and initiates a new route discovery.

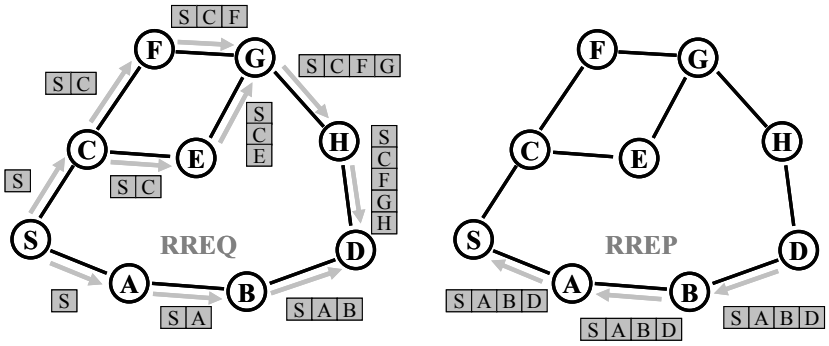

Figure 7: DSR routing scheme.

2.5 Simulation environment

Numerous different tools for ad hoc network simulations exist. However, some are unsuitable, because they are not supported any longer, like GlomoSim [79, 80], while others are not publicly available as open source, like QualNet [81] and OpNet [82]. The only tool which combines both characteristics is the network simulator ns-2 [83]. Therefore, the community widely supports and accepts it as research and as teaching tool. The original ns simulator was initially developed at the University of California at Berkeley in 1989 and is based on the REAL network simulator. In 1995, the DARPA supported the improvements through the VINT project which resulted in the development of the currently existing ns-2 program suite.

The ns-2 is an object oriented event driven, packet based network simulator. It is implemented in two different programming languages, namely C++ and OTCL. Due to performance considerations, the internal and static part of the program is based on the C++ class hierarchy. The internal C++ object hierarchy is in accordance to the ISO/OSI reference model. The dynamic and variable programming interfaces are implemented in OTCL. Therefore, OTCL scripts determine the general behavior of certain program modules as well as individually created simulation setups. The scripts allow a simple modification of simulation parameters at simulation start and during the simulation. Besides numerous other parameters, the script defines the network size, the utilized routing protocol, as well as the individual mobility behavior of all nodes. Example scripts for general IP-based simulation setups are presented in [84], while the examples in [85] focus more on mobile ad hoc networks. The network simulator sequentially processes the input file and generates output files with details about all simulation events. The output files are utilizable to analyze the performance and to visualize the ad hoc network with the network animator NAM. Figure 8 depicts the individual simulation steps.

The current version of the ns-2 program suite already contains numerous possibilities to simulate networks. However, as it is available in source code under an open source license, the program and single modules are continuously improved and extended. It is commonly used as simulator for local area networks (LAN), wireless LANs (WLAN),

and most interestingly ad hoc networks. It reproduces the behavior of certain network services like FTP and HTTP or the artificial packet generation with constant (CBR) and variable bit rates (VBR). Various different versions of the transport layer protocols UDP and TCP are supported. Furthermore, it is able to emulate the real time transfer protocol (RTP) as well as QoS extensions like IntServ, DiffServ and the ReSerVation Protocol (RSVP). Certainly, it supports the IP protocol but also various extensions like Mobile IP (MIP) or hierarchical MIP (HMIP) [86].

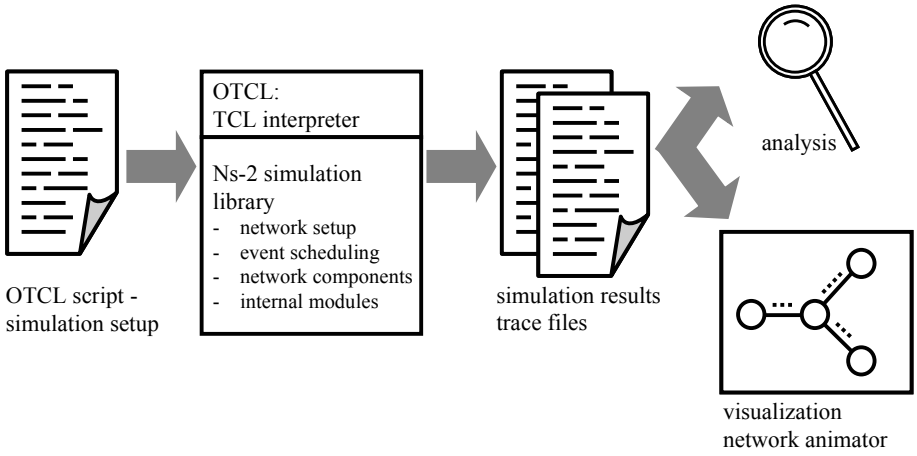

Figure 8: Operational process of ns-2 simulations.

Most important for ad hoc network simulations, four different routing protocols, namely AODV, TORA, DSR, and DSDV are already included. Numerous others are available, but must be manually added to the code source. As MAC layer protocol, ns-2 provides several implementations of standards for fixed network interconnection. For the simulation of ad hoc networks, ns-2 contains all functionalities of the IEEE WLAN 802.11b protocol and the Bluetooth stack is available as additional extension. Separated from the MAC layer protocol, different queuing schemes like Drop Tail, Random Early Detection (RED) und Class based Queuing (CBQ) are adjustable. For mobile wireless ad hoc simulations, the CMU monarch project [87] extended the ns-2 with wireless radio propagation models. Further details about the internal programming modules and a comprehensive description about all supported standards and protocols can be found in [88].

2.5.1 Mobility models

The particular movement of ad hoc nodes is not modeled during an ns-2 simulation run, but is previously determined with an external program and then linked to the OTCL setup script. Therefore, the utilization of arbitrary mobility models is possible. Most previous publications used the random waypoint mobility (RWP) model for ad hoc network simulations. It was initially invented to evaluate the DSR routing performance and [89] describes the model in detail. However, various publications [90, 91] point out that the RWP model causes unfavorable node distributions. Over time, the node density in the center of the simulation area becomes greater than at the borders. Further on, the rate of network topology changes due to node mobility decreases because of RWP's inherent characteristics. Both limitations prohibit a usage of RWP for reasonable evaluations. Figure 9 illustrates typical mobility traces for the RWP and the random

direction (RD) mobility model. It convincingly depicts the inability of RWP model to maintain a uniform node distribution.

Simulations and theoretical analysis in [92] illustrate that the RD mobility model [93] with bounce back at borders of the simulation area has the most valuable mobility characteristics. Due to this superiority of the RD model in comparison to the RWP model, it is used as basic mobility model for all evaluations within this thesis. In the following, it is described briefly. As with the RWP model, RD controlled nodes are initially uniformly positioned within the simulation area. Thereafter nodes independently choose a random direction $\varphi \in [0,2\pi]$, a random velocity $v \in [0,v_{max}]$, and a random movement time $t_m \in [0, t_{max}]$ with $t_{max} \ll T_{Sim}$ and T_{Sim} the simulation period. In case nodes approach borders, they bounce back into the simulation area and while keeping their movement angle. As a result, the average node density remains equivalent within all sectors of the simulation area while the dynamic of topology changes does not vary over time.

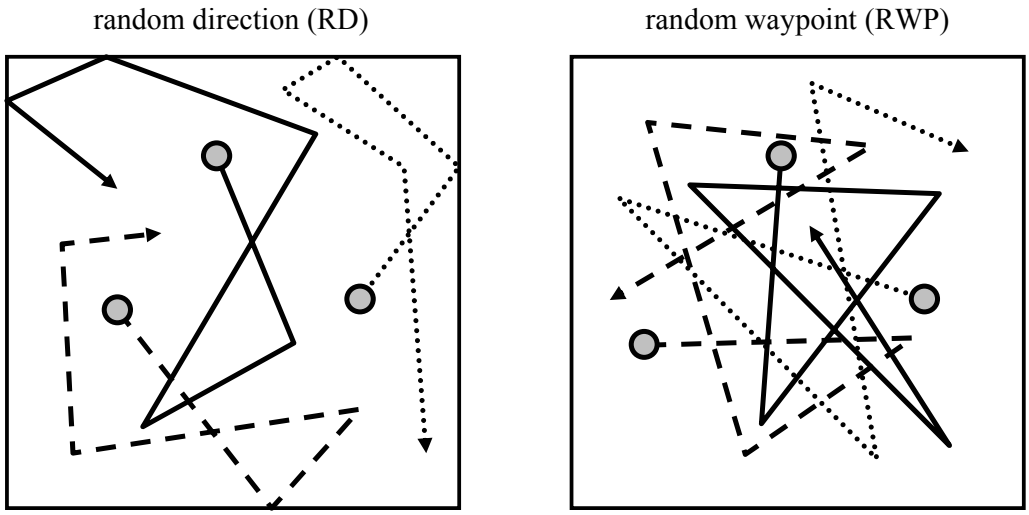

Figure 9: Typical mobility traces for the RD and RWP mobility model.

As already mentioned, several publications exist which comprehensively compared several ad hoc routing algorithms with the ns-2 simulation suite. They make qualitative statements about the performance of various protocols under certain simulation conditions and therewith illustrate the ability of the ns-2 to effectively compare the protocols. In the following the main contributions and the resulting conclusions are briefly presented.

2.5.2 Ad hoc network performance evaluations

Broch et al. [73] firstly use the ns-2 for performance measurements of ad hoc algorithms. They use the standard WLAN 802.11 as MAC layer and the two-ray ground (TRG) model for propagation predictions. They run simulations with 50 nodes. Their results reveal that all algorithms perform worst if they must cope with continuous node movements. DSDV and TORA as routing algorithms cause very high packet loss and routing overhead. Additionally, Broch's simulations reveal that DSR outperforms AODV in terms of packet loss and routing overhead.

Similar simulations were carried out by Johanson [77]. Initial simulations again utilized the TRG as propagation model. It reveals that AODV creates less packet loss in comparison to DSR within low load scenarios, and shows comparable packet loss results when utilized within high load simulations. In general, AODV requires more routing packets, while DSR causes more byte overhead. Subsequent evaluations used a very simple non-line-of-sight propagation model based on the TRG. In case an obstacle conceals the receiver of a transmission, the signal strength is set to zero, otherwise the TRG determines the propagation loss. Therewith, packet receptions are only possible if transmitter and receiver are within line-of-sight. The evaluations illustrate the ability of both algorithms to cope with frequent link breaks. Besides, their results do not reveal new insights. AODV and DSR perform almost equal.

The most recent evaluation was performed by Das et al. [78]. Their simulation setup slightly differs from previous evaluations. Results depict that AODV outperforms DSR within more challenging scenarios (higher network load, increased number of nodes, higher maximal node velocities). In less challenging conditions, DSR is more appropriate. Interestingly, the evaluations suggest that DSR is suitable for conditions, in which minimal routing overhead is crucial, e.g. when energy consumption must be limited. This is in contradiction to expectations, because DSR inherently creates more overhead by the usage of source routes.

2.6 Open research issues

Ad hoc networking is yet mainly a field of research, and only few universally utilizable solutions exist. While the enabling technologies already addressed some research issues, the above mentioned challenges are as well open topics. Therefore, many research issues remain unsolved, and the following section focuses on the most significant.

Figure 10 illustrates a categorization of the different topics with respect to the layer structure of the ISO/OSI reference model. Obviously, the challenges cover all ISO/OSI layers, and additionally introduce some multi-layer topics, like fairness and energy efficiency. Due to the unique characteristics of ad hoc networks, some yet unknown research issues emerged as well. Especially the constraint to develop solutions that work in a completely decentralized, variable, and self-organized manner is difficult to achieve. Due to the vast variety of applications, there cannot be a common and unified solution for all requirements. Varying scenarios require variable solutions.

Ad hoc networks and their participating nodes always suffer from scarce resources, like energy and frequency spectrum. Cross-layer approaches allow the optimization of network performance, although this violates the paradigm of minimal interworking between different ISO/OSI layers. The interworking with existing networks allows seamless services among different domains. Therefore, it presents another important research topic.

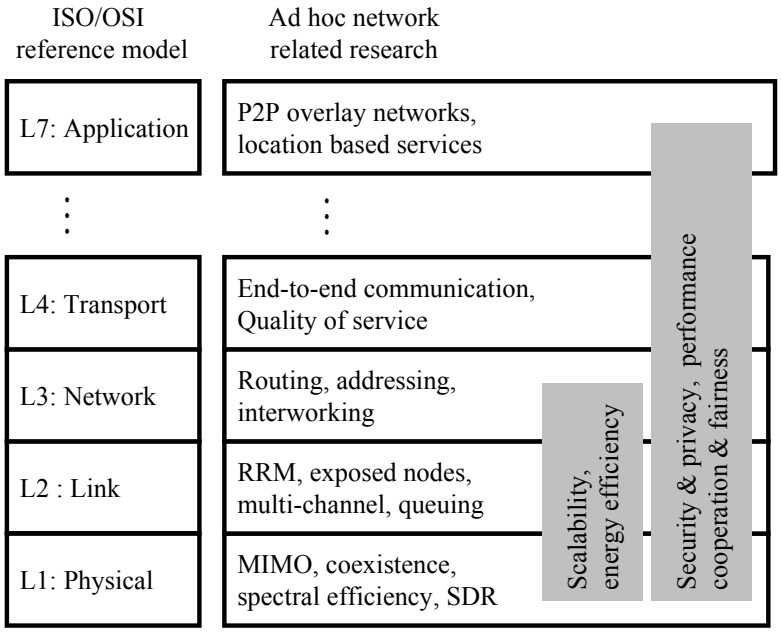

Figure 10: ISO/OSI reference model.

Physical Layer

The physical layer allows several improvements, requires certain design decisions and contains additional constraints. One of the major constraints is that the utilized frequency band is not arbitrarily selectable. This accompanies the design decision, whether an unlicensed (Industrial, Scientific, Medical: ISM) or a licensed band is preferable. The advantage of licensed bands is that national and international authorities grant exclusive access. For example, cellular providers solely use their GSM and UMTS bands. However, the charges for this service are significant. In contrast, unlicensed bands are free of charge. But in return, everyone has unrestricted access. Therefore, different protocols and standards may use the same frequency band. Standards have to consider the shared usage of the same frequency spectrum. Consequently, the coexistence of multiple standards in an uncoordinated manner is an important design issue.

The acceleration and optimization of the physical layer access is always a goal. The maximization of the spectral efficiency (in Bits per second per Hertz) by the use of new access methods and coding schemes allows the transmission of more bits within the same frequency spectrum. However, it always requires higher power consumption due to an increased signal processing complexity. The usage of MIMO together with orthogonal frequency division multiplex (OFDM) recently reached a spectral efficiency of about 10 Bit/s/Hz [48]. However, this does not seem to be the maximal, and therefore ongoing research might achieve even higher spectral efficiencies.

A yet open question is, if ad hoc networks require unique coding and modulation schemes. Or is a joint development for all kinds of wireless solutions optimal? Unquestionable is that different ad hoc scenarios and applications require different physical layer designs. Simple wireless single-hop connections to an AP have significantly different requirements than e.g. communication in vehicular environments. Physical layer necessities comprise high bit rates, large transmission ranges, robustness against bit errors and low power consumption. There will not be a general solution that

fits all circumstances. Optimizations are always application-dependent. However, the correct respective parameters are not yet evaluated and defined in all cases.

Modifiable coding and modulation schemes allow the reduction of the achievable bit rate in return for an improvement in robustness. However, continuously varying conditions make the optimal adjustment of the modulation schemes difficult. A possible solution is software defined radio (SDR) [94]. It is able to adapt to all network parameters, but causes additional computational complexity, and it is not yet able to rapidly change its parameters.

Medium Access Control layer

As the name implies, the function of the medium access control is the coordination of the medium access. Its main goal is an efficient and performance optimized handling. However, fairness among concurrent attempts from participating nodes is as well an issue. The uncoordinated network structure of ad hoc networks requires the usage of a distributed random access scheme. The 'carrier sense multiple access with collision avoidance' (CSMA/CA), used by the WLAN 802.11 suite is one option but others are proposed and discussed. To overcome the hidden terminal dilemma, the distributed coordination function (DCF) of WLAN 802.11 utilizes an additional in-band signaling scheme. While it allows a true distributed medium access, evaluations in [44] show severe performance degradations when used for multi-hop communication. Therefore, one of the most urgent challenges is the optimization and extension of existing MAC protocols to efficiently support multi-hop communication. Thereby the minimization or even prevention of unnecessary exposed nodes during data exchange takes a prominent role.

Out-of-band signaling offers the possibility to separate the channels for handshaking and data packet transmissions. This improves the efficiency, but also requires an additional radio transceiver. If multiple channels are utilizable, e.g. by the usage of OFDMA, a distributed radio resource management (RRM) and scheduling is necessary. Although some approaches [95, 96] are proposed, an optimal solution is not yet developed. Current MAC queues are usually organized in a FIFO manner. However, routing packets maintain the network management and therefore are important for the overall network performance. Therefore, the prioritization of these packets offers the possibility to achieve some performance improvements.

Network layer

The most commonly mentioned task of the network layer is routing. Due to the frequently changing topologies of ad hoc networks, existing routing approaches are unsuitable. Neither principles from the cellular network routing nor from the fixed network routing are applicable. The novel characteristics of ad hoc networks require completely novel approaches. For this reason, the IETF established the MANET working group [97] in 1997. Section 2.4 introduces the most important proposed and standardized ad hoc routing protocols. The three already standardized protocols only accomplish routing within pure ad hoc networks. They do not cooperate with existing fixed network infrastructure, nor do they profit from additional information sources. Therefore, reactive protocols always flood the network for a route discovery, while proactive protocols must continuously exchange topology information. This causes significant overhead and consequently reduces the available data rate. The usage of auxiliary information sources could prevent theses unintelligent network management

procedures and advantageous algorithms can perform similar tasks while saving bit rate. Particularly applications within vehicular environments greatly benefit from position and velocity information obtained from GPS and car sensors. Several position based routing protocols [98, 99] exist, but they are not yet standardized.

Sensor networks cover completely different applications. With respect to routing protocols, two different approaches are discussed. Either sensed data is continuously gathered by a central entity or nodes request to receive sensed information from other nodes. In this case, nodes do not address an explicit communication partner, but specify the sensor information category. Therewith, these sensor routing protocols need not to be address driven but content driven. While for the first sensor routing approach, an adapted, proactive routing protocol might be applicable, the second one requires novel approaches.

Besides routing, addressing is as well in the responsibility of the network layer. The challenge to uniquely address all participating ad hoc nodes is already considered in section 2.2. A global valid solution is not yet developed, and therefore it remains a field of research. Additional challenges arise when considering the necessity of interworking with existing fixed networks. In order to allow the access from the Internet, ad hoc nodes require IP addresses with the same subnet prefix as the gateway. Or the network address translation (NAT) protocol running on the gateways assigns local addresses to each node. In any case, the gateways must manage the address configuration. However, efficient protocols are not yet standardized nor even developed.

If participants require accessibility independent from their current location, the interworking with Mobile IP (MIP) [100], or its enhancement Hierarchical MIP (HMIP) is essential. While MIP and especially HMIP are optimized for nomadic mobility, they do not allow seamless handovers between different domains. Further improvements are necessary to address this issue [101, 102]. Another research topic occurs, when considering large scale ad hoc networks with several gateways. As described, gateways must manage the address configuration of ad hoc nodes in order to ensure their accessibility from the Internet. However, if multiple gateways are present, which gateway is responsible for which node? It gets even worse when considering node movement. Nodes could move into new domains or from one provider to another provider. Keeping the addressing consistent, always causes overhead and consequently must be done efficiently.

In order to allow interworking with fixed network infrastructure, a preliminary gateway discovery is essential. The routing protocol searches for explicit addresses for regular unicast communication. In contrast, the respective IP-address is dispensable for a gateway discovery, a certain node feature is important. Therefore, the search equals an anycast approach, and consequently existing ad hoc routing solutions are not applicable. One characteristic of anycast searches is that they may discover several independent gateways. The random selection of a gateway is simple but certainly not efficient. Therefore, the selection of the optimal gateway based on known network parameters is essential. This significantly reduces overhead and improves overall network performance.

The interworking with fixed network infrastructure does not only cause difficulties, it presents new possibilities as well. Large scale ad hoc networks containing several gateways also offer novel approaches for routing. Routes between two nodes do not have to remain completely in the ad hoc domain. Fractions of the path can be routed through

the fixed network infrastructure, minimizing the ad hoc network load and improving the path reliability. However, it is not always advantageous, and therefore exact characteristics are necessary to determine the optimal switching parameters and values.

The interconnection to cellular networks is also an open research topic. Ad hoc networks form a cost-efficient possibility to augment the services of existing cellular networks. The interconnection between both networks allows the extension of the coverage of cellular base stations. In addition, if the extending ad hoc networks overlap, a simple load balancing between cellular base stations is applicable. If base stations are overloaded and reject call attempts, the extended ad hoc networks may route the call to another base station within the proximity. As for the interworking with fixed networks, a gateway discovery algorithm is essential for the cellular interconnection. However, the cellular interworking also creates novel challenges. It must interconnect two completely separated and different networks. Cellular networks are designed to support highly mobile participants. Consequently, the extending ad hoc network must cope with them as well, although they represent unfavorable forwarding nodes. The ad hoc network only meets the QoS parameters of cellular networks, in case routing protocols circumvent these highly mobile nodes. However, the topic to discover the most durable paths is yet an open research issue, although some approaches exist [21, 103, 104]. Open topics are as well the handover between ad hoc networks controlled by different base stations or the possibility of fixed network originated call attempts. The consistent addressing of nodes presents again an inherent challenge.

Transport and application layer

Although most of the research issues are committed to the first three layers, some remaining issues are located in the higher layers. The most prominent open topic is the development of a TCP version, which efficiently supports ad hoc networks. Existing versions [105] efficiently support packet transport within fixed networks. They consider congestions within intermediate routers as main reasons for packet losses and delays. At least recent versions [106] achieve sufficient performance with single hops in the wireless domain. Besides the congestion in intermediate nodes, they must deal with wireless channel effects, like interference and fading, as additional reasons for packet losses. However, these solutions are not extendable to support true wireless multi-hop communication. For ad hoc networks, TCP must also consider network topology changes. Therefore, completely new approaches were developed [107, 108], but research is not yet completed.

Accompanying topics comprise the provisioning of QoS. Up to now, ad hoc network protocols only provide best-effort traffic. It is questionable, if protocols will be able to assure certain QoS parameters, comparable to the IntServ or DiffServ frameworks in fixed networks. Alternatively they could provide soft guarantees based on some novel statistical approaches.

Research does not focus on the development of services and applications solely for ad hoc networks, but existing ones have to work well within the novel environment. Evaluations [109] illustrate that the simple usage of existing fixed-network applications within ad hoc networks do not lead to reasonable results. It rarely scales and the induced overhead is significant. Existing applications are not aware of the new network environment. Therefore, they have to be tailored and adapted for seamless interworking. A possible solution is that network and application layer exchange status information to avoid duplicate network management tasks. Therefore, the ad hoc network application

research is always a cross layer issue. As described previously, application research comprises location based services and the cooperation with peer-to-peer networks [14, 24, 110].

Cross and multi-layer topics

As illustrated, most research issues mainly focus on a single issue in ad hoc networking, like routing or MAC layer protocols. Only recently, the research community started to discuss more integrated approaches [111], in which the packet transport, the routing, the link control, as well as the physical transmission cooperate. These cross layer methods violate the existing ISO/OSI reference model, as they support information exchange of non-adjacent layers. However, they achieve superior performance in comparison to existing approaches. This is necessary to widen the variety of possible application scenarios.

- Energy efficiency:
  Through the advances in miniaturization, nodes become tiny devices. However, the energy density of batteries does not keep up with the speed of miniaturization. Consequently, the overall available battery energy decreases and becomes the main limiting factor. Therefore, an energy efficient design is essential [112]. Particularly sensor networks suffer from the scarce energy resource [113], because participating devices are usually not rechargeable. Long lasting usage is only achieved with the help of energy saving algorithms and protocols. Although protocols on all layers can contribute, the true potential only arises when considering cross-layer approaches. Obviously, protocols achieve the maximal energy savings when they partly turn off nodes. However, there is always a trade off between overall network performance and individual node energy consumption.

- Performance improvements:
  Cross layer protocols spanning several OSI layers allow the exchange of status information and therewith allow the development of integrated solutions. These approaches are usually more efficient than their counterparts following the common OSI layer structure. However, supplementary modifications affect all participating layers and are thus more difficult to accomplish. Cross layer approaches avoid that protocols on different layers must perform similar tasks to maintain network integrity. Additionally, shortcuts between layers usually allow faster notification about network and topology changes. Therewith, time-consuming delay till timers expire is unnecessary and faster reactions are possible. The research focuses on the question, which layers have to exchange status information, and what kind of information is valuable. The physical layer can inform the network layer about the currently achievable data rates while the MAC layer notifies the transport and application layer about the link status. On the other hand, applications can inform the network and link layer about the type of service and expected data volume.

- Security and privacy:
  As described previously, the provisioning of security and privacy is one of the major challenges of ad hoc networks. True multi-layer research is necessary, as all layers are affected by the provisioning of security. Either protocols on certain layers demand it or they must cope with the additionally induced overhead. However, the open questions are complex: How can security be guaranteed without a central entity? How is it possible to establish a trusted network environment? Is the

authentication of certain participants possible? Another scaring scenario would be, if privacy within ad hoc networks is not guaranteed. Malicious participants could identify certain nodes and continuously track them. This would ultimately lead to profiles of movement traces and an entire loss of privacy. The prevention of denial of service (DOS) attacks is also a security issue. Protocols must prevent the interruption of network communication and therewith the selective disconnection of certain nodes. This comprises as well secure routes. Even the eavesdropping of route information must be impossible, e.g. via the exchange of protected network management information.

- Fairness:
  The ad hoc network paradigm requires that all or at least the majority of available nodes take over network tasks. Thereby the forwarding of packets for other connections is an essential function. If too many nodes refuse to participate, the whole network would collapse. Therefore, nodes generally have to participate, although it drains their batteries. Network protocols must ensure overall fairness among participants and exclude uncooperative nodes. These selfish nodes have to be permanently or temporarily banned from utilizing provisioned network services. A possibility is the introduction of a virtual currency [114]. Nodes purchase the cooperation of participating nodes and earn the currency for the provisioning of own assistance. However, this requires secure accounting and billing methods, and is difficult to achieve in a distributed ad hoc network. Obviously, it is directly related to the privacy and security topic described above. In provider managed networks, data forwarding can be directly beneficial through a refund or rebate policy of the provider. In any case, it forms an interface to business related topics. Additionally, protocols have to ensure that nodes share all available resources evenly. However, it also has to be guaranteed that none of the participants is overreached with respect to energy consumption or traffic load.

3 Path Lifetimes in Mobile Ad Hoc Networks

Mobile terminals spontaneously form self-organized MANETs. Routes between any source and destination consist of intermediate relays connected via wireless links. Generally, nodes move unrestrictedly and independent from each other. Within these continuously changing topologies, routing protocols create paths through the network. Stable routes last longer, and consequently the protocol must initiate less frequently route discovery cycles. For reactive routing protocols, the route discovery mechanism contains a network wide flooding. It generates notable overhead, causes long delays, and reduces the throughput of all other currently utilized connections. Increasing the time span between consecutive route discoveries would improve the overall network performance. Simulations with a reactive protocol as well as a theoretical analysis in [115] have verified that the basic behavior of ad hoc networks depends on the path lifetimes (PLT). It illustrates that the PLT is proportional to the throughput and inversely proportional to the induced overhead. Therefore, the overall network performance is directly related to the average PLT of routes. The average PLT of discovered routes is the main indicator for the quality of ad hoc routing protocols. Hence, the selection of stable routes with long PLTs is one objective of routing protocols. The challenge is the selection of promising routes, knowing only the current network topology.

This chapter gives a detailed insight into link and path lifetimes in mobile ad hoc networks. Section 3.1 introduces single- and multipath route creation strategies and gives protocol examples. Similar analytical approaches for predicting PLTs are discussed in section 3.2. Section 3.3 mathematically analyzes link and path lifetimes for a two hop route. The following section 3.4 extends the analysis to routes with an arbitrary number of hops and also discusses the benefit of disjoint multipath. It depicts the imponderability of predicting exact PLTs for arbitrary paths, and gives advice for reasonable routing protocol developments. An evaluation of different ad hoc routing strategies with respect to the average PLTs follows in section 3.5. An exact modeling and investigation of numerous routing strategies needs simulative evaluations instead of an analysis. Since multipath ad hoc routing strategies cause more routing overhead than single path algorithms, section 3.6 discusses this diversity overhead. The overhead considerations are based on existing ad hoc routing protocols. This includes an theoretical model to determine the induced overhead for arbitrary network topologies. Additionally it contains simulative evaluations to verify the theoretical analysis. Finally, section 3.7 indicates which routing strategy permits longest PLTs under most networking conditions.

3.1 Applied routing strategies and algorithms

Besides the theoretical analysis of PLTs, simulative considerations are necessary to determine the PLTs of certain multi-path algorithms. The particularly induced overhead for the routing protocols is important. For this reason, the following section introduces the various considered routing strategies examined by simulations. The section mainly focuses on multi-path algorithms, as the aspects of single path ad hoc routing algorithms are already described in section 2.4. However, the first subsection characterizes the underlying route creation strategies of the single- as well as the multi-path algorithms. The subsequent section briefly presents the protocol implementations of the examined routing algorithms. It presents the necessary information exchange between distant nodes and shows the particular design of the protocol messages.

3.1.1 Route creation strategies

This section introduces the before mentioned different route creation strategies. Besides the shortest path (SP) routing strategy, three multipath strategies are considered, namely the flooding based routing (FL), the disjoint multipath (DMP) as well as the non-disjoint multipath (NDM) routing. The section describes their general routing behavior and their key characteristics. Additionally, it briefly surveys the respectively created path sets.

- Shortest path:
  As first and most basic routing strategy, the shortest path (SP) routing is introduced. The source node chooses its communication partner and the routing algorithm discovers the SP between these two nodes. The SP is defined as the path between source and destination with the fewest number of hops. In case different paths require equal number of hops, the algorithm randomly chooses one of these. In particular, it does not consider any additional metrics to improve the path selection. The major characteristic of the SP algorithm is that it discovers only one single path.

  Ad hoc routing algorithms commonly use flooding to discover the shortest path. In protocol implementations, the SP is commonly defined as the path with the shortest response time, not with the fewest number of hops. The average delay to detect an unoccupied wireless medium is almost constant for individual hops. This delay together with the processing time within each node adds up to the major delay component of the response time. Therewith, the hop distance is again the most important parameter for the delay as it directly affects the response time.

- Flooding based multipath:
  As described above, reactive SP routing protocols utilize in general the flooding (FL) algorithm to create paths. Although the SP algorithm requires only a single path, the flooding algorithm often discovers multiple, non-disjoint, paths. The flooding algorithm extends the SP strategy with the yet unused non-disjoint backup routes. Characteristically, the flooding algorithm favors links on the shortest path. Therefore, backup routes often use the same first few links between source and destination as the SP. Only subsequent links are unique to the links on the SP. The trajectories of backup paths never cross the shortest path. Additionally, they never use the same links as the SP after they utilized unique links. Whenever links on the primary path and close to the source break, most backup routes break as well. Link

breaks close to the destination often do not affect the utilization of backup routes. Although, the found paths are suboptimal, the advantage of this strategy is its ability to keep the overhead almost as low as the SP algorithm. Further on, flooding based multipath route discovery algorithms generate less overhead than dedicated multipath algorithms.

- Disjoint multipath:
  In comparison to the flooding based routing extension, the disjoint-multipath (DMP) algorithm is a true multipath strategy. The main characteristic of a set of disjoint paths is that any two path of this set do not contain any link or node in common. This guarantees that paths break independently. In case the primary path of the set breaks, all subsequent paths are still valid. Therefore, the algorithm is able to switch to one of the remaining paths, whenever a permanent error occurs. AODV Multipath (AODVM) [116] is a working routing protocol which utilizes the DMP strategy and is described in detail in the following section 3.1.2.

  However, the limitation to link or even node disjoint paths reduces the number of maximal discoverable path M during a route request to the minimum number of neighbors of the source S_n and destination D_n ($M=min(S_n, D_n)$). In case the source has only a single direct neighbor, a link disjoint multipaths algorithm discovers only the shortest path.

- Non-disjoint multipath:
  The non-disjoint multipath (NDM) strategy uses a similar algorithm as the DMP, but allows the utilization of links several times within the same set of paths. The utilization of non-disjoint paths adds an additional degree of freedom to the route discovery algorithm. The algorithm is not restricted to discover disjoint paths, but favors them. Therewith, it is independent from the minimal number of neighbors of source and destination. It always discovers the proposed number of paths. However, it requires a certain limit of requested paths. Otherwise the algorithm tries to discover an infinite number of paths, unable to detect that the particular path was already discovered during a previous reply. Nevertheless, NDM is more flexible because it is able to reuse certain important links to setup additional backup routes. A variation of the AODVM algorithm discovers these NDM paths.

3.1.2 Routing algorithms

The evaluation in section 3.5 focuses on reactive routing algorithms rather than proactive ones. Therefore, the following section describes all simulatively evaluated protocols and gives details about their main path searching mechanisms. As described, reactive or on-demand routing algorithms do not keep track of the overall network topology. The algorithm generates only route discovery requests to distant nodes upon demand from an upper layer protocol or an application. As nodes usually need only few simultaneous connections to other nodes, this limits the overhead and saves available bit rate. Particularly in dynamic networks, with frequently changing topologies, the algorithms do not have to maintain an overall network view. They simply exchange topology information to keep their own connections working. Hence, they are unable to set up new connections without delay, but require preceding route requests before they are able to transmit data packets. For the creation of new routes towards destinations, they flood the network to search for an appropriate path. However, the route is only available after completing the route request – route reply cycle. Several performance

studies [56, 73, 77, 117] of MANETs depict that on-demand routing algorithms cause fewer routing overhead. Hence they are able to maintain more nodes with an equivalent information exchange effort and therefore networks scale better.

The examined five different routing protocols cover different aspects of reactive algorithms. Four algorithms are based on a table driven approach and one is a representative of the source routing paradigm. Further on, one algorithm strictly utilizes single path while the other four use different multipath search patterns.

AODV is the only examined single route protocol and additionally forms the basis for the flooding extension AODV-FL. It allows the discovery of non-disjoint backup routes during route request (RREQ) floodings. The third algorithm is the dynamic source routing protocol (DSR) with the same ability to utilize flooded backup routes (DSR-FL). AODV-Multipath (AODVM) allows the detection of disjoint backup routes, and, as the name implies, is based on AODV as well. In contrast to AODVM, the non-disjoint multipath AODV (AODV-DMP) searches for non-disjoint backup routes.

All evaluated algorithms have in common that they utilize network wide flooding to discover paths to the destination. In order to ensure loop free routes, every node maintains a monotonically increasing sequence number for itself. In combination with node IDs, this allows temporally and spatially unambiguous assignments of routing information. Every node maintains the highest recognized sequence number for each known node as destination sequence number (DSN). Source nodes send all routing messages together with a novel, increased sequence number. Receiving nodes discard all information with tagged DSNs lower than the DSN stored in the routing table. Nodes remove outdated routes and setup new routes whenever they receive information tagged to fresh DSNs. Every time nodes receive information in combination with DSN equal to the stored one, they update routes in case the information depicts a shorter route. Nodes discard copies of route requests and therewith ensure the loop freedom.

The following subsections describe all five observed protocols, but particularly emphasize the four not yet introduced protocols. Besides, it discusses the necessary characteristics of the embedded protocol implementation.

I. Ad hoc On-Demand Distance Vector Routing

The AODV routing protocol is described in detail in section 2.4.4 and therefore only a short introduction is given to conceive the following considerations. AODV combines an on demand route discovery mechanism with pure shortest-path routing. Additionally, sources do not have information about the exact path through the network, as every forwarding node only maintains next hop information. Therefore, route information within data packets is unnecessary. Since intermediate nodes discard duplicate RREQ messages, the RREQ flooding generally does not detect possible additional node-disjoint paths. The gray edges in Figure 11 show links never considered as possible parts of routes, because receiving nodes discard these RREQ messages as duplicates.

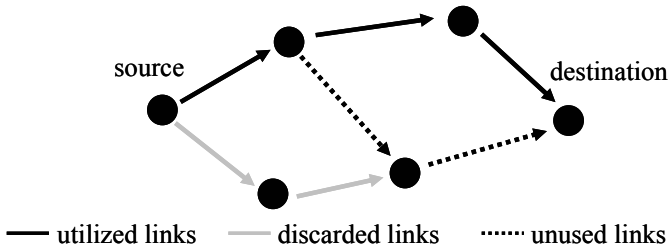

Figure 11: AODV path search query.

II. AODV and DSR multipath flooding extension

An extension of simple shortest path protocols are flooding based multipath algorithms. Figure 12 shows possible paths of this protocol extension. Backup-paths are usually non-disjoint. As described above, the flooding mechanism prevents the discovery of disjoint paths. Although these paths have links and nodes in common, they still improve the reliability while causing only little additional overhead. However, this extension is not usable in combination with AODV. It does not support the differentiation between multiple paths within intermediate nodes. An AODV flooding extension (AODV-FL) would require additional path-ID numbers together with the destination address within RREP packets. For different path-IDs, intermediate nodes must store the next hop information separately. Therewith, the source is able to switch to subsequent paths in case of errors on the primary path. However, as sources do not have knowledge about the exact path through the network, they are unable to determine which link is broken. Possibly, they switch to backup paths utilizing the same broken link. This in turn would cause further RERR packets and force the sources to switch to the next backup paths.

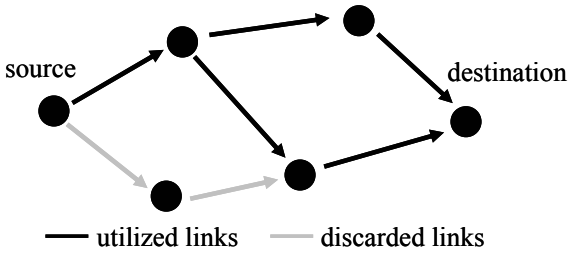

Figure 12: Discovered paths with a flooding based multipath extension.

To circumvent this unnecessary broken-link reutilization, protocols can use an extension of DSR, which supports the source routing algorithm. It is also an on-demand protocol, but in contrast to AODV, it does not require link tables within nodes. All packets contain the full path between source and destination. The route-discovery starts with RREQ flooding messages. Every forwarding node adds its own address to the routing header, after it checked for already forwarded duplicates. The destination possibly receives multiple RREQ packets. Each packet contains unique path information between the source and itself. The DSR flooding extension (DSR-FL) uses these additional RREQ messages to set up additional backup paths. For each received RREQ, destinations generate a new RREP message utilizing the reverse paths of the RREQ. The RREP packets contain the entire path information through the network. The source receives these messages and picks the primary path. Data packets always contain the full path within the routing header. Consequently, the routing overhead increases with increasing path length, while intermediate nodes do not need to store their next-hop neighbors. The

failure procedure of DSR is equivalent to the processing of AODV. An intermediate node realizing a permanent link error towards a downstream neighbor informs the source node with an RERR packet about this failure. As described, DSR sources know the exact paths to destinations and therewith can omit backup paths utilizing the detected broken links. This avoids usage attempts of already broken routes as it would occur with the AODV extension.

III. Disjoint Multipath AODV

The authors in [116] propose an AODV extension to discover multiple node-disjoint routes. As described, AODV uses a route request – route reply mechanism to discover a path between communication partners. Nodes drop additional copies of flooding messages. Therewith, intermediate nodes discard messages showing possible backup paths. The novel AODV multipath (AODVM) algorithm circumvents this disadvantage. Figure 13 illustrates the discovered disjoint paths.

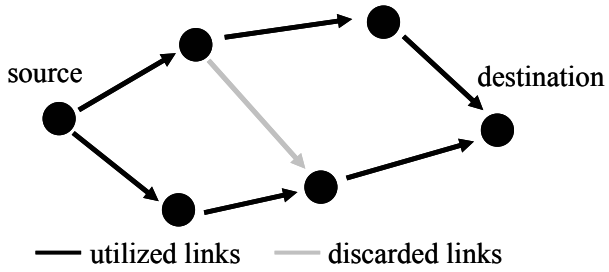

Figure 13: AODVM disjoint multipath discovery.

Intermediate forwarding nodes save all received RREQ messages. These messages contain the number of hops back to the source as additional information. The destination generates for each received RREQ an RREP message and returns it to the source, utilizing the path with the fewest hops. Each intermediate node, which receives the RREP message, retrieves the RREQ information with the fewest hops to the source. It forwards the message to the next hop neighbor stated in this particular RREQ. Afterwards the node deletes any stored information about the utilized link from its RREQ message table, which guarantees node-disjoint paths. In case intermediate nodes do not have any further stored RREQ information, they generate a route discovery error (RDER) as negative acknowledgment and return it in the direction of the destination. Upon reception of an RDER, intermediate nodes try to forward the message via a different path. After the reception of an RREP message, the source generates a route confirmation message (RRCM) as positive acknowledgement. It returns this to the destination to confirm the established path. The RRCM forces all intermediate nodes to discard any additional stored RREQ information and requests the destination to generate another RREP message. To reduce the network load, the source piggybacks the RRCM message onto the first data packet.

IV. Non-Disjoint Multipath AODV

As the name implies, AODV-non-disjoint multipath (AODV-NDM), utilizes AODV as basis and discovers non-disjoint routes from source to destination. AODV-NDM is a novel extension of AODVM and utilizes the same messages and follows the same basic ideas. Despite the original algorithm, nodes do not delete RREQ table entries after using them as outgoing links for RREP messages. Instead, they increase the hop metric for the

utilized links and keep them within the table. Subsequent RREP messages favor unused links. However, nodes use previously utilized links in case yet unoccupied links are not available anymore. Therewith, the algorithm generates mostly disjoint paths, while it is able to reuse some important links. As an example, Figure 14 depicts the additionally discovered non-disjoint path. In case, the algorithm generates only as many RREP messages as it received RREQ packets, it could at most generate as many paths as the legacy AODVM protocol. The other option would be, to create a predefined number of paths, independent from the received RREQs. In any case, the algorithm must limit the maximal number of return messages. The source receives all RREP messages and consequently additional error procedures during route setup are unnecessary. Therewith, the forwarding of RRCM messages or the creation of RDER packets back to the destination is avoidable. The destination delays all but the first individual RREP message for some predefined period to prevent inconsistencies within forwarding nodes.

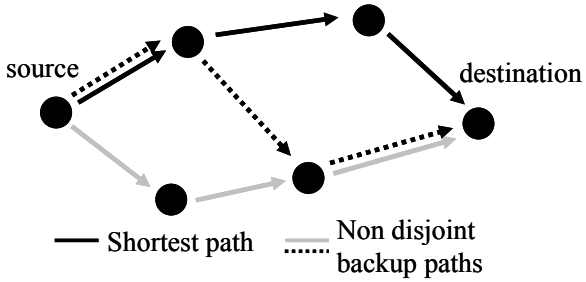

Figure 14: Route discovery with the non-disjoint extension of AODVM.

As described, AODV-NDM does not require any additional packets besides the RREQ, RREP, and RERR packets. Furthermore, the source is unable to determine if an occurred error on a path affects subsequent paths as well. It must use all M discovered paths subsequently and intermediate nodes report M individual errors for all paths before the source initiates a new RREQ cycle. Therefore, from a routing information exchange perspective, it is comparable to AODV-FL.

3.2 Related work

3.2.1 Theoretical path lifetime considerations

Several research groups investigate the link and path availabilities in MANETs. McDonald and Znati presented the first results in [118]. Their analysis utilizes an epoch-based mobility model based on the random waypoint mobility (RWP) model. To allow statistical evaluations, the mobility parameters speed and movement time are exponentially independent and identically distributed (IID), instead of uniformly distributed as proposed within the RWP model. Their definition of the link availability predicts the future status of wireless links and therewith depicts the probability over time that two nodes are in direct transmission range at time t_2, given that there is an active link at starting time t_0. The availability distribution follows the Kummer-Confluent hypergeometric function. However, this evaluation does not ensure that the link exists at time t_1 with $t_0 < t_1 < t_2$. And therefore no forecasts are possible about the time interval in which links exist uninterruptedly. The particular path availability is the product of all independent link availabilities within each path. However, the theoretical analysis does

not consider that intermediate nodes participate in two links and therewith link availabilities within multihop paths are not independent from each other.

The authors in [115] investigate the impact of different node mobility models on link and path lifetimes. They run excessive simulations with the network simulator ns-2 under various network conditions. Simulations show that the link duration for the RWP model approximately follows negative exponential distributions, while other mobility models have a multi-modal behavior. Interestingly, PLTs always follow an negative exponential distribution, independent from the underlying mobility model. Accomplished simulations show the impact of the path length h, the average relative speed v and the radio range R on the average PLT. While R linearly increases the PLT, the reciprocals of h and v are proportional to the PLT. Their simple first order analytical model for PLTs depicts that the distribution function is $f(x) = \psi \cdot e^{-\psi \cdot x}$ with $\psi = \lambda_0 \cdot h \cdot v/R$ and λ_0 a constant. They verified their assumption with the Kolmogorov Smirnov test. The second analytical model covers the relation between PLTs and network performance. The Pearson correlation test suggests that there is a linear relationship between the reciprocal of the average PLT and the network performance in terms of routing overhead. The same test shows that the throughput is proportional to the negative reciprocal average PLT. Therewith, the PLT is a good indicator to predict the performance of reactive routing algorithms. Although the authors introduce simple analytical models to predict average PLTs, they do not show a comprehensive theoretical analysis.

3.2.2 Further existing multipath routing protocols

The On Demand Multipath Distance Vector Routing (AOMDV) [119] is another multipath routing algorithm based on AODV. Unlike AODVM, it discovers link disjoint routes rather than node disjoint ones. Link disjoint routes may have nodes in common, and therewith do not guarantee that routes fail independently. In case the first hop of each path is unique, the flooding algorithm guarantees node disjoint paths between source and destination. AOMDV-RREQ packets contain a first hop field indicating the ID of the node after the source. Intermediate nodes maintain a list of unique first hops for each RREQ, representing disjoint paths to the source. The destination replies to the first arriving request and to a predefined number of additional requests received via unique neighbors, regardless of the first hop field. Intermediate nodes, receiving more than one reply, forward the packet via different return paths. This guarantees multiple link disjoint paths between source and destination. However, this certainly does not generate node disjoint paths, as the trajectories of different paths cross in intermediate nodes. The evaluation shows that the frequency of RREQs is significantly reduced in comparison to the legacy AODV. An improved packet delivery ratio and lower overall routing load comes along with it. Additionally, the average packet delay is reduced by more than 50%, because nodes do not have to interrupt packet delivery due to a route reestablishment.

The optimal number of disjoint paths to maximize the overall, combined paths lifetime is limited. The achieved benefit in path lifetimes shrinks with any additional succeeding path. Any backup path must be valid when used. Due to continuous topology changes, this becomes more unlikely for increasing times between route discovery and usage. The authors in [120] observed that the optimal number of paths within a set of paths is three for link-disjoint paths.

The caching and multipath (CHAMP) routing protocol [121, 122] is a non-disjoint multipath algorithm. It is based on the temporally-ordered routing algorithm (TORA) [76] and therewith is a pure on-demand routing algorithm utilizing direct acyclic graphs (DAG). Every node maintains not less than two paths to every active destination. In case of link breaks on both paths, nodes reply RERR packets containing the header of the undelivered data packet to the previous node. This node uses the still cached packet and its remaining second paths for continuous packet delivery without the necessity to inform the source about the path break. During the RREQ, nodes create a DAG with the source as root node. Each node maintains a table with the number of forwardings for every source-destination pair and the node IDs of previous hops towards the destinations. CHAMP only utilizes paths with hop length equal to the length of the SP. It expects nodes to forward packets to the least used next hop neighbor in order to spread packets in a round robin fashion. This behavior optimizes the usefulness of caches in intermediate nodes and nodes keep their routes fresh. However it also induces notable overhead and causes difficulties with standard TCP versions. Due to the variable packet run times over different routes, destinations receive packets out of sequence and initiates congestion avoidance mechanisms to reduce the data rate. CHAMP does not utilize optimized multipath strategies during the route discovery process and therewith the strategy is comparable to AODV-FL. Intermediate nodes utilize RREQ as well as RREP packets to update their distances to sources and destinations. However, both packets only reveal the shortest path to the originating nodes. It remains unknown, how intermediate nodes get knowledge about additional paths. As described, the algorithm necessitates that all contributing nodes utilize at least one additional route to each destination. This is difficult to fulfill, especially when considering that backup routes must use an equal number of hops compared to the shortest path. MANETs often have just a single route between source and destination, and therefore the requirement of equal hop numbers remains unconsidered.

The authors of [123] evaluated different multipaths routing strategies for wireless sensor networks (WSN). As mentioned, sensor devices usually are stable in their position and therewith significantly different to mobile ad hoc nodes. Therefore, the utilizable routing algorithms are optimized for stable networks and link and node failures are not necessarily isolated. However, the proposed simulation results show some new insights. The authors consider resilience, energy efficiency, and overhead as the most important parameters for WSNs. Braided (non-disjoint) routing strategies are more resilient than disjoint strategies. They improve the resilience by 50% for isolated failures and by 30% for patterned failures. Additionally, braided routing strategies require less maintenance overhead and therewith consume less energy than their disjoint counterparts. As described, WSN differ from MANETs and therefore the shown results are not completely adaptable to MANETs.

The Split Multipath Routing (SMR) protocol [124] is based on DSR. It creates two maximally disjoint paths which are not necessarily of equal length. Both paths are used in a round robin fashion for data packet delivery. SMR utilizes flooding for route discovery. Unlike DSR, it prevents intermediate nodes from replying cached routes back to the source and it does not always drop duplicate RREQ messages. Instead a node forwards every RREQ packet, on the condition it has received the duplicate from a different neighbor, but with an equal hop length compared to the path of the original RREQ message. Therewith, the destination certainly receives multiple RREQ messages over non-disjoint paths. It chooses the path of the first incoming message as primary

paths and chooses a second one utilizing maximally disjoint nodes and with the fewest number of hops. The evaluation depicts that the initiation of a route recovery after the first route break is unfavorable. The additional induced overhead of more frequent route discoveries cause more packet loss and even larger overall packet delays. As shown in previous comparisons of MANET routing algorithms, the source routing of DSR and SMR causes notable overhead. Additionally, the forwarding of RREQs multiple times worsens the difficulties with RREQ broadcast storms and therewith induces even more overhead and packet drops. Whereas the above proposed AODV based multipaths routing algorithms improve the network performance, while they simultaneously circumvent all theses shortcomings.

The prediction of link lifetimes saves overhead, in case permanent route errors occur more seldom and consequently routing algorithms must initiate RREQs less frequently. A straight forward approach is the utilization of GPS information to improve route stability. However, the utilization of geocast [125] routing algorithms like GeoTORA [126] or GeoGrid [127] requires the approximate position of the destination. In case the destination positions are unknown, the results in [8] depict that the necessary efforts to obtain position information exceeds the advantage of exact link lifetimes.

In order to circumvent the utilization of peripheral information, the authors in [128] introduce an algorithm to predict the future status of wireless links without the help of external sensors. They investigate different link caching schemes to improve the overall network performance. The algorithm is able to distinguish between low and high mobility nodes. It certainly prefers reliable nodes with many uninterruptedly available links. The continuous examination of connectivity to neighboring nodes enables the algorithm to predict the future movement behavior. The selection of reliable nodes for paths through the network improves paths lifetimes and therewith saves considerable network resources. However, the authors only introduce an algorithm and verify its usefulness with simulations, but do not describe any analytical model. The paper does not discuss the relation of paths lengths on route lifetimes and the impact of different routing strategies on the overall network performance.

The authors in [129-131] propose a framework to use a set of multipath routes simultaneously. The algorithm splits information among the multitude of paths. It adds redundancy information to each packet and transmits them over all available paths in a round robin fashion. Even in an error prone environment with numerous topology changes the destination receives some packets and therewith is able to reconstruct the complete information. Obviously, the algorithm causes more overhead than single paths algorithms and therewith restricts the processing to fewer simultaneous connections. The proposed scheme does not include a routing algorithm, but expects the presence and knowledge about multiple node disjoint paths. The papers do not evaluate different routing strategies. The authors depict that their algorithm performs best with up to 30 node-disjoint paths. This setup increases the probability of successful transmissions by 25%. However, following the evaluations in [132], the optimal throughput and minimal routing overhead is reached with about ten nodes on average within direct radio range. Simulations in section 3.5 show that networks with this node density rarely contain more than five node disjoint paths. Increasing the node density causes a significant increase in overhead due to broadcast storms [133] during route discoveries.

3.3 Path lifetime analysis for two-hop connections

The generally induced overhead within MANETs occurs during route discoveries. The flooding approach of reactive routing algorithms requires that all participating nodes forward the request once. Therefore, the lifetime of a route is a crucial parameter for the overhead calculation. The path lifetime (PLT) is defined as time from path setup between source and destination and the occurrence of a permanent failure on any utilized link. A route failure occurs whenever any of the utilized links breaks. This definition is based on the description given in [6]. Consequently, the link lifetime (LLT) is defined as the period between the setup of a route and the occurrence of a permanent error of this particular link. Thereby, the point in time of the initial emergence of a link is unimportant, as reactive routing protocols use links only on demand. Within multipath sets, the PLT represents the time between the route discovery and the route break of the most durable path. As link errors occur statistically independent, backup paths may break before source nodes intend to utilize them. Therefore, the most durable path is not necessarily the last one within the set. As described, the LLT and PLT distributions depict the probability that an arbitrary link or route (set) exists after a certain time. It can be represented as a complementary cumulative density function (cdf). The probability distributions are one at startup and monotonically decrease over time. Theoretically they reach zero at time $t \rightarrow \infty$.

The utilization of local repair algorithms as proposed in all ad hoc routing protocols is irrelevant, as they only allow a temporary improvement. Repaired routes usually are longer than the original paths and they reutilize most of the previously used links. The algorithms do not induce network wide route reestablishments and therefore path breaks occur more quickly than with recently discovered routes. Additionally, as all protocols are able to utilize local repair algorithms, none of the protocols improve their performance in comparison to other algorithms.

The first examination focuses on simple ad hoc scenarios with only two hops. The scenario is derived from the observation that a mobile user wants to start an online application on its wireless terminal. For simple operation the user stops moving, takes the device out, and starts the application. In order to setup a connection, the routing protocol of the terminal searches for a local ad hoc–Internet gateway in its proximity.

Protocols supporting uni-directional links usually create non-symmetric routes between sources and destinations, as nodes may not receive any information from downstream neighbors. However, common ad hoc routing algorithms require bidirectional links between neighbors, otherwise they fail to function properly. Therefore, the following analysis only considers bidirectional links. This implies that all nodes transmit radio wave signals with identical energy. Consequently the radio transmission range R is equal and constant for all nodes. Additionally, a simplified channel model is considered. The coverage area around each node forms a perfect disk with radius R. Nodes within the coverage are able to directly communicate with this node, while greater distances prevent direct links.

The user terminal n_2 searches for an AP within its neighborhood. In the scenario depicted in Figure 15, the distance a between the node n_2 and the AP n_0 prevents a direct node-to-node connection. The limited radio transmission range R ($R < a$) is insufficient to achieve direct access to n_0, and therefore n_2 reaches n_0 only over a two hop path.

Since the Internet gateway is certainly fixed in position and the user does not move as well, source and destination are stationary.

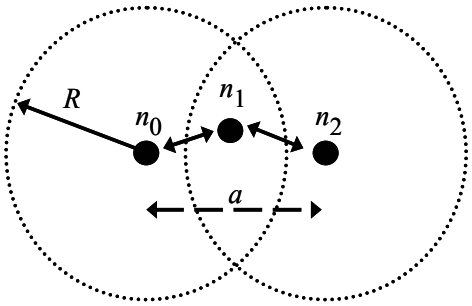

Figure 15: Connection between the nodes n_0 and n_2. Node n_1 relays the connection.

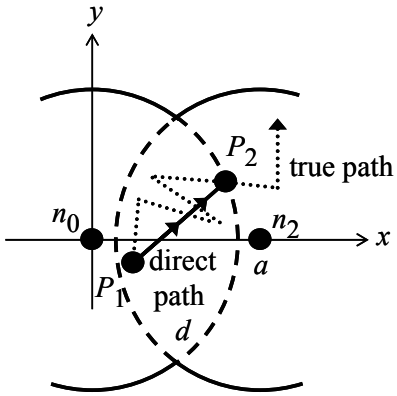

Figure 16: Relation between true movement pattern and the direct path nodes.

Terminal n_2 uses its routing algorithm to discover the shortest path to the AP. All existing routing algorithms perform equal in this simple scenario, independent whether they are based on a proactive or a reactive scheme. Nodes utilizing proactive routing constantly receive topology update messages. Therefore, they are aware of the route over node n_1 in advance. In contrast to that, reactive routing algorithms usually do not cache information about certain routes and therefore create an RREQ. Source n_2 forwards it to the intermediate node n_1. It receives the packet and forwards the request to the destination n_0. The destination receives the RREQ and creates an RREP back to n_0, using n_1 as relay. Thus, node n_2 gets knowledge about the route to n_0 as well. Following packets bound for gateway n_0 always use n_1 as relay. Considering the channel model, the intermediate node n_1 must be inside the coverage of the source node n_2 and the destination node n_0 to act as relay. Concerning the shape of the combined coverage area, this is called the "eye of coverage" (see Figure 15).

In order to obtain reasonable PLTs, knowledge about realistic node movements is necessary. However, mobility algorithms to model real world ad hoc node movements are currently not available in the literature. The vast variety of scenarios makes it impossible to create one model which covers all environments. For simulations of MANETs, numerous different approaches exist reproducing human movement under various conditions [134]. The movements of general mobile ad hoc nodes are independent from other nodes and they are not restricted in any way. Therefore, nodes within the following scenarios choose their velocities, moving angles and pause times independent from other nodes. As most ad hoc network simulations utilize mobility generators based on the random direction (RD) model, the following evaluations utilizes it as well (see Figure 16). More details about the mobility model can be found in section 2.5 or e.g. in [135].

Within the proposed scenario, the relaying node is not restricted in its position. Hence it moves independently, and after a certain period of time T it leaves either the proximity of n_0 or n_2. However, while node n_1 is within the transmission range of both other nodes, it relays the connection. Following the definition of PLTs, the moving node n_1 starts relaying the connection between n_0 and n_2 at time T_1 at the random initial position P_1.

The relaying node moves with constant velocity v_1 in direction φ_1 for a period of time τ_1, remains stationary for pause time τ_p and starts moving again with the parameter triple $[v_2, \varphi_2, \tau_2]$. This continues infinitely in the model.

The analyzed connection of n_2 to n_0 breaks as soon as n_1 leaves the proximity of either node n_0 or n_2, at the exit point P_2 and time T_2 with $T_2 = T_1+T$. The movement pattern of node n_1 depends on the average node velocity v_{av}, the average movement period τ_m of node n_1, and the pause times τ_p between consecutive movements. However, it is independent from the movement angles φ, as they are uniformly distributed with $0 \leq \varphi \leq 2\pi$. In order to allow a straightforward analysis, the proposed simplified model [118] does not contain turns during movements. Therefore, node n_1 moves linearly with the constant velocity v. It starts moving from P_1 in direction θ and leaves the proximity of either node n_0 or n_2 at the exact same exit point P_2 as in the original model.

Both scenarios are similar, in case only restricted areas of interest (the eye of coverage) are considered and simple scenarios are modeled. To create comparable models, the moving speed v of node n_1 within the simplified scenario must be adapted to the mobility parameters $[v_{av}, \tau_m, \tau_p]$ of the previous scenario. Due to the pause times τ_p and the zick-zack movements of RD controlled nodes, velocities for the simplified model are below the velocities of an equivalent RD movement. For any further calculations the evaluation uses the simplified scenario. Within this simplified model, the coordinates x and y of the initial point P_1 as well as the moving direction θ are uniformly distributed.

The following section theoretically analyses the probabilities that node n_1 leaves the eye of coverage after a certain distance d. It includes scenarios with fixed distance a between source node n_0 and destination n_2 as well as a uniformly distributed distances a. In order to obtain an universally valid analysis for this scenario, the subsequent section 3.3.2 considers arbitrary node movement velocities as well.

3.3.1 Leaving-probabilities for certain node movement distances

As depicted in Figure 15, R is the maximal distance for a single hop connection. Further more a is the distance between both stationary nodes n_0 and n_2, with $R < a < 2 \cdot R$. The coordinates of the initial node position P_1 (see Figure 16) lie in the eye of coverage. They are uniformly distributed over all possible values x and y.

At the exit point P_2, node n_1 leaves either the scope of node n_0 or n_2. The exit point can be any point on the dashed lines in Figure 16. It depends on the moving direction of node n_1 and therefore depends on the movement angle θ. The value of the length d is the Cartesian distance $\overline{P_1P_2}$ between the start point P_1 and the exit point P_2 of node n_1's movement within the scope of both other nodes. In order to calculate the distribution function for the PLTs, the distance d and the initial node position P_1 are assumed to be known. The probability p that the node remains within the scope of both nodes is proportional to the angle $\gamma = 2 \cdot \pi - \alpha$ of the segment of a circle c_P around P_1 with radius d, which does not intersect with both coverage circles (see Figure 17 right). In case both circles do not intersect, γ is maximal and is given through $\gamma_{max} = 2 \cdot \pi$. Therefore, the exact probability is given through

$$p(d, P_1) = \frac{\gamma}{\gamma_{max}} = \frac{2\pi - \alpha}{2\pi} \quad (1)$$

As described, the node may leave the coverage of either node n_0 or n_2. Due to the symmetry of the scenario, the probability that n_1 leaves the coverage of n_0 is equivalent to the probability that n_1 leaves node n_2's neighborhood. Therefore, without loosing generality, the theoretical analysis omits the calculation of the probability that n_1 leaves node n_2's proximity. As a result, the maximal angle γ_{max} alters, and is not the full circle $2\cdot\pi$ anymore. The new maximal angle γ_{max} is limited to the opening angle formed by the points Φ_1 and Φ_2 together with P_1. The coordinates of the intersection points Φ_1 and Φ_2 of both areas of coverage are given through

$$\Phi_{1/2} = \begin{pmatrix} a/2 \\ \pm d_{max}/2 \end{pmatrix} \text{ with } d_{max} = \sqrt{4 \cdot R^2 - a^2} \tag{2}$$

whereas the initial node position P_1 has the distance r from the origin with $r^2 = x^2+y^2$. Figure 17 clarifies the relation. The lines R, r and l_1 form a triangle, with l_1 the distance between Φ_1 and P_1 and δ_1 the opening angle between r and l_1 ($\delta_1 = \sphericalangle(r, l_1)$). With the help of the law of cosine, δ_1 is calculated as

$$\cos(\delta_1) = \frac{R^2 - r^2 - l_1^2}{2 \cdot r \cdot l_1} \tag{3}$$

The other angle δ_2 is calculated accordingly. As depicted in Figure 17, the new maximal angle γ_{max} depends on the intersection points Φ_1 and Φ_2 and is given through $2\cdot\pi-\delta_1-\delta_2$. Additionally, the combined striped angle in Figure 17 depicts the new angle γ, which is described as $2\cdot\pi-\alpha-\delta_1-\delta_2$. Consequently, the calculation for the probability p changes as well. The probability p is now calculated as

$$p(d, P_1) = \frac{\gamma}{\gamma_{max}} = \frac{2\pi - \alpha - \delta_1 - \delta_2}{2\pi - \delta_1 - \delta_2} \tag{4}$$

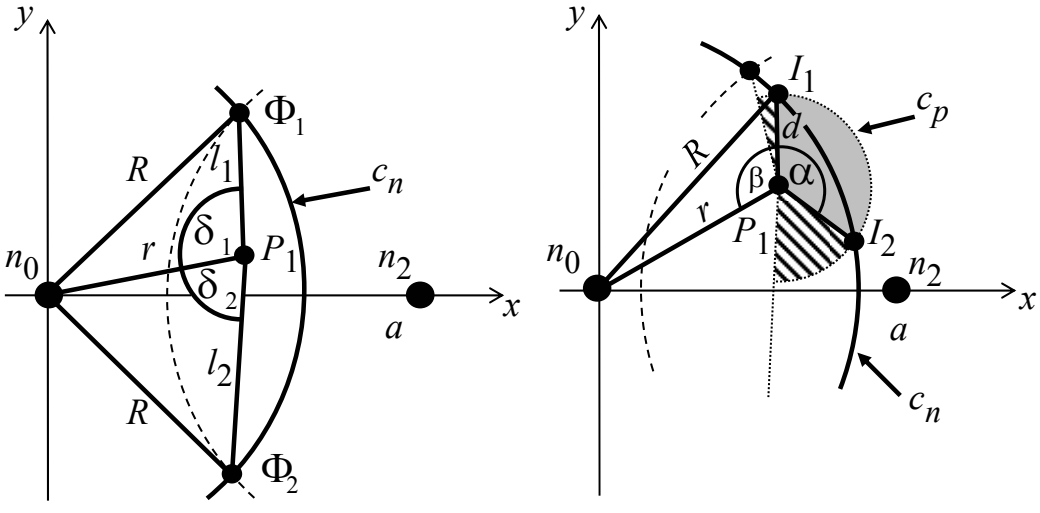

Figure 17: Deviation of the probability p that node n_1 stays within the eye of coverage for known P_1 and d.

To calculate the angle γ, the knowledge about the intersection points between the circle c_P with radius d around P_1 and the circle c_n, depicting the coverage area around n_0, are necessary. Depending on the initial node position P_1, the equation to calculate the intersection points I_1 and I_2 is either without solution, has one or has two solutions.

The discussed scenario is symmetric to the x-axis as well. Therefore, and without loosing generality, calculations only consider initial node positions P_1 with $y>0$. For the evaluation of the PLT-probabilities, three different calculations are necessary. The appropriate scenario depends on the considered distance d. Figure 18 shows all scenarios. In case $a<6 \cdot R/5$, the term $(2 \cdot R-a)^2$ is greater than $R^2-(a/2)^2$ and therewith the middle scenario in Figure 18 is unnecessary. For all other cases, all three scenarios are required.

For distances d smaller than $2 \cdot R-a$, the scenario on the left is essential. In case $r<R-d$, the circle c_P does not intersect with c_n .The grey region in the left picture depicts this area. Circles c_P with larger distances r from the origin basically have two intersections with c_n. However, one intersection I between c_P and c_n may be out of the observed range of angels. These areas are striped as grey regions in Figure 18. The center of this area is always the intersection point Φ_1 with radius d. All points P_1 within the remaining region of the eye of coverage generate two valid solutions for the intersection calculation. Both other figures are illustrated accordingly.

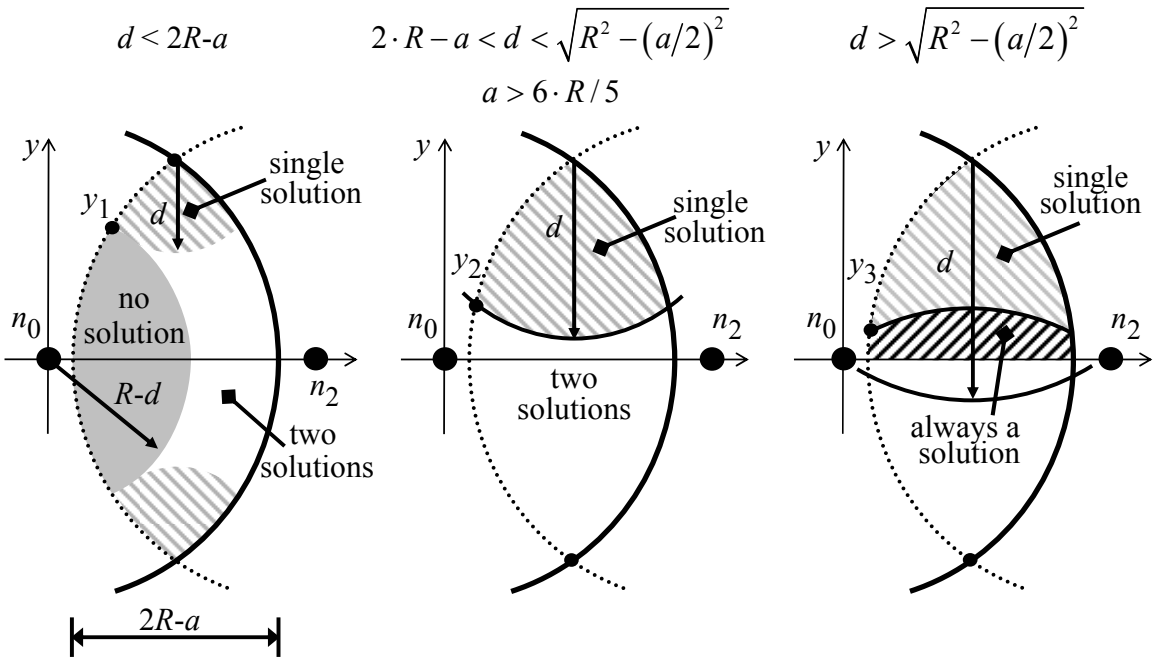

Figure 18: Necessary scenario differentiations for variable distances *d*.

The complementary probability $q = 1-p$ that n_1 leaves the scope of n_0 when moving a distance d is α/γ_{max} (see the grey shaded area in the right picture of Figure 17). Together with the radio range R, r and d form a triangle. Again, with the help of the law of cosine, the opening angle β between r and d ($\beta = \sphericalangle(r, d)$) is

$$\cos(\beta) = \frac{R^2 - r^2 - d^2}{2 \cdot r \cdot d} \tag{5}$$

Both triangles for both intersection points I_1 and I_2 are equivalent and therewith β as well. In case, both intersections are within the valid sector of c_n the probability q_2 for two solutions is

$$q_2(d,P_1) = \frac{2\cdot\pi - 2\cdot\beta}{2\cdot\pi - \delta_1 - \delta_2} \tag{6}$$

Otherwise one angle β must be replaced with δ_1 and the single solution probability q_1 is calculated as

$$q_1(d,P_1) = \frac{2\cdot\pi - \beta - \delta_1}{2\cdot\pi - \delta_1 - \delta_2} \tag{7}$$

For large distances d with $d^2 > R^2-(a/2)^2$, both intersections may be outside of the observed sector of c_n. In this case, both angels β must be replaced with $\delta_{1/2}$. Therewith, the contrary probability is constant and

$$q_{max}(d, P_1) \equiv 1 \tag{8}$$

The black striped region in the right picture of Figure 18 depicts the area where q is equal to one. The area borders y_1, y_2, and y_3 within the respective scenarios can be calculated solving simple rationale equations of the second order. With the knowledge about the probabilities q for certain triples (x,y,d) and the belonging areas within the eye of coverage, the distribution function $f(d)$ for the PLTs is the 2D-integral over all possible initial node positions $A(x,y)$.

$$f_{d,a}(d,a) = 1 - \iint_{A(x,y)} q(d,x,y)\,dxdy \tag{9}$$

A closed form solution of this integral is not possible. Further evaluations are based on numerical calculations. The formula in (9) is only valid for a known distance a between both communication endpoints. Obviously, the exact distance a is unknown but the probability for a certain distance is uniformly distributed. Therewith, the general distribution function for arbitrary distances between both communication endpoints can be calculated as

$$f_d(d) = \int_{a=R}^{2R} f_{d,a}(d,a)\,da = 1 - \int_{a=R}^{2R} \iint_{A(x,y)} q(d,x,y)\,dxdy\,da \tag{10}$$

After the exact description of the used theoretical model, some results follow for various relative distances a/R between both communication endpoints as well as for the variable case. The radio range R is constant and one. Figure 19 shows the distribution functions that the node n_1 remains within the proximity of the source and the destination node after a certain distance d. The left picture depicts the probabilities for some distinct a/R ratios, whereas the right figure depicts results for the variable case. Both graphs show simulation results as well. The simulations for certain fixed a/R ratios are based on 10^6 independent results whereas the simulations with variable a/R contain 10^8 replications.

The graphs in both figures clearly depict that the simulation corresponds with the theoretical analysis. The differences are marginal. The results in [11] proof that the mean E as well as the variance σ is proportional to the radio range R ($E \sim R$ and $\sigma \sim R$). Therefore, the behavior of both parameters is independent from the radio range.

The first theoretical analysis examines the characteristics of $f_{d,a}(d,a)$ for a fixed a/R ratio. As expected, both values E and σ have their maximal with a minimal ($a = R$). From their maximals, both parameters decrease and turn to zero at the maximal distance between n_0 and n_2 ($a = 2\cdot R$). The characteristic of the deviation show that randomly created

movements can heavily differ from the calculated average. Especially if $r \rightarrow 2 \cdot R$, σ is greater then E ($\sigma > E$).

Obviously, the largest probabilities $f_d(d)$ occur at $d = 0$. The probability equals one that the node remains within the eye of coverage if the node is stationary ($d = 0$). Greater values of d become more unlikely. The probabilities near the possible maximal distance d_{max} are almost zero. The occurrence of a distance $d = d_{max}$ is only possible twice. The initial node position P_1 must be at Φ_1 or Φ_2 and the node movement must be exactly parallel to the y-axis towards the opposite coverage intersection point $\Phi_{1/2}$.

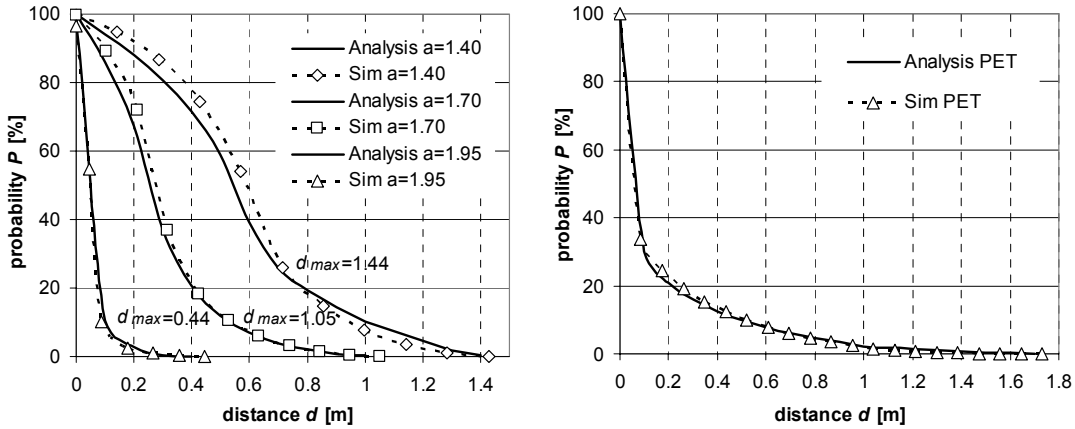

Figure 19: Complementary cdf distributions that node n_1 remains within the proximity of both other nodes after a certain distance d.
Left: constant a/R ratio, Right: variable a/R ratio.

The probability that a randomly measured distance d is greater than the average $E(d)$ ($P(d > E(d))$) decreases with increasing a/R ratios. Distances larger than the mean E are more unrealistic, even if the mean $E(d)$ decreases with increasing a/R as well. This behavior is verifiable in all distribution functions $f_d(d)$, independent from the used a/R ratio. For small a ($a \rightarrow R$), distances greater than $d_{max}/2$ have higher probabilities than for large values of a ($a \rightarrow 2 \cdot R$). This indicates an over-proportional reduction of probability-values from 0 to d_{max} with increasing a/R ratios. The shape of the eye of coverage causes this behavior. For small a/R ratios, the eye is approximately a circle, and large distances d are common. Whereas for large a/R ratio, the eye has more the shape of a rectangle with unfavorable width-to-height ratios. Thereby, large values of d become almost impossible.

After the evaluation of the behavior of $f_{d,a}(d,a)$ for a fixed a/R ratio, an evaluation with a uniformly distributed a/R ratio follows. As in the previous section, the calculated values are independent from the used radio transmission ranges R and are valid in general. The distribution function $f_d(d)$ has an extreme progression. Distances d close to zero are highly probable, whereas values greater then $d_{max}/2$ are unlikely. The probability for distances close to d_{max} are almost zero. Despite the high probabilities for small values of d, the mean $E(d)$ is $0.29 \cdot R$. The deviation $\sigma(d)$ is nearly as high as the mean $E(d)$, and has the value $0.27 \cdot R$. Therefore, it is very likely that a randomly measured distance heavily differs from the mean $E(d)$. The probability of measuring smaller values than $E(d)$ is much higher, than the probability of receiving larger values. The probability to measure distances greater than the mean E is only about 40%. Consequently, particular

lifetimes will always differ significantly from a statistical mean value and paths will frequently have lifetimes below the average.

3.3.2 Path lifetimes for arbitrary node movements

The above introduced functions only concern the random variable d. They fail to make any assumptions about the time a path exists between the source node n_0 and the destination node n_2 using the node n_1 as relay.

Hence, some further calculations are necessary to get some results regarding the PLT t of a connection within this scenario. The time t is calculated as the division of d with the velocity v of node n_1. Fortunately, the random variables d and v are independent of each other, thus the calculation for the PLT function $f_t(t)$ can be split into two steps, the first one for $f_d(d)$ and the second one for $f_t(t)$. Therefore, $f_t(t)$ can be expressed as

$$f_t(t) = \int_{d=0}^{d_{max}} \int_{v=v_{min}}^{v_{max}} \frac{d}{v} f_v(v) f_d(d)\, \partial v\, \partial d \tag{11}$$

The random variable v is always positive ($v > 0$), as negative velocities are equivalent to positive velocities in the opposite direction. A velocity equal to zero leads to infinite PLTs. Therefore, the minimal velocity v_{min} must always be above a minimal threshold value. Within the following calculations, velocities are uniformly distributed with the finite limits v_{min} and v_{max}, and have an average $E(v)$ of

$$E(v) = 1/2 \cdot \left(v_{min} + v_{max}\right) \tag{12}$$

As described above, the assumed velocities within this model are below the velocities used for a regular RD scenario. To calculate the minimal and maximal velocities of the simplified model, pedestrian speed is assumed. The maximal velocity for the RD model would be $v_{max} \approx 5$ km/h = 1.39 m/s. Hence the decreased velocity parameters to calculate (11) are set to $v_{max} = 1.2$ m/s and $v_{min} = 0.2$ m/s with $E(v) = 0.7$ m/s.

For a reasonable approximation of PLT times and corresponding probabilities a constant radio transmission range R of 100 meters is assumed. Further on, the mentioned parameters for the maximal and minimal velocities are considered. Table 3 shows various characteristics for certain distances a between source and destination node. The time $T_{max} = d_{max}/v_{min}$ is the maximal time a connection can last with the above made assumptions, the minimal time T_{min} is always zero. $P(t < ½\ E(t))$ is the probability that a randomly measured time t is smaller than half of the mean $E(t)$, whereas $P(t > E(t))$ is the probability that a measured time t is larger than the mean $E(t)$.

Table 3: average PLT for different distances *a* between source and destination.

a in [m]	t_{max} in [s]	$E(t)$ in [s]	$P(t<½E(t))$ in [%]	$P(t>E(t))$ in [%]
100	866.0	131.8	15.7	36.7
140	714.1	83.9	19.3	36.7
170	526.8	44.2	23.0	34.0
195	222.2	8.3	26.1	29.3
variable distance a	866.0	34.1	54.9	30.0

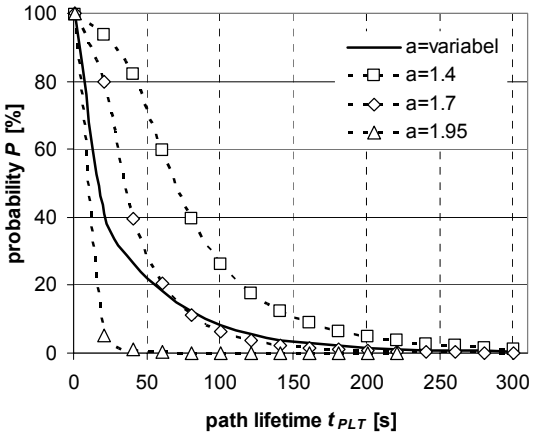

Figure 20: PLT distributions for constant and uniformly distributed distances *a* between source and destination node.

As shown in Table 3, the average PLT is 34.1s, for a variable distance *a* between source and destination. This implies that a two hop connection between two stationary users and a mobile relay will last for about 35 seconds on average. After that time, the link from the relay to either the source or the destination node usually breaks and a new route discovery process is necessary. However, the possibility that a link will last for more than 35 seconds is only 30%, whereas the possibility that a link break happens within the first 17 seconds is about 55 %. Figure 20 shows the PLT- function for the variable *a*/*R* case and for several constant a/*R* cases. The y-axis depicts the probability that an arbitrary path is still functional after a certain amount of time.

The unpredictable terminal mobility of the moving node n_1 prevents reasonable predictions of PLT times even within this simple scenario. However, even this basic two hop ad hoc network scenario permits basic statistical investigations. The expected failure times show large deviations and the probabilities for PLTs below the average are more likely. The average achievable path lifetime is unexpectedly low (34s), when keeping in mind that the source as well as the destination is stationary and the relay only moves with pedestrian speed. As this scenario is not valid in general, the following section contains an extended model to theoretically analyze path lifetimes with an arbitrary number of participating mobile nodes.

3.4 Path lifetime evaluation for arbitrary multi-hop connections

The previous evaluation only covers a simple and particular scenario. The theoretical analysis in this section extends the scenario and investigates the PLT for general MANET connections with an arbitrary number of hops. Therefore, all nodes within the recent scenario are mobile, and are initially uniformly distributed within a simulation area. As described in section 3.3 and in [11], only the time of the first link break is relevant for the determination of PLTs. This limitation simplifies the model behavior, as further calculations beyond the first link break are unnecessary.

In contrast to the previous mobility model, not only the relaying node moves, but all nodes. However, the simplified mobility behavior remains equivalent. Nodes move on a straight line with a uniformly distributed direction between 0 and 2π as well as a uniform distributed velocity between 0 and v_{max}. Therefore, any two nodes will leave each other's proximity after a certain time Δt. In case both nodes have an initial velocity equal to zero ($v_1 = v_2 = 0$) or both nodes are moving in the same direction with identical velocities ($|v_1| = |v_2| \cap \varphi_1 = \varphi_2$), the LLT between these two nodes will turn to infinity. This behavior leads to unrealistic lengthened LLTs, but only for single hop networks. In multihop scenarios, the probability of multiple infinite LLTs between any two nodes contributing to a route is almost zero, and therefore negligible.

3.4.1 Link lifetime distributions of the initial hop

The following section 3.4.2 will illustrate that the lifetime probability distributions of different links within routes are not identical. The lifetime probability for the first hop is unique, while the lifetime probabilities for all subsequent hops are identical. Therefore, a differentiation is necessary and consequently this section initially focuses on the link lifetime distribution of the first hop.

The difficulty with two independently moving nodes can be converted into an approach with one node in motion and one stationary node [118]. Therefore, without loosing generality, the relative difference between both velocity vectors $\vec{v}_1$ and $\vec{v}_2$ is

$$\vec{v} = \vec{v}_1 - \vec{v}_2 \tag{13}$$

with v the absolute value of the velocity vector $\vec{v}$.

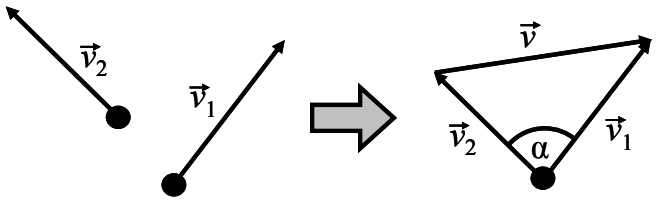

Figure 21: Relative velocity vector *v*.

Following [136], the probability density function $f_v(v)$ of the absolute value v is

$$f_v(v) = \frac{1}{\pi} \iiint_{v_1, v_2, \alpha} \delta\left(v - \left|\vec{v}_1 - \vec{v}_2\right|\right) dv_1\, dv_2\, d\alpha \text{ with } \left|\vec{v}_1 - \vec{v}_2\right| = \sqrt{v_1^2 + v_1^2 - 2 \cdot v_1 v_2 \cos(\alpha)} \tag{14}$$

in which δ(…) describes the Dirac delta function. The function gives the probability that two randomly chosen velocity vectors result in a subtracted absolute value v, with v_1 and v_2 the corresponding absolute values of the initial vectors $\vec{v}_1$ and $\vec{v}_2$ and α the angle between both. The integral parameters v_1, v_2 are within the range $[0..v_{max}]$ while α covers $[0..\pi]$.

The formula in (14) is a general description. To solve the integral a different approach is necessary. The probability distribution $f_v(v)$ solely depends on the velocity v. To calculate $f_v(v)$, v is initially set constant. In case, v_1 is set constant as well, the ranges of the remaining velocity v_2 and the angle α forms a disk with radius v_{max}. The disk is colored grey in Figure 22. With known v and v_1, the arrangement has solutions located on the circle with radius v. The respective solution L is the length of the covered

segment of the circle v. The author in [137] illustrates that uniformly distributed random variables on a circle do not rise linearly but are proportional to their square root. Therewith, L is

$$L = \frac{1}{\pi} \cdot \beta \cdot \sqrt{v} \tag{15}$$

Due to the symmetry of the configuration, only the upper part of circle v is considered. Depending on the parameters v and v_1, three differentiations are necessary. Table 4 shows the respective cases. The disk with radius $v_2 = v_{max}$ either covers the full circle v, covers parts (as depicted in Figure 22) or does not intersect at all. In case the solutions describe a circle segment, the lines v, v_1 and v_2 form a triangle with β_{seg} describing the opening angle between v and v_1. With the law of cosine, β_{seg} is

$$\beta_{seg} = \cos^{-1}\left(\frac{v_1^2 + v^2 - v_{max}^2}{2v_1 \cdot v} \right) \tag{16}$$

Table 4: Case differentiations for the angle β.

$\beta = ?$	$v < v_{max}$	$v > v_{max}$
$v_1 \in [0, \lvert v_{max}\text{-}v \rvert]$	π	0
$v_1 \in [\lvert v_{max}\text{-}v \rvert, v_{max}]$	β_{seg}	β_{seg}

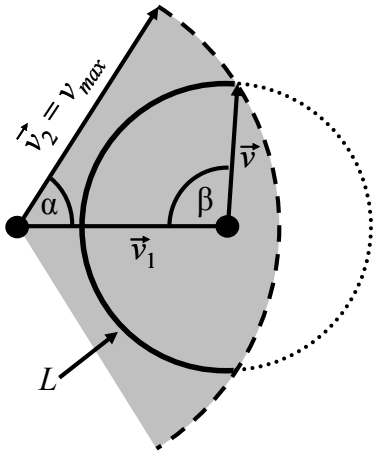

Figure 22: Velocity distribution.

Therewith, the remaining integral is

$$f_v(v) = \int_{v_1} L \, dv_1 \tag{17}$$

As the integral (17) has no closed form solution, a transformation of (16) with the help of a Taylor series is necessary. A fifth order approximation shows sufficient accuracy. Therefore, the function $\cos^{-1}(x)$ has the Taylor series

$$\cos^{-1}(x) \xrightarrow{Taylor} \frac{\pi}{2} - x - \frac{1}{6}x^3 - \frac{3}{40}x^5 + O(x^6) \tag{18}$$

Therewith, a closed form solution is possible. Solving the integral in (17) is straight forward, but creates a large formula, and therefore is not presented. The resulting pdf function $f_v(v)$ of the new velocity vector $\vec{v}$ is shown in Figure 23. It has been calculated with a maximal node velocity $v_{max} = 1$. For displaying purposes, the following pdf functions $f_v(v)$ is scaled in order to meet the requirement $max(f(x)) = 1$. The figure additionally shows simulation results. The simulation includes 10^6 independent replications with randomly generated initial velocity vectors. The simulation uses the

random number generator presented in [138]. The simulative results of Figure 23 are shown in [16] as well.

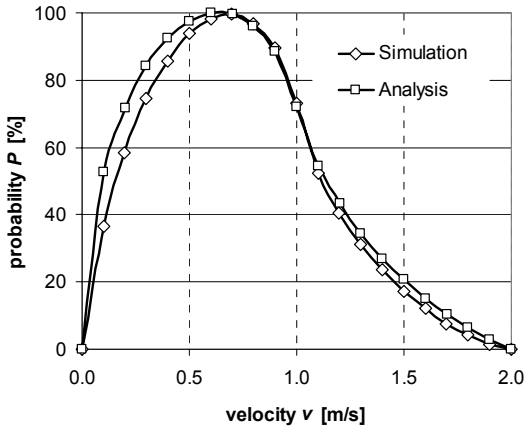

Figure 23: Theoretical analysis and simulation results for the relative velocity distribution $f_v(v)$ between two moving nodes.

The maximal difference of $v = 2 \cdot v_{max}$ is reached, in case both nodes move in opposite directions with the maximal velocity v_{max}. The figure shows that about 80% of the values lie in the interval between $[0.3 \cdot v_{max}, 1.3 \cdot v_{max}]$. The mean E is calculated as

$$E(v) = \frac{1}{\pi} \iiint_{v_1, v_2, \alpha} \left| \vec{v}_1 - \vec{v}_2 \right| dv_1 \, dv_2 \, d\alpha = \int_{v=0}^{v_{max}} v \cdot f_v(v) = 0.725 \cdot v_{max} \tag{19}$$

with $\left| \vec{v}_1 - \vec{v}_2 \right|$ from (14). Solving the integral reveals $E(v) = 0.725$. This perfectly matches the results illustrated in Figure 23.

As described in [139], the LLT for a single hop path is the time, when the distance between two nodes, initially within each other's proximity, becomes greater then R and consequently, both nodes are unable to communicate directly with each other anymore. Figure 24 shows this correlation.

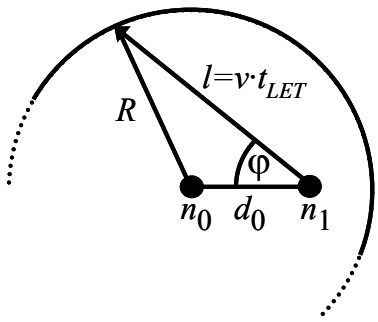

Figure 24: Calculation of t_{LLT} with known mobility parameters d_0, φ, and v.

The LLT time t_{LLT} between any two ad hoc nodes obviously depends on their initial distance d_0 and the relative velocity vector $\vec{v}$, with an absolute value v and a movement angle φ. With the law of cosine, the LLT depends on

$$R^2 = d_0^2 + \left(v \cdot t_{LLT} \right)^2 - 2 \cdot d_0 \cdot v \cdot t_{LLT} \cdot \cos(\varphi) \tag{20}$$

Solving (20) to t_{LLT}, the general description for the distribution of the LLTs $f_1(t)$ between a node assumed stationary and one node with the relative absolute velocity v is

$$f_1(t) = \iiint_{d_0, v, \varphi} \delta(t - t_{LLT}) \cdot f_\varphi(\varphi) \cdot f_{d_0}(d_0) \cdot f_v(v) \; \partial d_0 \, \partial v \, \partial \varphi \tag{21}$$

with $f_\varphi(\varphi) = \frac{1}{\pi}$, $f_{d_0}(d_0) = 2\frac{d_o}{R^2}$ and $f_v(v)$ from (17).

Obviously, the integral cannot be described in closed form. In order to calculate this, another approach is necessary. The author in [136] depicts that two random variables and their respective probably distributions depend on each other in a way that

$$f_x(x)dx = f_y(y)dy \;\Leftrightarrow\; f_x(x) = f_y(y)\left|\frac{\partial y}{\partial x}\right| \tag{22}$$

Therewith, the probability distribution $f_1(t)$ is

$$f_1(t) = \frac{2}{R^2 \cdot \pi} \cdot \iint_{v, d_0} \left|\frac{\partial \varphi}{\partial t}\right| \cdot d_0 \cdot f_v(v) \; \partial d_0 \; \partial v \tag{23}$$

with φ from (20) as

$$\varphi = \cos^{-1}\left(\frac{d_0^2 + (v \cdot t)^2 - R^2}{2 \cdot v \cdot t \cdot d_0}\right) \tag{24}$$

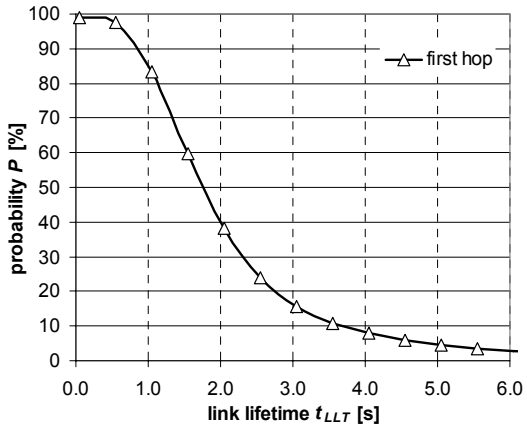

Figure 25: LLT-distribution function showing t_{LLT} for the first hop.

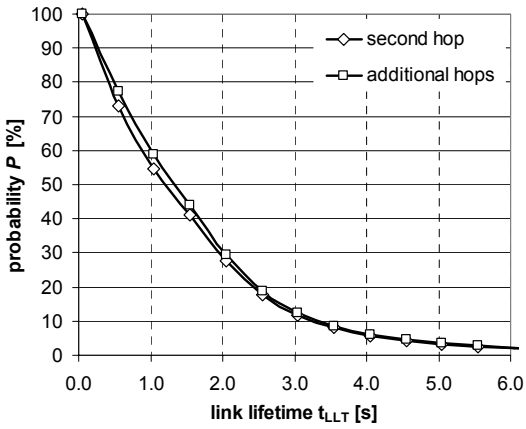

Figure 26: LLT-distribution function showing t_{LLT} for the second and any following hop.

The differentiation of φ as well as the integration of d_0 is straightforward, while the outer integration with v requires a differentiation for t greater and smaller than $2 \cdot R/v_{max}$. Due to the complexity of $f_v(v)$, the complete formula is not presented. Figure 25 shows the corresponding LLT distribution function for the first hop. The figure depicts the graph for $R = 1$ m and $v_{max} = 1$ m/s. It can be observed that the probability rapidly decreases with increasing times t_{LLT}. The proof in [11] shows that the LLT times are proportional

to R and inverse proportional to v_{max}. Therefore, the LLT probability for varying basic network parameters can be expressed as

$$f_1'(t) = f_1\left(t \cdot \frac{R}{1m} \cdot \frac{1m/s}{v_{max}}\right) \tag{25}$$

3.4.2 Distribution of link lifetimes for subsequent hops

The flooding mechanism of reactive routing algorithms usually discovers the shortest path from source to destination. Any node receiving a request for the second time discards this message. Whereas, a node receiving the flooding message for the first time is certainly not a neighbor of nodes, which have previously forwarded the request. Otherwise, it already had received the message from one of these nodes. This behavior is system immanent when assuming an error free radio transmission. Unlike the description in [16], where LLT times of arbitrary hops are equivalent, the LLT times depend on the link position within the path. The reason for this necessary differentiation is the varying position probability of link neighbors. Positions of downstream nodes within the path depend on the positions of previous nodes. The position of the first downstream neighbor after the source is uniformly distributed within the coverage of the transmitting node. In contrast to that, the positions of subsequent nodes n_{i+1} (with $i > 0$) are limited. A node, receiving the message for the first time and not directly from the source, can only be located inside an area, covered by the radio range of its predecessor n_i, minus the area covered by the transmission disks of any upstream node n_j (with $j \leq$ i-1) which already forwarded the request message. Figure 27 clarifies the relation between the positions of neighboring nodes. Consequently, the distance relationship between nodes participating in a route discovery can be expressed as

$$\left|\overrightarrow{n_{i+1}}\text{-}\overrightarrow{n_i}\right| \leq R \;\cap\; \left|\overrightarrow{n_{i+1}}\text{-}\overrightarrow{n_j}\right| > R \;\forall\; j \leq i-1 \tag{26}$$

with $\overrightarrow{n_x}$ being the position of node n_x, and i, j the node numbers relative to the source. Therefore, flooding based route discoveries cause different behaviors of link lifetimes for the first hop and any subsequent hop.

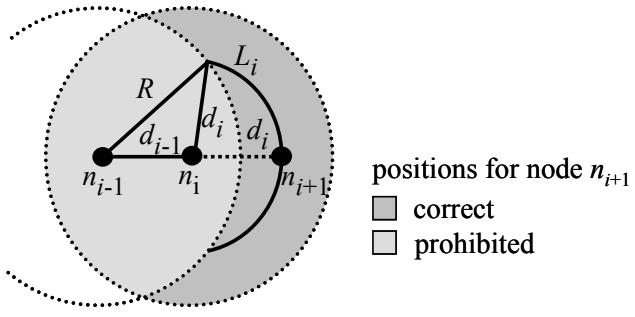

Figure 27: Possible area for a next nodes location, depending on the positions of previous nodes.

Additionally, the node position probability varies slightly with the overall network node density. Due to the numerous node selection possibilities in dense networks, the routing algorithm creates paths with nodes having maximal distances between each other. Therefore, in highly dense networks, paths are shorter, and the probabilities of positions

for the following relay node are usually not uniformly distributed. As the effect occurs only for high node densities, it is neglectable.

The flooding algorithm prevents the selection of nodes violating the second constraint in (26). After the route discovery between source and destination, intermediate nodes are free to move into areas previously prohibited. The path breaks, whenever the distance between adjacent nodes exceeds the maximal transmission range R. In order to allow statements concerning the expiration times of links between intermediate node n_i and n_{i+1}, the probability distribution describing the distances between them is necessary.

Figure 27 depicts equidistant circles around node n_i. The lengths L_i of these circle segments determine the distance probability between node n_i and n_{i+1}. The length depends on the distance d_i and on the distance $d_{i\text{-}1}$ between the nodes n_i and $n_{i\text{-}1}$. With the law of cosine, the radial length L_i of a circle segment is calculated as

$$L_i = 2 \cdot d_i \cdot \cos^{-1}\left(\frac{R^2 - d_{i-1}^2 - d_i^2}{2 \cdot d_{i-1} \cdot d_i} \right) \tag{27}$$

The probability function for the distance between any two nodes can be calculated as integral over the distance L_i together with a probability factor $f_{d_{i-1}}(d_{i-1})$. This factor determines the probability that node n_i has a certain distance $d_{i\text{-}1}$ to its own predecessor. Therefore, the integral for the distance probability function is defined as

$$f_{d_i}(d_i) = \int_{d_{i-1}=R-d_i}^{R} L_i \cdot f_{d_{i-1}}(d_{i-1}) \; \partial d_{i-1} \tag{28}$$

Equation (28) is recursive and therefore certainly not memory free. The distance distribution between every two nodes depends on the distance distributions of all preceding nodes. Since the second node within a path can utilize the full transmission disk of the fist node, the pdf function for the distance between the first two nodes is again the distribution $f_{d0}(d_0)$.

Furthermore, with the help of this initial node distance distribution, any following distance distribution can be calculated. The evaluation of (28) again requires the Taylor series transformation of the $cos^{-1}(x)$ –term given in (18). The computation shows that only the pdf-functions for d_0 and d_1 significantly differ from following distance distributions d_i. For $i \geq 2$ the function reaches its steady state, which is shown in Figure 28 as well.

Hence, following the calculation in (21), the LLT function for a further hop within an already existing path from source to the predecessor node can be determined as

$$g(t,i) = \frac{1}{\pi} \cdot \iiint_{v,d_i,d_{i-1}} \left| \frac{\partial \varphi}{\partial t} \right| \cdot L_i \, f_{d_{i-1}}(d_{i-1}) \cdot f_v(v) \; \partial d_{i-1} \, \partial d_i \; \partial v \quad with \; i > 1 \tag{29}$$

Due to complexity reasons, the evaluation does not take into account the possibility, that other nodes than the one before the direct neighbor covers the transmission disk of the current node. In other words, it does not consider any node position limitations caused by nodes more then two hops away, as requested in (26).

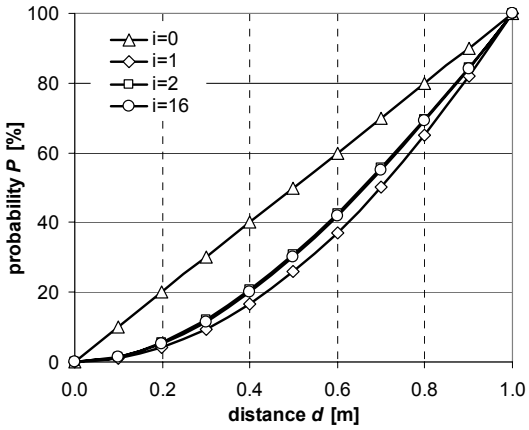

Figure 28: Distance distribution between any two nodes for different hop values *i*.

The graphs in Figure 26 indicate that the LLT distribution for the second hop $g(t,2)$ is similar to distributions $g(t,i)$ for links with more than two upstream nodes ($i > 2$). An additional analysis shows that the induced error, caused by using only the distribution $g(t,3)$ for any hop but the first one, is about 0.7% for a two hop scenario, and decreases with increasing hop numbers. The induced error is already neglectable for paths with three or more hops. Therefore, it is unnecessary to distinguish between the second and any further hop. Using the distribution of the third link $g(t,3)$, the special function $g(t,i)$ simplifies to a general $g(t)$ for all possible values of i.

3.4.3 Distribution of path lifetimes for arbitrary hop numbers

After the evaluation of the LLT for a first hop and for any subsequent hop, the following theoretical analysis evaluates the distribution of path lifetimes with arbitrary hop lengths. As described in [12], the PLT for a path with h hops and h+1 participants is the minimum of h LLTs of all individual links, and can be described as

$$t_{PLT} = min(t_1, t_2, t_3, \ldots, t_h) \tag{30}$$

with t_i being the LLT of the i'th hop. As all link breaks are independent of each other, the distribution for a path with h hops $f_h(t)$ is sum of all conditional probabilities

$$P = \frac{1}{h}\sum_{i=1}^{h} P\left(link\ i\ breaks \,\middle|\, all\ other\ links\ exist\right) \tag{31}$$

The distribution function $f_h(t)$ depicts the respective distributions of PLT. Therewith, the pdf-function for any path with h hops is created when utilizing

$$f_h(t) = 1 - \frac{\left(1-f_1(t)\right)\cdot\left(1-G(t)\right)^{h-1} + \left(h-1\right)\cdot\left(1-F_1(t)\right)\cdot\left(1-G(t)\right)^{h-2}\cdot\left(1-g(t)\right)}{h} \forall\ h>1 \tag{32}$$

with $F_1(t)$ and $G(t)$ being the complementary cdf-functions of the corresponding LLT distributions $f_1(t)$ and $g(t)$.

As described in section 3.4.2, the PLT-function *g(t)* is used for all links but the first link. Figure 29 shows that the probability of large PLTs for paths with more than two hops decreases rapidly. The probability of maintaining a single hop path for more than 2 seconds (with the conditions of R = 1 m and v_{max} = 1 m/s) is around 25%. However, the

probability to reach an equivalent PLT with two hops already decreased to 6 %, whereas it drops to only 0.4% with 4 hops.

3.4.4 Validation of the theoretical assumptions

This section compares results from the previous theoretical analysis with simulations to verify the accuracy of the made assumptions and to show that the introduced simplifications are valid.

The source node creates a path to a destination over h hops. Every downstream node n_i ($0 < i < h+1$) is uniformly and randomly placed within node n_{i-1}'s coverage area, but outside of any other upstream node. This consecutively forms a random path from source to destination and ensures that the simulation only considers paths, which are initially utilizable. This procedure is comparable to a flooding request within real networks. A real node forwards a request by sending it to any node within its transmission range, while the simulation creates a new node to receive this request. The simulation uses the same mobility model and parameters as described in section 3.4.1. They have sufficient runtimes and the evaluation includes 10^6 independent simulation runs.

Figure 29 shows the differences between selected graphs from the theoretical analysis and the simulations. Solid lines depict the theoretical analysis, whereas the dotted graphs show the corresponding simulative results. The largest difference between simulation and theoretical analysis occurs for the two-hop path. This arises from the simplification made for the velocity distribution $f_v(v)$ in section 3.4.1. While this simplification perfectly matches for a single hop path, it does not hold for two or more hops. The velocity of a relaying node is certainly equal for the incoming as well as for the outgoing link. However, the simplifications do not consider this dependency, and consequently assume both hops as independent. However, Figure 29 illustrates that the impact lessens with increasing numbers of hops.

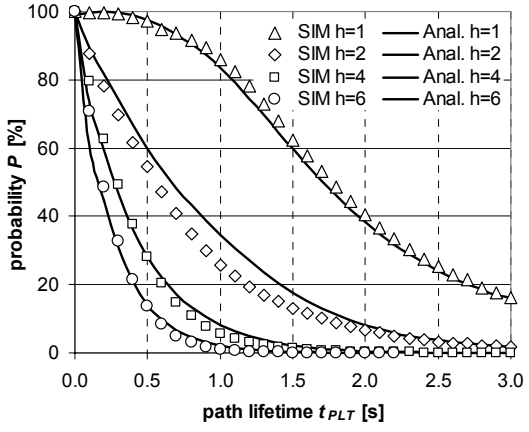

Figure 29: Comparison between PLT distributions for various hops h. Theoretical results are depicted as lines while points indicate simulation results.

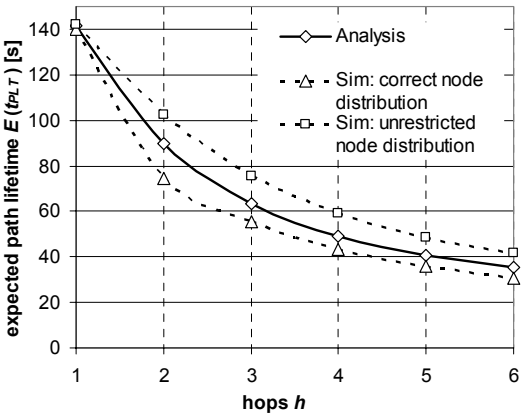

Figure 30: Expected PLT times with a transmission range R = 100, comparison between correct and unrestricted node positioning and theoretical analysis.

As described in section 3.4.2, another simplification contains a small additional error. It only considers the restricting node positioning effect caused by the node before the

direct neighbor n_{i-1}, and not caused by all previous nodes. This error is not neglectable, although it happens seldom. However in the rare cases of its occurrence, it considerably reduces the area for the next node's position. As Figure 29 presents, its disregard slightly falsifies the path lifetime distributions.

For every additional hop the PLT distribution changes and significantly reduces the average PLT. Unpredictable node movements within mobile ad hoc networks prevent forecasts about path lifetimes of particular paths. The large degree of freedom of node movements prevents forecasts of individual PLTs. However, the statistical method gives an estimate for PLT in case external informations are unavailable. Especially longer paths have unfavorable PLT distributions. The possibility is greater than 50% that particular paths do not remain valid as long as the average lifetime

Figure 30 shows the average PLT times $E(t_{PLT})$ for various path lengths h. The figure contains the theoretical evaluation, simulations with the correct as well as simulations with unrestricted node positioning. The expected mean path lifetime $E(t_{PLT})$ is in accordance to

$$E(x) = -\int x \cdot \frac{\partial f(x)}{\partial x} \, \partial x \tag{33}$$

with $f(x)$ the complementary cdf function of the PLT distribution. In contrast to previous evaluations, the radio transmission range is set to $R = 100$ m in order to create realistic network conditions. The maximal node velocity is kept constant with $v_{max} = 1$ m/s.

Despite the theoretical analysis overestimates the probabilities for a certain PLT and hop number, it gives a good estimate for $E(t_{PLT})$. The introduced error increases with increasing hop numbers. With two hops, the divergence is only 3% while it overestimated the simulative result by 15% for six hops. However, with six hops, the average PLT time already dropped below 40 seconds, with an unfavorable PLT distribution. The probability that those paths break within the first 20 seconds is almost 40%. As a conclusion it can be stated that routes with six or more hops are ineffective. The average PLT is too short, and reasonable communication is impossible. Frequent route requests or reestablishments would cause network congestion and high packet loss rates.

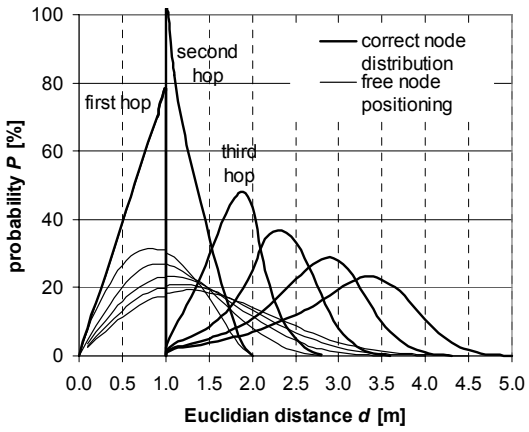

Figure 31: Euclidian distance between source and destination for *h* hops.

Completing this section, Figure 31 presents simulation results for the Euclidian distance distribution between source and destination for a certain path length. It again assumes a radio transmission range of $R = 1$ m. It illustrates the distributions for the first nine hops and compares two different node distributions approaches. The first distribution set considers that node positions depend on the position of previous nodes. The second function set presents the behavior of the Euclidian distance for an unrestricted node positioning.

As described in section 3.4.2, the distance distribution for single hop paths is equal in both cases. Yet it differs for all other path lengths. Assuming an accurate node distribution, the Euclidian distance between source and destination is always greater than R for paths with more than one hop. At route setup, all relaying nodes, but the first one, must be beyond the coverage of the source. When considering an accurate node positioning, longer path lengths create larger Euclidian distances. In contrast to that, results derived from simulations with an independent node distribution, illustrate that positions of destination nodes remain within closer proximity to the source node's position.

3.4.5 Multipath Connections

The knowledge about probability distributions of certain path lengths allows the calculation of disjoint multipaths distributions as well. As described in section 3.1.1, disjoint multipath ad hoc routes do not have any links in common. Therefore, routes fail independently and the probability functions of the path lifetime for individual paths do not correlate. The combined PLT time is the maximal PLT of all m paths, and therewith

$$t_{PLT} = max(t_1, t_2, t_3, \ldots, t_m) \tag{34}$$

The function $f_{h1\ldots hm}(t)$ describes the PLT distribution for a multipath system with m routes consisting of $h_1, \ldots, h_m$ individual path lengths. With $f_{hi}(t)$ the PLT function for a path with h_i hops and $F_{hi}(t)$ the respective complementary cdf function, $f_{h1\ldots hm}(t)$ is calculated as

$$f_{h_1,\ldots,h_m}(t) = 1 - \frac{\left(1 - f_{h_1}(t)\right) \cdot F_{h_2}(t) \cdot \ldots \cdot F_{h_m}(t) + \ldots + F_{h_1}(t) \cdot \ldots \cdot F_{h_{m-1}}(t) \cdot \left(1 - f_{h_m}(t)\right)}{m} \tag{35}$$

Assuming that all m routes have equal path length h, the calculation in (35) can be simplified to

$$f_{m\times h}(t) = 1 - \frac{1}{m}\left(1 - f_h(t)\right) \cdot F_h(t)^{m-1} \tag{36}$$

Figure 32 contains PLT distributions for up to five disjoint paths with a path length $h = 2$, while Figure 33 depicts the respective behavior for $h = 4$. Both figures use the parameter set with $R = 100$ m and $v_{max} = 1$ m/s. The pictures depict the possible improvement of PLTs for a certain probability $P = 50\%$.

Obviously, the achievable additional lifetime improvement decreases for increasing numbers of paths within a set. The possibility that an additional path remains valid longer than all other previous paths becomes more unlikely for larger numbers of paths.

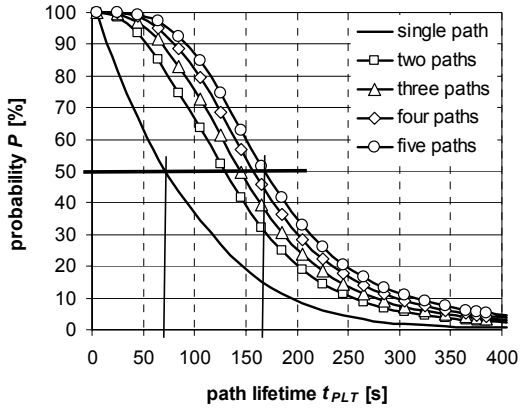

Figure 32: Distribution of lifetimes for a multipath system with *m* paths, utilizing two hops.

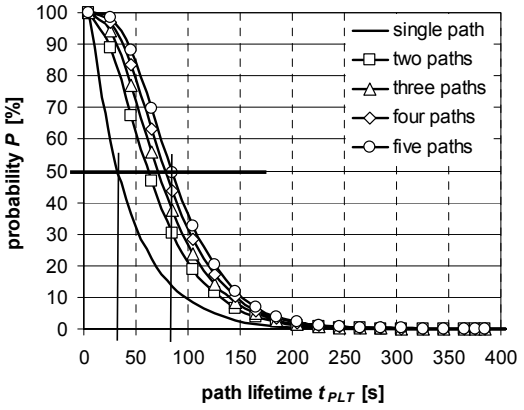

Figure 33: Distribution of lifetimes for a multipath system with *m* paths, utilizing four hops.

An evaluation of the expected PLTs for different route lengths and numbers of paths reveals that the achievable relative improvement is maximal for a single backup route. Within a two hop multipath system, it achieves about 45% longer combined PLTs, while it permits almost 60% for a four hop system. The variation of relative improvements for varying route lengths is almost insignificant Subsequent paths do not achieve comparable improvements. The relative improvement is about 10% for the third path, while the fourth and fifth path allow for only 6% and 4% superior results respectively. The variations between different path lengths are insignificant. Figure 34 shows the behavior of the average PLT for path sets with different route lengths and an assumed radio range $R = 100$ m. Instead of enhancing the overall achievable average PLT with additional backup paths, they could as well improve the possibility that at least one path within a set is still utilizable after a certain period of time. Therewith, backup routes improve networking with respect to certain connection availability requirements. The network layer is able to ensure minimum lifetimes for a set of routes with a certain probability.

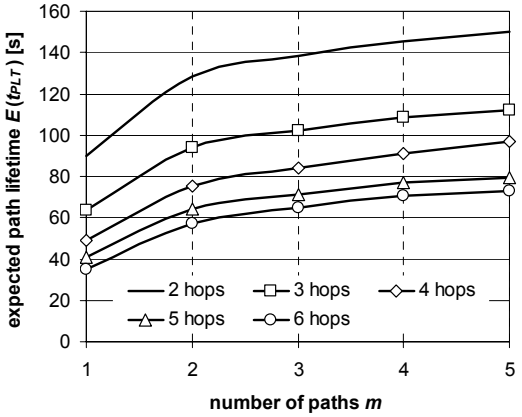

Figure 34: Behavior of average PLT for path sets with different individual route lengths.

By definition, backup routes are at least as long as their primary paths. However, additional backup routes are often longer than any previous route, and therewith their individual probability distributions $f(t)$ are more disadvantageous. Consequently, the achievable lifetime improvement is not as good as shown in Figure 32 and Figure 33. Both cases imply optimal conditions, where backup routes have the same length as the primary path. Any additional path improves the PLT, however the necessary effort to discover additional backup routes rises as well. Therefore, the optimal number of routes is certainly limited. Section 3.5.3 discusses the optimal number of backup routes while section 3.6 focuses on the respective route diversity overhead.

3.5 Simulative evaluation of multi-hop and multi-path scenarios

After the theoretical analysis of link and path lifetimes for SP and disjoint multipath, the following section focuses on the performance of the remaining non-disjoint routing strategies. However, theoretical analysis attempts are not considered in this work, and therefore simulations have been carried out as presented in [19] to evaluate the performance of these algorithms in comparison to the previous two strategies.

As introduced in section 3.1, the following evaluation considers four different types of path creation algorithms: the shortest path (SP), disjoint multipaths (DMP), the non-disjoint multipaths (NDM) and the flooding algorithm (FL). Figure 35 illustrates the different discovered paths.

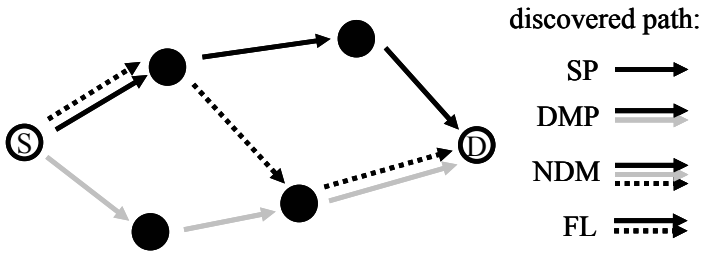

Figure 35: Route discoveries for single path, flooding extension, disjoint, and non-disjoint multipaths algorithms.

3.5.1 Modeled routing strategies

Section 3.1 introduces the behavior of routing strategies in real world environments. However, for an also realistic simulative approach, an efficient and accurate modeling of the routing strategies is necessary. Therefore, the following section gives a short introduction, how the examined strategies are included:

- *Shortest Path (SP):* The simulator has an overall network view, and therefore it utilizes the Dijkstra [140] algorithm to discover the shortest path between source and destination. The Dijkstra shortest path is equal to the path discovered with the regular ad hoc flooding algorithm, but requires less computation effort than the simulation of an entire flooding mechanism.
- *Flooding (FL):* To generate multipath sets, equal to the sets discovered by the flooding strategy, the Dijkstra algorithm is unsuitable. It is unable to model the

characteristic path trajectories obtained with the RREQ flooding mechanism. In order to generate suitable multipaths sets, a detailed implementation of the flooding mechanism is necessary. The included extension follows exactly the AODV RREQ flooding as proposed in [74]. It uses the same mechanisms and maintains the same routing tables. Obviously, this requires much more computational effort than the Dijkstra algorithm, but guarantees exact results.

- *Disjoint multipath (DMP)*: In contrast to the implementation of the flooding mechanism, the Dijkstra algorithm together with a connectivity matrix is used as basis to model the DMP routing strategy. The virtual connectivity matrix contains all links between neighboring nodes. Due to the fact that the simulator has full knowledge about the topology, it is able to create the matrix. In order to generate multiple node disjoint paths, the DMP model executes the Dijkstra algorithm multiple times. The first call of the Dijkstra subroutine again discovers the SP path. After each execution, the DMP algorithm removes all utilized links and nodes from the previously created connectivity matrix. Therewith, the connectivity matrix forms a new virtual network of nodes and links not yet utilized by any previously discovered path. The following call of the Dijkstra procedure discovers the SP of the remaining virtual network. The DMP algorithm repeats these steps until it has discovered the designated number of routes or the source or the destination becomes disconnected within the virtually formed connectivity network.
- *Non-disjoint multipath (NDM):* The model for the NDM algorithm is a variation of the DMP strategy. With the first route discovery call, it again determines the SP. However, in contrast to the DMP algorithm, it does not remove all the utilized links from the connectivity matrix. The links remain within the matrix but the simulation program increases the weights of all utilized links. Therewith, subsequent Dijkstra calls inherently try to prevent these links due to their increased costs, while they are still able to reuse them. In case yet unutilized links are inappropriate, new routes reuse already utilized links again. Therewith, subsequent paths are not necessarily node and link disjoint to previous paths. The initial weights of all links are one and the additional virtual costs Δw_{NDM} for all utilized links is constant. In contrast to DMP, the source never becomes disconnected form the destination within the created virtual network. Therefore, NDM sets always contain the designated number of backup paths.

Obviously, DMP and NDM path sets outperform the SP with respect to path availabilities. The DMP and the NDM path sets always contain the SP. As described in section 3.1.1, additional backup paths always lengthen the path expiration time in comparison to the single path. Therefore, their combined path lifetimes are always superior to those of the SP.

3.5.2 Description of the PLT Simulator

The PLT-simulator developed at the Lehrstuhl für Kommunikationsnetze is based on the *L*ibrary of *E*fficient *D*ata Types and *A*lgorithms (LEDA) [141]. LEDA contains the necessary data types and procedures, e.g. the Dijkstra algorithm, to perform the required evaluations easily. Additionally it provides a graphical interface to visualize the discovered paths within arbitrary network topologies. The left picture in Figure 36 depicts the established paths with the flooding algorithm at startup time, whereas the right snapshot shows the discovered path obtained by the DMP strategy. The optimized

implementations of algorithms in LEDA allow numerous repetitions with varying starting parameters within reasonable time.

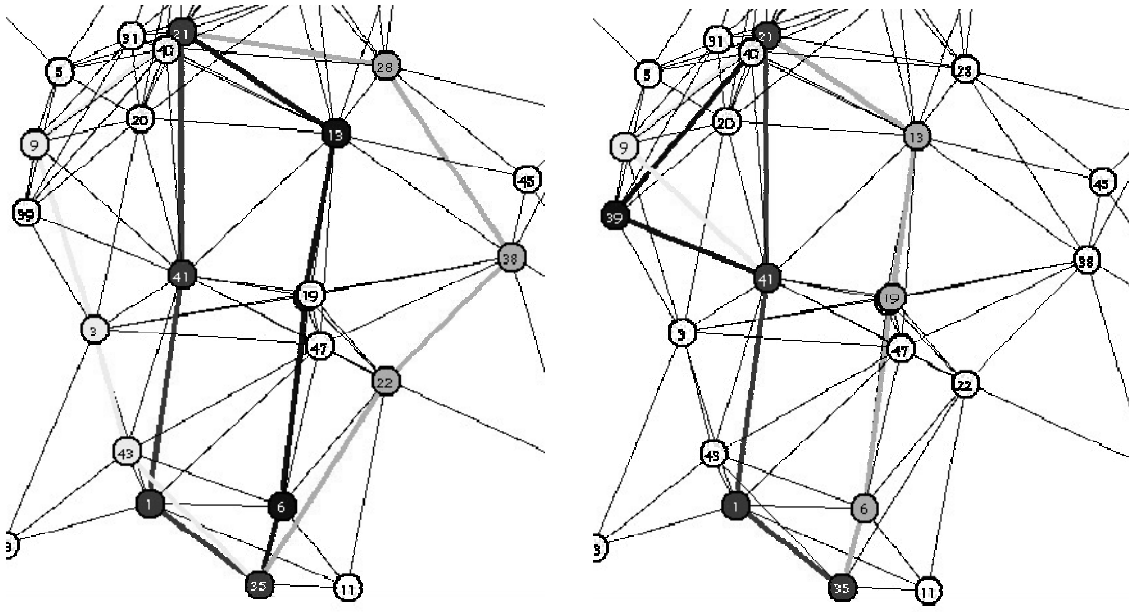

Figure 36: Snapshots of the PLT simulator.
Left: the DMP strategy. Right: the FL algorithm.

To compare the theoretical results with results obtained from simulations, the simulation environment again uses a simplified real world model in terms of the used mobility model. The variable parameters are the number of nodes N, the average number of neighbors n, the maximal node velocity v_{max}, and the radio transmission range R. As a simplification, the transmission area again forms a perfect disc around each node.

Neighbors are nodes within the maximal transmission range and therewith capable to setup direct node-to-node communication. The definition of the average number of neighbors n is shown in Figure 37. It is derived from the overall node density ρ within the network. The average number of neighbors n is equivalent to the node density ρ multiplied with the covered area of a radio transmission $\pi \cdot R^2$ and subtracted by one, to consider the transmitting node within the covered area. Therewith, the number of neighbors is

$$n = \rho \cdot \pi \cdot R^2 - 1 \tag{37}$$

Obviously, the square of the radio range R is proportional to n for large n ($R^2 \sim n$ *for* $n \gg 1$).

The overall node density ρ is defined as the fraction of the number of nodes N with the size of the simulation area a^2

$$\rho = \frac{N}{a^2} = \frac{n+1}{\pi R^2} \tag{38}$$

Assuming N and n as given, the edge length of the simulation area is constant and a is computed as

$$a = \sqrt{\frac{N}{\rho}} = R \cdot \sqrt{\pi \frac{N}{n+1}} \tag{39}$$

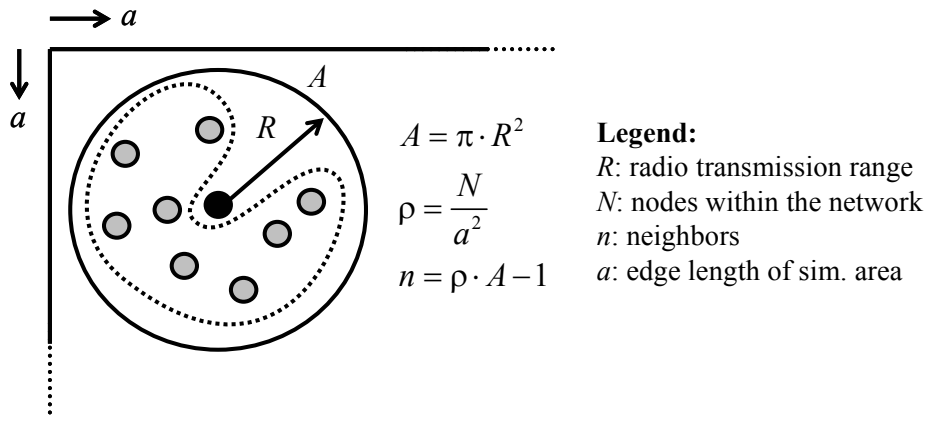

Figure 37: Definition of the neighbor node density *n*.

As node mobility model, the simulations utilize a simplified random direction (RD) mobility model [93]. A random process assigns all nodes an initial position within the simulation area, and an initial velocity vector. Node $i \in [0..N\text{-}1]$ moves towards direction $\theta_i \in [0..2\pi]$ with speed v_i. The velocity is uniformly distributed between 0 and v_{max}. Therewith, the average node velocity v_{av} is $v_{max}/2$. After start-up, all nodes immediately start moving towards their direction θ_i with their constant velocity v_i. The mobility model does not induce nodes to change their velocities. Additionally, nodes do not change directions or pause between consecutive movements. Only in case nodes approach the borders of the simulation area, the movement algorithm reverses the velocity vectors at the perpendicular. This guarantees that the overall node density is constant throughout the complete simulation.

The parameters n and v_{max} depend on R. The number of link changes (emerging links and disconnections) per time interval directly depends on the average time t_{av} a node requires to move through another node's radio coverage. With [142], the average distance L_{av} of two random points on the circumference of a circle with radius R is calculated as

$$L_{av} = 4 \cdot R / \pi \tag{40}$$

With $t_{av} = L_{av}/v_{av}$, and t_{av} set constant, R is proportional to v_{av} and therewith proportional to v_{max} as well. As already depicted, the radio range R^2 is proportional to n. Hence, the neighbor node density as well as the dynamic of a network is changeable without modifying the radio range. Consequently, the radio range as additional parameter is unnecessary and consequently is kept constant within all evaluations.

Although the proposed model is very simple, it does not significantly vary from the simulation environment of the standard ns-2 networks simulator. An ns-2 simulation contains a predefined number of nodes and all nodes have the same radio coverage. Commonly, this coverage is a perfect disc around each node. Therewith, link breaks only occur when the distance between two nodes exceeds the radio range.

Nodes move in accordance to the RD model. Therefore, nodes controlled by this model pause and turn after consecutive movements to obtain some variations in the network topology. However, on a short time scale, the differences to the above described

movement model are marginal. As the examined simulation periods are always short, the made simplifications are valid. While ns-2 allows in depth evaluations of protocols on the network layer or above, the possibility to evaluate the node connectivity is limited. Hence, the underlying connectivity models of both simulators do not significantly differ.

For each simulation run, the simulator creates a new arbitrary network. With the knowledge about the initial positions of all nodes and the radio transmission range R, the simulator calculates all available links between nodes. Hereafter, it randomly chooses a source and a destination node and calculates all routes between both endpoints. The characteristic and the number of calculated paths depends on the currently utilized routing strategy. The required route setup time is much smaller than the PLT ($t_{route\text{-}setup} \ll t_{PLT}$) and therefore the simulator neglects movements during route creation. The term PLT is used for SP as well as for multi paths simulations. Within a multi paths set, PLT depicts the failure time of the most durable path, which is not necessarily the last path within the multipath set.

As all calculations use exact node positions and all links are operational, the simulated routing scheme always generates the optimal available paths. In order to achieve realistic network conditions, the simulator is able to use error prone links. With the error threshold e_{link}, a random process decides about link errors, and the simulation removes defective links from the network. With a link error probability of 50%, statistically every other link shows errors. Therewith, the tested routing strategies do not always discover the shortest physically available path and consequently the average number of hops between source and destination increases.

As described, the edge length a of the simulation environment depends on the radio range, the average number of neighbors, and the number of nodes. Therewith, the initial simulation area varies for different simulation setups. The common parameter set defines a network with $N = 100$ nodes, assumes a radio range of $R = 100$ m, and a node density n of ten neighbors on average. Further on, the maximal node velocity v is 2 m/s and the network does not contain defective links ($e_{link} = 0$). The multipath algorithms try to create up to five paths and the NDM paths calculation uses a link-weight increment Δw_{NDM} of 0.5 for each utilized path. If not otherwise stated, all following simulations use this common parameter set.

The evaluation of a certain parameter set contains 10000 independent simulations in order to allow statistically significant results. Additionally, the 95% confidence intervals are computed. However, due to the large number of simulations, the sizes of the confidence intervals are neglectable, and therefore the following figures do not contain the intervals.

3.5.3 Evaluations

The first simulation compares the path calculation algorithms with respect to different network sizes. Figure 38 shows the behavior of PLTs for networks with 50 and 500 nodes. The average path length increases with increasing network sizes, as source and destinations are uniformly distributed within the simulation area. Certainly, the PLT depends on the path length. If all other simulation parameters are constant, smaller networks have more favorable PLT distributions than large networks. It is also evident that connections with multiple backup routes allow longer lifetimes in comparison to the single path. Obviously, the simulated PLT distributions of the SP and the DMP sets

depict behaviors comparable to those theoretically determined. Only the scaling varies due to different assumed radio ranges and maximal node velocities.

With 50 nodes within the simulation, 30% of all SPs are still valid after 20 seconds and at least 50% of the multipath connections. This turns worse for 500 nodes networks. The probability of an unbroken SP after the same time (20s) is only 10% and 20% for multipaths systems. For all network sizes, the NDM route selection outperforms the DMP and the FL selection. For small networks, the FL algorithm significantly outperforms the SP, although it never reaches the performance of both other multipath algorithms. Within large networks, the improvement in comparison to the SP algorithm is not as considerable. As described, FL backup paths often utilize the same first couple of links as the primary path. Therefore, link breaks close to the source often destroy complete path sets. The relative number of disjoint links within backup paths is significantly higher within networks consisting of 50 nodes than within a 500 nodes network.

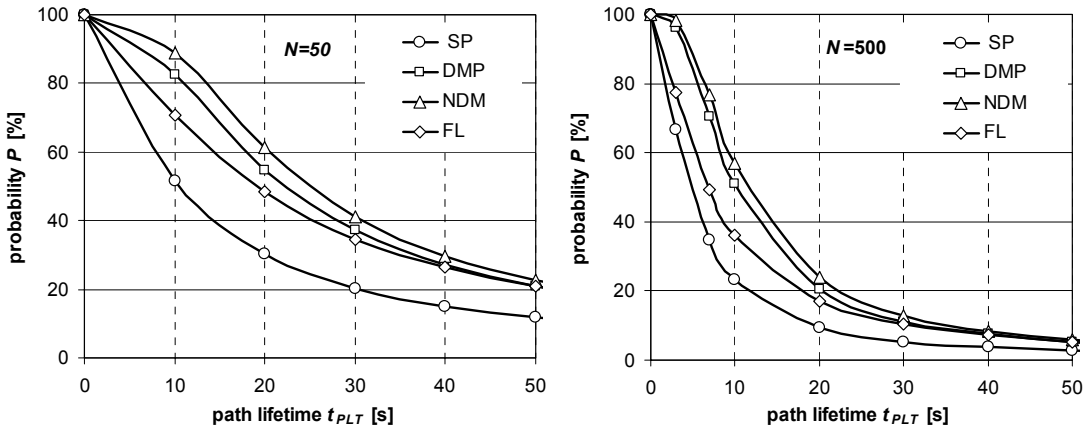

Figure 38: Path lifetime distributions for 50 and 500 nodes networks and for different routing strategies.

In order to emphasize this correlation, Figure 39 shows the average path lifetimes for varying network sizes. As described, the node density is ten neighbors on average and remains constant. Within 50 node networks, NDM increases the average connection lifetime by more than 50% compared to SP. With 500 nodes, NDM allows 80% longer lifetimes than SP. The improvement of NDM to DMP is always about 8% and independent from the network size. Within small networks, the FL algorithm achieves an equivalent average PLT as the DMP strategy. However, it fails to achieve comparable results for larger networks. Consequently, multipaths algorithms show good results in large networks, in which numerous backup routes are available. However in small networks the simple flooding protocol already achieves sufficient PLT improvements with significantly reduced routing efforts.

The higher flexibility of the NDM routing strategy allows better performance than the DMP algorithm, although routes have links in common and multiple paths can break simultaneously. The degrading factor of common links within subsequent paths is smaller than the advantage of greater flexibility during route discovery.

With constant node densities, the diameter of networks increase proportional to $\sqrt{N}$. The average length of the shortest path rise with this factor and consequently the average

path length of all three multipath strategies rise equivalently. The average path length of a multipath set is the arithmetic mean of all individual path lengths at startup. Figure 39 shows the average path length for different network sizes and for all four routing strategies.

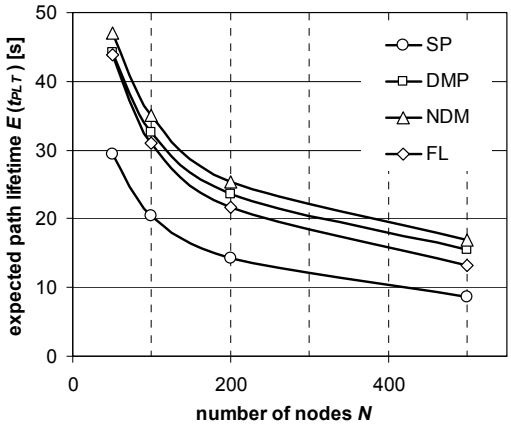

Figure 39: Average path lifetimes as function of simulation sizes.

Obviously, SP routes have the shortest paths. The average length of DMP paths is about 15% greater than NDM or FL routes, whereas NDM and FL require only about 10% more hops than the SP. The differences are independent from the network size. The greater flexibility of NDM route discovery allows shorter average paths than the DMP algorithm. The DMP algorithm mandatory requires disjoint paths and therefore DMP backup routes in general are longer than NDM routes. Subsequent paths are significantly longer, because the route discovery must circumvent all previously utilized links. As shown in section 3.4, longer routes statistically break earlier than routes with fewer hops. Therefore, the combination of greater flexibility during path selection together with short backup routes allows NDM superior performance.

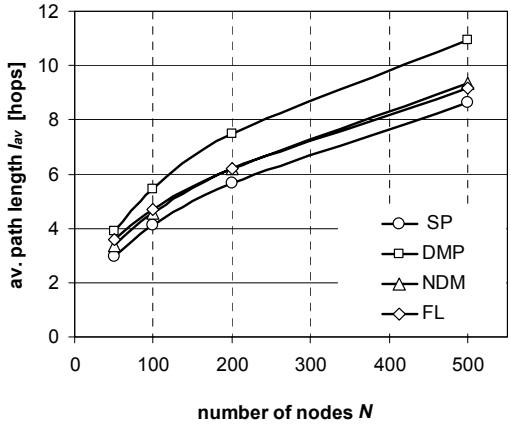

Figure 40: Correlation between network size and average path length.

Figure 41 shows the behavior of the average path length with respect to PLTs. Within this case, the average path length is the average of the individual path lengths of all still functional paths at time t_{PLT}. The initial lengths of paths in 500 nodes networks are about three times larger than routes in 50 nodes networks. After 40 seconds, the average path

length of unbroken routes is almost equal in both networks. Connections with five or less participating nodes contain unbroken routes, whereas routes with more hops rapidly show errors within all discovered paths. The network acts as “low pass filter” for path lengths.

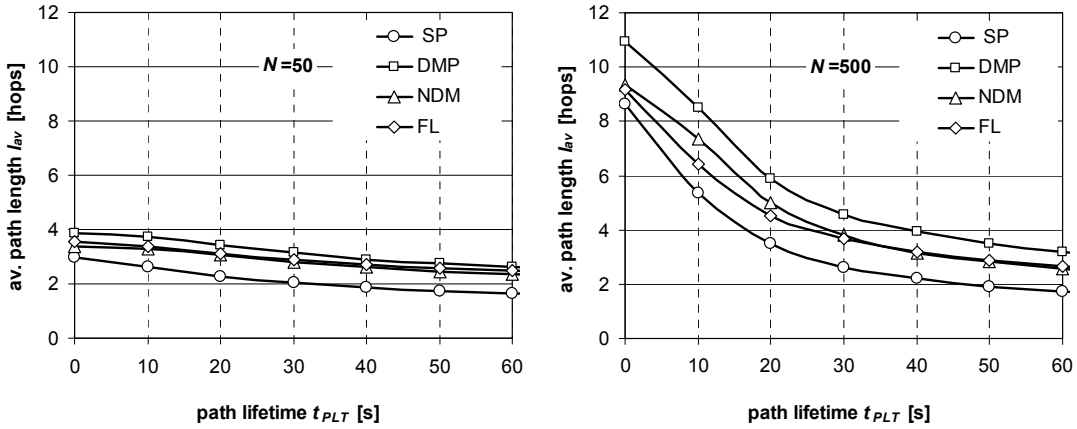

Figure 41: Average path lengths as a function of path lifetimes for 50 and 500 nodes networks.

As depicted, connections with more than five hops break within the first 20 seconds (for this parameter set). The connection between sources and destinations requiring routes with many hops significantly downgrades the overall network performance. The higher number of connection reestablishments congests the network and constrains even connections with fewer hops. A possible solution can be the usage of a time-to-live (TTL) function within route request packets would limit the maximal route length. It excludes connections requiring longer routes. As a consequence, it avoids these network performance degrading effects.

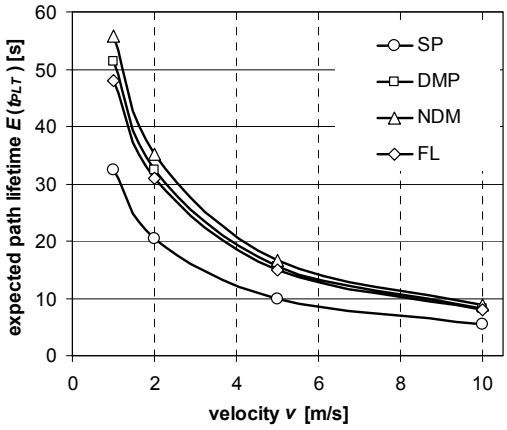

Figure 42: Average path lifetimes with respect to maximal node velocities.

As depicted by Figure 42, the expected PLT decreases with increasing maximal node velocities. The necessary movement period for a certain distance decreases inversely proportional to v, and therefore the average PLT decreases with the same relation. The probability that backup routes still exist, in case primary routes show permanent errors is higher in low mobility networks than within highly dynamic environments. Therefore,

the performance benefit of backup routes is greater under low mobility conditions. In this case, the usage of the NDM route discovery improves the path lifetime by 80%. Within highly dynamic environments, the improvement drops to 50%. The benefit of NDM in comparison to DMP is about 8%, again more favoring low mobility scenarios. Additionally, the FL algorithm performs almost as good as the DMP strategy.

Figure 43 clarifies the benefit of multiple backup routes. With up to two backup routes (three paths) and a node density of ten neighbors on average, the DMP algorithm is as good as the NDM strategy. The requirement of disjoint links does not avoid the discovery of short, favorable paths. With lower node densities or more routes, NDM again shows better results than DMP. In these cases, the DMP algorithm is unable to create routes with short path lengths or even does not find enough disjoint backup routes at all. The FL algorithm always performs about 5-10% worse than DMP. Even so it does not use any kind of multipath routing strategy the average PLT improvements are significant.

With a node density of ten neighbors, the DMP algorithm is able to calculate only four disjoint paths on average. Therefore, the expected PLT does not increase further with more than four requested paths within a multipath set. With a reduced node density of only five neighbors, DMP just discovers two routes. Only with 20 or more neighbors, it is able to calculate five disjoint paths. The NDM algorithm is more flexible within all kinds of network environments. It allows the creation of numerous backup routes. Figure 43 depicts that all backup routes improve the average PLTs of NDM, although the improvement rate decreases with an increasing number of backup routes.

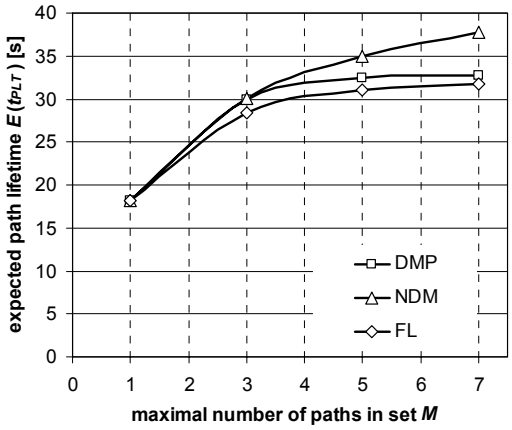

Figure 43: Average path lifetimes as a function of number of simultaneous paths.

Figure 44 shows the PLTs as well as the average path length with respect to the node density. As described, with only few neighbors, the path setup is difficult. Even the NDM algorithm is unable to create the necessary number of different non-disjoint backup routes. With three neighbors, the mean PLT is higher than within networks with five neighbors. Although this reduction seems unexpected, the sparse node distribution limits the path length. The network environment only allows the discovery of short routes with only few hops. Networks are often disconnected, and therefore favor connections with few hops. Consequently, the majority of found routes is short and quite durable. The short average path length emphasizes this behavior. The average path length is lower with three neighbors than with five neighbors. However, the average paths lengths would decrease with further decreasing node densities. Within very sparse

networks, the SP is the most useful path calculation algorithm, as other strategies are unable to setup valuable backup paths.

With five neighbors, the network discovers longer routes and consequently the overall average PLT decreases. The average path length is about seven hops for all three multipaths routing algorithms. Higher node densities lead to a probability of almost zero that disconnected networks occur and the average path lengths decreases rapidly.

Between five and ten neighbors, the achievable improvement of NDM in comparison to the DMP algorithm is most significant. With respect to the belonging SP, NDM increases the average path length only by about 20%, whereas it improves the mean PLT by at least 50%. Further on, the NDM algorithm improves the PLT by 15% in comparison to both other multipath protocols. This additional achievable PLT improvement is maximized for networks with five neighbors, whereas ten neighbors enables the algorithm to setup numerous additional paths with almost optimal path lengths.

With further increasing node densities, networks allow numerous and short disjoint multi paths and the benefit of non-disjoint paths are neglectable. Therefore, with more than 20 neighbors on average, DMP path calculation is equivalent to the NDM algorithm with respect to the average PLT. As the simulations suggest, the DMP is the optimal path calculation algorithm in crowded networks.

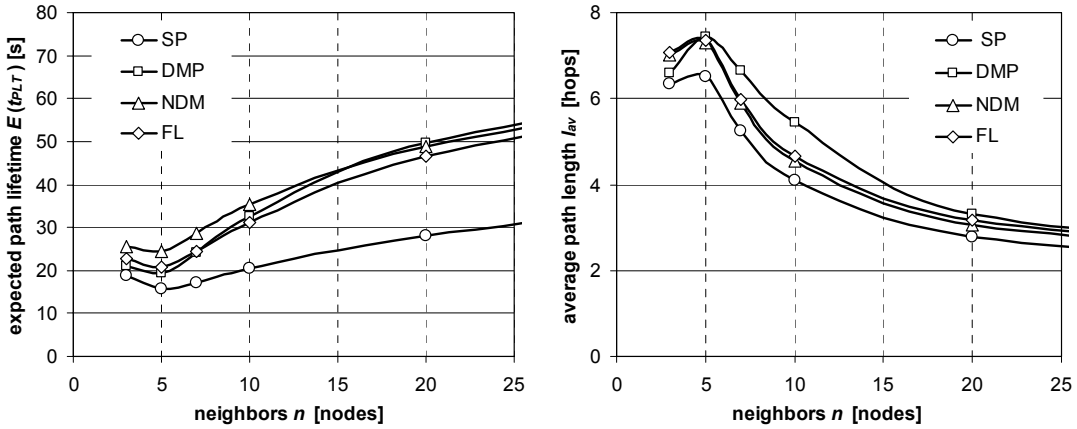

Figure 44: Expected PLT and average path length as a function of node densities.

The expected PLTs for varying link error probabilities are presented in Figure 45. For low link error probabilities (up to 20%), the PLTs do not degrade. Only with even higher error rates, the PLTs of multipath algorithms shorten slightly, whereas the SP route discoveries are not affected from these errors. The DMP lifetimes degrade significantly but only by about 17% with a 50% link error probability. While NDM outperforms DMP by 9% for error free scenarios, it even shows 14% improved PLTs within highly error prone scenarios. With fewer available links, the more flexible NDM algorithm is able to create a sufficient number of backup paths with reasonable hop lengths. DMP must use long path or even does not detect enough backup routes and therewith its PLTs degrade faster with increasing link error probabilities. The flooding algorithm is able to cope with all kinds of error probabilities and its average PLT is almost constant.

With increasing link error probabilities, the path lengths of NDM increase faster than the lengths of DMP paths. Without errors, NDM connections have 20% shorter paths than

DMP, whereas this advantage drops to only 8% with every other link broken. This is unexpected, as the inflexible DMP algorithm should calculate much longer backup routes within error prone scenarios than the NDM algorithm. The DMP algorithm often does not even find a single backup route. This in turn is the reason for the slow rising path length average. The average path length of DMP with only a single path is always the optimal/shortest path. However, this does not benefit the PLT. In contrast to DMP, the NDM algorithm finds multiple backup routes. They are certainly longer than the primary path and increase the average path length, but improve the PLT as well.

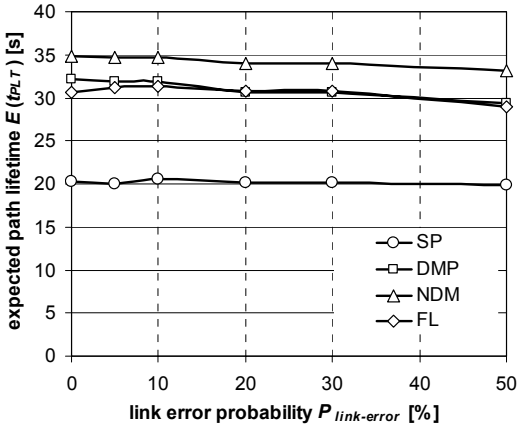

Figure 45: Average path lifetime for varying link error probabilities.

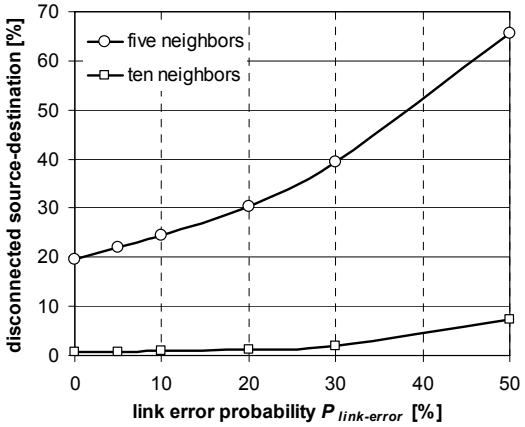

Figure 46: Inaccessible destinations as a function of link error probabilities.

The average number of neighbors is an important parameter for reasonable ad hoc networking. Figure 46 shows the probability P_{dc} that arbitrary sources and destination are disconnected for various node densities and link error probabilities. This probability is closely related to the network connectivity described in [143]. The network connectivity depicts the theoretical probability that parts of a network become separated. The probability that a network is disconnected is higher than the probability that any two nodes are unable to discover a path between each other, because even if the network is disconnected, both nodes can still be within the same part of the network. Therefore, from a user's point of view, only the probability that his node becomes disconnected from its destination is of importance.

With only five neighbors, 20% of all requested connections are impossible due to partitioned networks. With ten or more neighbors on average, networks are usually fully connected. With five neighbors, the number of impossible paths rises from 20% without link errors to almost 70% for 50% link error probability. With ten neighbors, the number of disconnected node pairs increases from zero to only 8% with every other link broken (Figure 46). For denser node distributions (more than 20 neighbors on average), the link error probability has no impact on the number of inaccessible destinations.

An average of only three neighbors already results in 70% of all destinations inaccessible. In case of coincidental link errors, the network is not functional anymore. Networks with node distributions smaller than five neighbors are unsuitable for reasonable ad hoc networking. On the other side, with high node densities, the possibility of undisturbed communication between adjacent nodes decreases rapidly. Nodes often block transmissions of neighboring nodes and the overall performance of

the network decreases again. Therefore, a node density of ten neighbors is a good trade off between both contrary requirements.

For all analyzed simulation setups and with assumed transmission ranges of 100 m, PLTs last usually much shorter than 100 seconds. Paths with many hops break even more premature. As only short paths allow reasonable PLTs, the usage of TTL extensions is proposed, in order to limit the maximal path length. This certainly limits the usability of MANETs. However, the unrestricted route setup with distant nodes increase the combined routing overhead while it does not increase the overall network performance. As described, DMP shows best results with up to four consecutive paths, whereas NDM slightly increases the PLTs with each new path. However, the achievable improvement decreases with every additional path.

For varying network conditions, different routing algorithms are most suitable. In very dense networks, the DMP algorithm outperforms NDM, as numerous disjoint paths exist. In networks with sparse node distribution, the SP is the best alternative. It simply discovers a single path and performs almost as good as the other algorithms. Under these conditions, all multipath strategies fail to discover reasonable backup paths to improve the average PLT.

In all other cases, the NDM algorithm shows better results, due to its greater flexibility. It improves results with respect to PLTs and average path length. Especially the reduced average path lengths also decrease the energy consumption for data packet transmissions. It also copes better with error prone networks, as it is able to reuse certain important links, rather than avoiding these links, as the DMP routing strategy demands. NDM is equivalent to the DMP within dense networks and equivalent to the SP in sparse networks. Hence it shows best results within all kinds of network environments. The flooding algorithm significantly outperforms the SP, while it never reaches the performance of the DMP strategy. Even so, as it does not cause any additional routing overhead, it is a good alternative within congested networks.

3.6 Routing overhead analysis

In order to allow estimations about the necessary route diversity overhead, the following section briefly describes and evaluates the caused overhead for a complete route request (RREQ), route reply (RREP) and route error (RERR) cycle. The analysis considers packet and byte overheads as well. The packet overhead O_P takes into account all packets solely bound for routing issues. Data packets with additional routing information are not considered as packet overhead. Whereas the byte overhead O_B adds up all message sizes of routing packets plus the additional routing information within data packets.

As additional packets cause much more overhead on the physical and medium access layer than piggybacked information in existing packets, the packet overhead O_P affects the overall network performance more significantly. In contrast to that, the byte overhead depicts the overall complexity of a routing protocol. Therewith, small O_P values are an indicator for the ability of a routing protocol to scale well with increasing network sizes while small O_B values show the capability of the algorithm to minimize the necessary transferred information between all involved nodes on a route.

The evaluation simplifies the routing behavior, to reduce the complexity of the theoretical overhead analysis. The network is always connected, and therewith at least one route exists between source and destination. Additionally the theoretical analysis omits the case that paths break during route setups. RREPs always contain valid routes back to the source and they are theoretically able to reach the source.

In order to keep the number of variables within the calculations transparent, a reduction to essential network parameters is necessary. Common ad hoc networks can be differentiated using two major network inherent parameters, their size, and their average neighbor node density, which is equivalent to the average node connectivity. Therefore, the considerations are based on these two parameters.

Although, the comparison includes different routing algorithms and extensions, the main function blocks (RREQ–RREP–RERR) of all proposed routing algorithms are comparable. Therefore, the theoretical analysis basically utilizes the packet structures of AODV as described in the IETF RFC 3561 [74].

Table 5: Necessary parameters for the theoretical overhead analysis.

Name	Description
O_P	Packet Overhead
O_B	Byte Overhead
O_{SR}	Source routing overhead
S_{RREQ}	Size of route request packet
S_{RREP}	Size of route reply packet
S_{RERR}	Size of route error packet
N	number of nodes within the network
n	average number of neighbors per node
M	number of discovered multipaths
m	number of used multipaths
h_i	hop length of path i
τ	Number of data packets over a set of path
ΔS	Size of additional information, like node ID or hop number
η	Proportional number of forward ratio for all route discovery error packets
F	Number of forwardings during a route request ring search cycle

Additionally, the evaluation considers the expanding ring search mechanism as advanced route request mechanisms within the evaluations. This algorithm does not perform a full network wide search for the destination node in the first place. However, it starts with e.g. a hop maximal of four hops. Each forwarding node decrements the time-to-live field (TTL) within the route request. In case the TTL field reaches zero, the node discards the packet. If the destination is not within the four hop radius, the source initiates a new search with a larger hop limit. Only as final option, it launches a network wide route request. In case the source node previously was connected to the destination and thus knew the hop distance, it can reinitiate a route search with a hop limit slightly greater than the original hop distance. The possibility is high that the destination has not moved outside this radius. This significantly reduces the number of necessary forwardings F during a route request. However, the route request overhead still depends on the number of nodes within the network, as the average route length depends on the network size.

Table 5 presents the parameters necessary to describe the different overhead calculations for all examined ad hoc routing algorithms.

To simplify the formulas for multipath algorithms, the average path length $\overline{h^{\psi}}$ for a number of paths ψ in a set is introduced as

$$\overline{h^{\psi}} = \overline{h(\psi)} = \frac{1}{\psi}\sum_{i}^{\psi} h_i \tag{41}$$

3.6.1 Theoretical routing protocol overhead

The following subsections explain the calculation of the respective overheads and show the particular formulas.

AODV

A full cycle of AODV messages contains the flooding, the route reply, the data packets, and the route error message. The packet overhead for a cycle thus includes all RREQ messages, the RREP, and the route error message.

As described above, the route discovery utilizes the ring search mechanism and the combined number of RREQ packets is F. Obviously, the average number of necessary hops to return the RREP to the source is equal to the average path length $\overline{h}$ between arbitrary nodes in the network. In addition, the average distance between a broken link and a source is half of the complete path length. Therewith, the RERR error message causes only half of the packet overhead compared to the route reply. The byte overhead O_B contains the same packets as the packet overhead O_P, multiplied with the size of the routing packet. AODV data packets do not contain additional routing information and therewith do not increase the overhead. Following the AODV RFC, a route request requires 24 Bytes, a route reply contains 20 Bytes and the size of a route error message S_{RERR} is 12 Bytes.

$$O_{P,AODV} = F + h + \frac{1}{2}\cdot h \tag{42}$$

$$O_{B,AODV} = F \cdot S_{RREQ} + h \cdot S_{RREP} + \frac{h}{2}\cdot S_{RERR} \tag{43}$$

AODV Flooding Extension

The packet and byte overhead of the AODV flooding extension contains the same basic parts as the original AODV overhead calculation. However, as the AODV extension utilizes each additional path, the number of discovered paths M is equivalent to the number of used paths m ($M = m$). Therewith, M is the basic path set for the overhead calculation. It solely depends on the number of received RREQ packets.

$$O_{P,AODV-FL} = F + M \cdot \overline{h^M} + \frac{M}{2}\cdot \overline{h^M} \tag{44}$$

$$O_{B,AODV-FL} = F \cdot S_{RREQ} + S_{RREP} \cdot M \cdot \overline{h^M} + S_{RERR} \cdot \frac{M}{2}\cdot \overline{h^M} \tag{45}$$

DSR Flooding Extension

The overhead calculation for the flooding extension of DSR is more complex. DSR only utilizes a subset of originally discovered paths and therewith both numbers M and m are not the same. While the route reply overhead depends on M, the route error overhead depends only on the number of used paths m with $M \geq m$. The number of discovered paths M again depends solely on the number of received RREQ packets. Additionally, the byte overhead for the DSR protocol must consider the source route overhead. As described, the ring search prevents network wide searches for destinations. Therewith, only the necessary numbers of RREQ forwardings F cause additional source routing overhead. Each node adds its own node address to the source route, and the path length increases by one for each additional hop. The combined additional byte overhead for the source route during the complete RREQ ring search cycle is O_{SR}, which is a function of the average RREQ hop length $\overline{h}$ and the structure of the search tree. A large fraction of the additional routing information is occupied by IP addresses. Therefore, the size of the additional information field ΔS is adjusted to the size of IP addresses. Consequently, ΔS encloses 32 Bit or 4 Byte respectively. Beside the regular routing information, DSR RREP packets include the source route twice, once as routing information back to the source, and for the second time within the payload field. In contrast to the RREQ, the routing information within these packets has a constant size. The route error packets generally have only 50% the length of the original source routes, and therewith add only half to the overhead calculation. All data packets contain source routes as well. The additional byte overhead ΔS is multiplied with the average number of data packets τ per set of routes times the square of the average path length of all used paths. The first multiplication in (47) is for the additional size of each individual packet, while the second multiplication is to represent the necessary number of hops between source and destination.

$$O_{P,DSR-FL} = F + M \cdot \overline{h^M} + \frac{m}{2} \cdot \overline{h^m} \tag{46}$$

$$\begin{aligned} O_{B,DSR-FL} = {} & F \cdot S_{RREQ} + \Delta S \cdot O_{SR} + \left(S_{RREP} + 2 \cdot \Delta S \cdot M \cdot \overline{h^M} \right) \cdot M \cdot \overline{h^M} \\ & + \left(S_{RERR} + \Delta S \cdot m \cdot \frac{\overline{h^m}}{2} \right) \cdot m \cdot \frac{\overline{h^m}}{2} + \Delta S \cdot m^2 \cdot \overline{h^m}^2 \cdot \tau \end{aligned} \tag{47}$$

Disjoint Multipath AODV

The packet overhead calculation for AODVM includes the same fundamental parts as the legacy AODV, and additionally the new component $\eta \cdot M \cdot \overline{h^M}$. This component models the additionally introduced packet overhead for route discovery errors. The variable η is the proportional RDER packet forward ratio. It depicts the proportional number of forwarded RDER packets traversing through the network during a complete routing cycle. It is the multiplication of the proportional number of hops an RDER traverses through the network η_L and the average number of unsuccessful route setup attempts η_M. Both values are defined as $0 < \eta_M, \eta_L < 1$. The overall number of necessary forwardings of these packets clearly depends on the average route length $\overline{h^M}$ and the number of discovered paths M. Since route confirmation information are piggybacked on data packets, they do not influence the packet overhead. Therefore, the packet overhead is calculated as

$$O_{P,AODVM} = F + M \cdot \overline{h^M} + \eta \cdot M \cdot \overline{h^M} + \frac{M}{2} \cdot \overline{h^M} \tag{48}$$

The first data packet traversing a newly discovered path contains piggybacked confirmation information. The part $S_{RRCM} \cdot M \cdot \overline{L^M}$ within (49) models this additional byte overhead. The authors of the AODVM protocol do not specify the size of the RRCM information, but their description in [116] implies that the RRCM packet size contains 4 Byte of overhead. Following previous overhead calculations, the byte overhead for RDER packets is computed as number of routing packets multiplied with the size of a single RDER packet ($S_{RDER} \cdot \eta \cdot M \cdot \overline{h^M}$). Again, the protocol description does not contain any exact values for the RDER packet size but the explanation implies 16 Byte of extra overhead.

$$\begin{aligned} O_{B,AODVM} = {} & F \cdot S_{RREQ} + S_{RREP} \cdot M \cdot \overline{h^M} + S_{RERR} \cdot \frac{M}{2} \cdot \overline{h^M} \\ & + S_{RDER} \cdot \eta \cdot M \cdot \overline{h^M} + S_{RRCM} \cdot M \cdot \overline{h^M} \end{aligned} \tag{49}$$

Non-Disjoint Multipath AODV

Following the previous calculations of AODV overheads, the non-disjoint extension of AODV requires the following packet and byte overheads to maintain the network.

$$O_{P,AODV-NDM} = F + M \cdot \overline{L^M} + \frac{M}{2} \cdot \overline{h^M} \tag{50}$$

$$O_{B,AODV-NDM} = F \cdot S_{RREQ} + S_{RREP} \cdot M \cdot \overline{h^M} + S_{RERR} \cdot \frac{M}{2} \cdot \overline{h^M} \tag{51}$$

As shown in the description of the routing algorithms in section 3.1.2, AODV-FL and AODV-NDM require equal routing information to maintain the network. Additionally, both protocols must consecutively utilize all M available paths. They are unable to detect prematurely broken routes. Therefore, they forward equal routing information and consequently cause the same amount of packet and byte overhead.

3.6.2 Simplifications

In order to verify the following assumptions and simplifications, simulations should validate the theoretical analysis. The simulation environment is simplified to allow sufficient repetitions and to receive statistical reasonable results. It contains N randomly placed nodes. The program chooses a random source and calculates the shortest path to each node. It utilizes a simplified route-request mechanism to discover each route. The source emits a request token with a certain time stamp. Each node within transmission range receives it and adds its own supplementary period at which it will forward the token again. The supplementary period is a uniformly distributed random number between 0.99 and 1.01. This emulates the internal processing time within each network node. The chosen range of the delay period prevents that the program discovers paths between source and destinations with more hops than the shortest path.

As with regular RREQ packets, nodes forward these tokens only once, and discard any subsequent received packets. After each node received the token, the program evaluates the average hop distance, the necessary RREQ overhead and the tree depth with respect to the neighbor density and the number of nodes within the network. It evaluates the

average Euclidian distance between the source and any destination for various node densities as well. To prevent border effects, the simulation considers only those paths with nine or less hops within networks with 900 or more nodes. For each individual set of node numbers and densities, the evaluation contains 5000 independent simulations. Therewith, the sizes of the 95% confidence intervals are always below 0.1% of the respective absolute value and therewith neglectable.

The average shortest path length

Obviously, the length h of the shortest path is equivalent to the average route length $\overline{h}$ within a random network. The average route length of all discovered paths $\overline{h^M}$ within a set of routes is in general larger than the length of the utilized multipaths $\overline{h^m}$. The used paths are mostly the shorter paths, while the discovered paths contain unused paths as well, and therefore these paths tend to be longer. Nevertheless, it is assumed that both path sets have equal average route length. The same relationship is present for the average multipath length and the length of the shortest path. The shortest path is by definition the path with the fewest hops. Therefore, any set of multipaths must utilize more forwarding nodes between source and destination on average. However, in case the number of multipaths within a set is small, the difference between the average path length of the shortest path and those within the multipath set is small. Therefore, we can assume with minor simplifications that

$$\overline{h^M} \approx \overline{h^m} \approx h \approx \overline{h} \tag{52}$$

The probability density function (pdf) $f_s(s)$ of the Euclidian distance s between two random points on a square plane of size $A = a \times a$ is

$$f_s(s) = \begin{cases} 2a \cdot s\left(s^2 - 4s + \pi\right) & \text{for } 0 \le s \le a \\ 2a \cdot s\left(4\sqrt{s^2-1} - \left(s^2 + 2 + \pi\right) + 4 \cdot \tan^{-1}\left(\sqrt{s^2-1}\right)\right) & \text{for } a \le s \le a\sqrt{2} \end{cases} \tag{53}$$

The complete description is in [144] and in [145]. Further on, the average distance $E(f_s)$ between two points on a square of size $A = a \times a$ is

$$E(f_s) = \frac{a}{15}\left(\sqrt{2} + 2 + 5\ln\left(1 + \sqrt{2}\right)\right) \approx 0.5214 \cdot a \approx \frac{1}{2}a \tag{54}$$

Consequently, $E(f_s)$ depicts the average distance between any two nodes within the simulation field. With (38), the size of the simulation square a^2 can be substituted with the number of nodes N within the network divided by the node density ρ ($a^2 = N/\rho$). Therewith, the average Euclidian distance $d(N, \rho)$ solely depends on N and ρ. In case, ρ is sufficiently large, the network is dense enough to approximate the ad hoc shortest path between two nodes with the distance $d(N, \rho)$. Therewith, $d(N, \rho)$ is equivalent to the average path length $\overline{h}$ multiplied with the average neighbor distance D

$$E(N, \rho) = D \cdot \overline{h} = \frac{1}{2}\sqrt{\frac{N}{\rho}} \tag{55}$$

To further simplify (55), the average neighbor distance D is necessary. The pdf-function for the radial neighbor location probability is $f(d) = 2 \cdot d/R$ with the distance $d \in [0,R]$ and R describing the maximal radio transmission range of each node. The average distance D

between two adjacent neighbors within a uniformly distributed network is the expected value $E(d)$. Solving the integral in (56), the result for the average normalized distance D is $2/3 \cdot R$.

$$D = E(d) = \frac{1}{R} \int_{d=0}^{R} d \cdot f(d)\, \partial d = \frac{1}{R^2} \int_{d=0}^{R} 2 \cdot d^2\, \partial d = \frac{2}{3} R \tag{56}$$

However, this result is only valid for a single hop communication. The Euclidian distance D_E for a path with more than one hop does not follow the simple rule $D_E = 2/3 \cdot h \cdot R$, with h describing the number of hops within the route. As known from previous sections, D_E additionally depends on the node positions of all previous relays and on the neighbor node density.

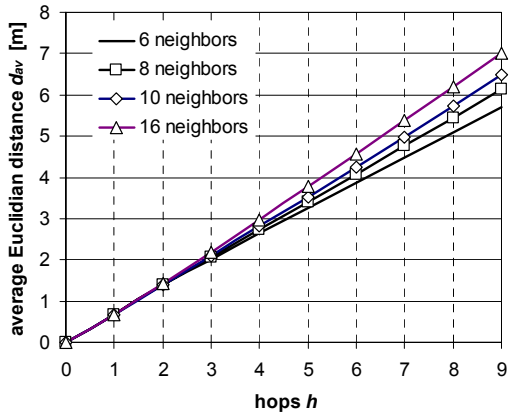

Figure 47: Average hop distance for various node densities.

As described, simulations determine the average distance between nodes for various numbers of hops and neighbor densities. Figure 47 shows these results. As expected, the first hop distance is 2/3. For all subsequent hops, the average distance varies with the node density. Increasing node densities lead to larger distances for each hop as well. The majority of neighbors causes forwarding nodes to be more far away from the sender during the route request flooding process. As nodes forward flooding messages only once, shortest paths tend to cover long distances between adjacent nodes. Within very dense networks, neighboring nodes must be close to the maximal transmission range, otherwise it is very likely that they have already received the flooding message. This correlation shows the graph in Figure 47. However, although the simulation considered a wide range of node densities, the absolute differences in Euclidian distances between different node densities are very small. Consequently, it is acceptable to use the results from (56) as a first approximation. Therefore, the average Euclidian distance can be approximated as

$$D = \frac{2}{3} R \tag{57}$$

for all network sizes and reasonable node densities. With $d(N, \rho)$ from (55), the average hop distance $\bar{h}$ between two terminals is

$$\bar{h} = \frac{E(N,\rho)}{D} = \frac{3}{4 \cdot R} \sqrt{\frac{N}{\rho}} \tag{58}$$

To further simplify (58) the definition of the average number of neighbors replaces the node density ρ. Consequently, ρ in (58) is substituted with $(n+1)/(\pi \cdot R^2)$. The formula in (59) depicts the average number of hops between any source and destination node within an ad hoc network with given number of nodes and known neighbor density

$$\bar{h} = \frac{3 \cdot \sqrt{\pi}}{4} \sqrt{\frac{N}{n+1}} \tag{59}$$

The necessary condition is that the node density is large enough so that a straight line (Euclidian distance) approximately matches the route of the shortest path. To fulfill this requirement, the number of neighbors must be greater than five.

The left graph in Figure 48 shows the theoretical results, the right depicts results obtained from simulations. Obviously, the theoretical results match the simulations very well. Due to the simplification with the average Euclidian hop distances, the theoretical analysis underestimates the average path length for small node densities, while it overestimates the route length in dense environments. Therefore, the introduced theoretical forecast of average route lengths within a broad range of network environments is utilizable for the following theoretical overhead analysis.

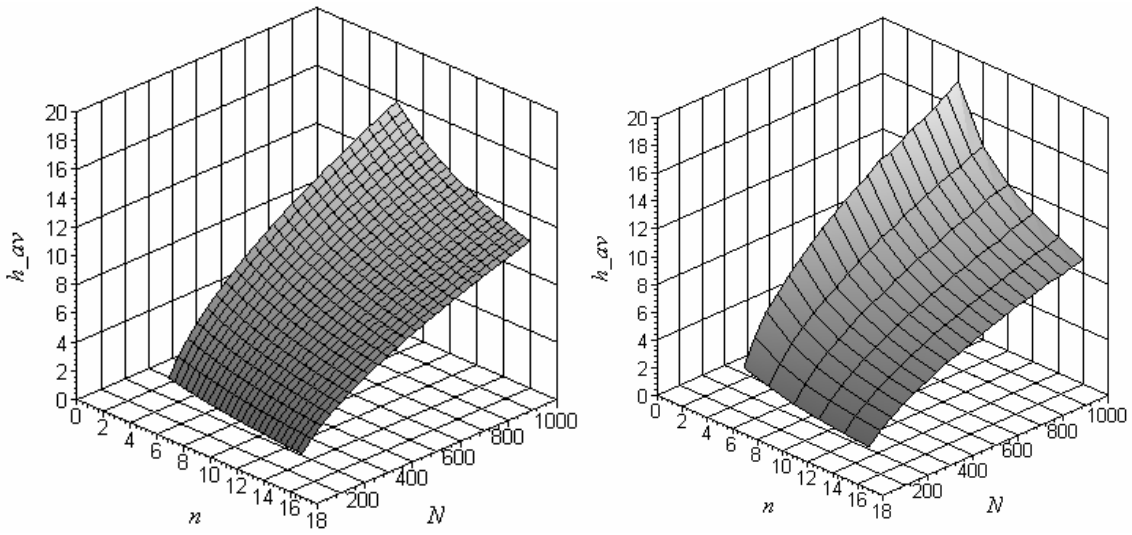

Figure 48: Average hop length $\bar{h}$ between source and destination for various network environments. Left: theory. Right: simulation.

The route request ring search overhead

The expanding ring search mechanism decreases the overall routing overhead during a route request. In order to determine the remaining overhead of the ring search instead of a full network wide search, a new approach is necessary. To keep the complexity of the deviation moderate, only a single ring search request is modeled. The average route length $\bar{h}$ determines the limit for the request as well. The number of necessary packet forwardings for a ring search is determined by *F*.

The sources of RREQ messages are the roots of virtual trees. All, but the destinations forward the requests exactly once. Hence, all nodes receive the RREQ messages, except those nodes solely accessible via a path through the destination. Additionally, nodes within close proximity to the destination may receive the RREQ over a path omitting the destination, although the shortest path would utilize the destination as relay. Therefore, virtual RREQ trees do not always represent the shortest path between sources and each other node within networks. Nevertheless, for simplification, it is assumed that each node receives the flooding message over the shortest possible path. Figure 49 depicts such a tree. Therewith, the average number of RREQ forwardings depends on the created tree structure during the route search and on the average route length $\overline{h}$.

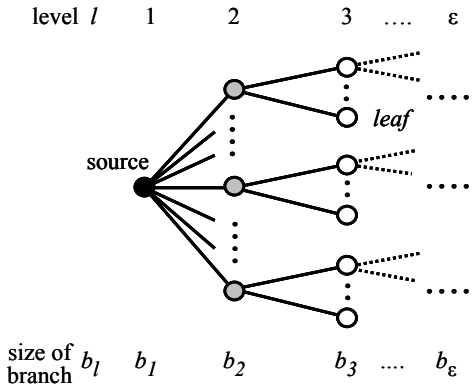

Figure 49: Generated tree structure of RREQ message.

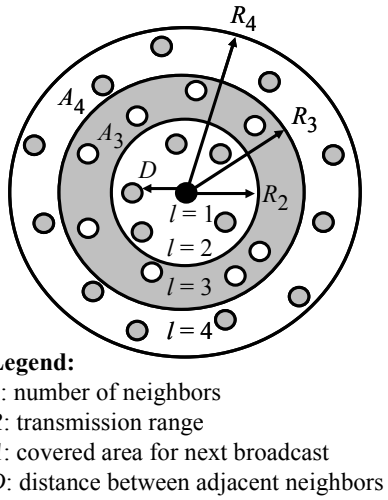

Figure 50: Coverage area for a RREQ forwarding.

As described, the first RREQ broadcast from the source towards its neighbors reaches about n nodes in general. Each subsequent broadcast does not reach a comparable number of new neighbors. Some nodes already received the message via a different path. Figure 50 highlights this relationship. In case nodes are uniformly distributed within the network, the number of undiscovered neighbors and the yet uncovered area of the RREQ broadcasts correlate.

With the above made assumptions, the RREQ tree is symmetric for all levels but the first two levels. Obviously, the first level contains only the source node, and the number of branches for level 2 depends solely on the number of neighbors n. For succeeding levels l, the number of nodes $b(l)$ within each level depends on the size of the covered area A and the number of neighbors. It is assumed that within all levels equal or greater than three, one third of the available nodes do not reach successors with their RREQ broadcasts. Consequently they are leafs within the tree. Following the approximation in (57), nodes have an invariant distance from their predecessors. The average distance is two thirds of the radio transmission range R. Consequently, $A(l)$ is an annulus with a width of $2/3 \cdot R$. The outer radius is

$$R_{out}(l) = R\left(1 + \frac{2}{3}(l-2)\right) \quad for\ l \geq 3 \tag{60}$$

while the inner radius is

$$R_{in}(l) = R\left(1+\frac{2}{3}(l-3)\right) \; for \; l \geq 3 \tag{61}$$

Therewith, the size is

$$\left.\begin{aligned} A(l) &= \pi \cdot \left(R_{out}{}^2 - R_{in}{}^2\right) \\ &= \frac{8}{9}\pi \cdot R^2 \cdot (l-1) \end{aligned}\right\} for \; l \geq 3 \tag{62}$$

Utilizing (62) we can compute the number of nodes $b(l)$ within each level of the routing tree. It depends on $A(l)$ and on the average node density. As described, within the first and second level, all nodes are forwarding nodes. For all subsequent tree levels, two thirds of the nodes are forwarding nodes (non-leaf nodes), with

$$b(l) = \begin{cases} 1 & l = 1 \\ n & for \quad l = 2 \\ \frac{16}{27} n \cdot (l-1) & l \geq 3 \end{cases} \tag{63}$$

While the remaining one third of the nodes within each level are leaf nodes, and consequently b_{leaf} is defined as

$$b_{leaf}(l) = 8/27 \cdot n \cdot (l\text{-}1) \quad for \; l \geq 3 \tag{64}$$

The tree depth is the average route length $\bar{h}$ and consequently F is determined as follows

$$F = \sum_{l=1}^{\bar{h}} b(l) = 1 + n + \frac{16}{27} n \cdot \sum_{l=3}^{\bar{h}} (l-1) \tag{65}$$

Solving (65), F can be simplified to

$$F = 1 + n + \frac{16}{54} n \cdot \left(\bar{h}^2 - 5 \cdot \bar{h} - 2\right) \tag{66}$$

with $\bar{h}$ from (59). Again assuming $N \gg n$, F can be simplified to

$$F = \frac{1}{6}\sqrt{\pi} \cdot N - \frac{2}{9}\sqrt{\pi} \cdot \sqrt{n \cdot N} \tag{67}$$

Therewith, F is still proportional to N, while the ring search reduces the overhead by about 70% compared to a full network wide search with N necessary RREQ forwards. In order to verify the RREQ tree model, simulations determine the average tree depth of arbitrary networks. To compare these results with the theoretical analysis, the maximal number of tree levels is necessary. Therefore, all nodes within the network receive the broadcast messages, and consequently the number of messages is constant N. With ε the maximal number of tree levels, N is

$$N = \sum_{l=1}^{\varepsilon} b(l) = 1 + n + \frac{16}{27} n \cdot \sum_{l=3}^{\varepsilon} (l-1) \tag{68}$$

With the help of the finite geometric series

$$\sum_{i=1}^{k} i = \frac{k \cdot (k+1)}{2} \tag{69}$$

ε is calculated to

$$\varepsilon = \frac{1}{2} + \frac{3}{4} \sqrt{\frac{6N-6}{n} - 2} \tag{70}$$

The second theoretical solution of the parabolic equation always has negative results, and therewith is invalid. The tree depth $\overline{h^N}$ is equivalent to the average number of hops to all leaf nodes within the RREQ flooding tree. Therewith, the average tree depth is the sum of the path length to all leaf nodes $b_{leaf}(l)$ divided by the number of leaf nodes

$$\overline{h^N} = \sum_{l=3}^{\varepsilon} l \cdot b_{leaf}(l) \Big/ \sum_{l=3}^{\varepsilon} b_{leaf}(l) \tag{71}$$

with $b_{leaf}(l)$ from (64) and ε from (70).

It is possible to simplify (71), for network environments satisfying the requirement that the number of nodes N is much greater than the neighbor node density n ($N \gg n$). Therewith, the average tree depth $\overline{h^N}$ is

$$\overline{h^N} \approx \sqrt{\frac{3}{2}} \sqrt{\frac{N}{n}} + 2 \quad for\ N \gg n \tag{72}$$

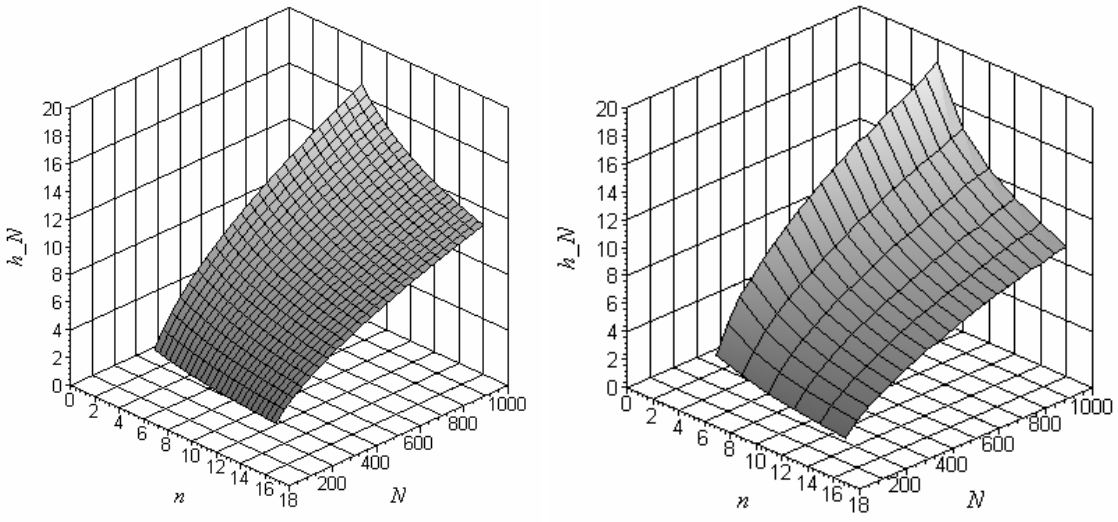

Figure 51: Average depth of an RREQ flooding tree for varying node densities *n* and network sizes *N*. left: theory, right: simulation.

The graphs in Figure 51 show the theoretical outcome from (72) and the simulation results. The results consider a different aspect of the network than those in Figure 48. They solely depict the average route length to all leaf nodes within the network, and not

the paths to all nodes. For large N, the average tree depths is proportional to the square root of network nodes ($\overline{h^N} \sim \sqrt{N}$), while it is inversely proportional to the square root of neighbor nodes. Although, the proposed model is simple, it meets the expectations. The simulative obtained results and the theoretical results match well. This proofs the validity of the simplified ring search model as well.

The source routing overhead

The additional overhead O_{SR} for source routing based requests depends on the depth of the request tree and the branching of the tree $b(l)$. The tree level l depicts the size of the source route within each RREQ packet, and the branch size $b(l)$ represent the number of forwarded messages within this tree level.

$$O_{SR} = \sum_{l=1}^{\bar{h}} l \cdot b(l) = 1 + 2 \cdot n + 3 \cdot b(3) + ... + \bar{h} \cdot b(\bar{h}) \tag{73}$$

Rewriting the above equation leads to

$$O_{SR} = 1 + n + \frac{16}{27} n \cdot \sum_{l=3}^{\bar{h}} l \cdot (l-1) \tag{74}$$

containing an altered finite geometric sum. With the help of a combination of two geometric series,

$$\sum_{i=1}^{k} \left(i^2 - i \right) = \frac{1}{3} \left(k^3 - k \right) \tag{75}$$

and a substitution of the sum in (74) to fit the series in (75), the equation (76) with $\bar{h}$ from (59) can be computed.

$$O_{SR} = 1 + \frac{1}{81} \cdot n \left(66 - 16 \cdot \bar{h} + 16 \cdot \bar{h}^3 \right) \tag{76}$$

Figure 52 shows the results of the source routing overhead for different network conditions. As expected, the overhead depends heavily on the depth of the flooding tree. The overhead rises with increasing number of nodes, although it does not follow the $\sqrt{N}$ proportionality of the tree depth, but is proportional to $N \cdot \sqrt{N}$ for large N.

For network sizes much greater than the average node density, the simplified overhead formula is

$$O_{SR} = \frac{N}{12} \cdot \sqrt{\pi^3 \frac{N}{n}} \quad for\ N \gg n \tag{77}$$

Therewith, the number of nodes within the network has a much greater influence on the overhead than the node density. For small networks, the overhead is neglectable, as only few packets traverse the network, and the size of the source route within each packet is short. Obviously, for large networks, the utilization of source routing algorithms accompanies with a severe disadvantage in routing overheard compared to protocols based on link-state routing.

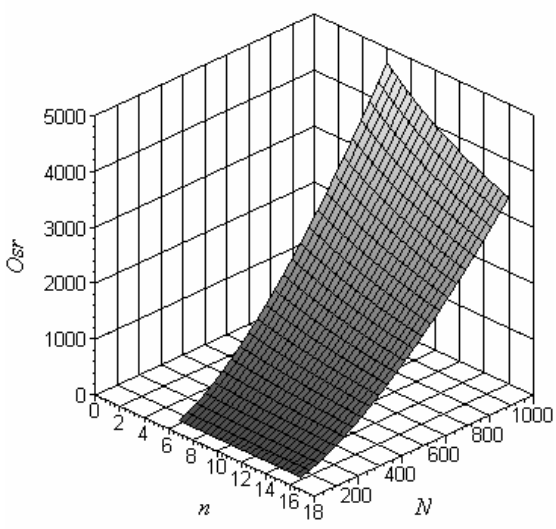

Figure 52: Theoretical analysis of the source routing overhead for an RREQ ring search flooding for various network configurations.

To consider the additional source routing overhead introduced by data packets, the average number of packets for a set of paths is necessary. The evaluation assumes a data rate of 100 kBit/s and a data packet payload size of 512 Bytes. This leads to 25 packets per second on average. From the previous section, the average path lifetimes for different network conditions is known. The path lifetimes depend on the route length and therewith is a function of the number of nodes and the node density Therefore, a combined path lifetime of 50 seconds for a complete set of routes is considered. Consequently, the average number of packets per path set is 1250 ($\tau = 1250$).

3.6.3 Overhead results

The following packet overhead considerations are based on the description of packet sizes and structures within the AODV RFC 3561 [74]. Table 6 contains all necessary values to accomplish the estimation. Additional overhead introduced by other routing algorithms is directly considered within the respective calculations.

Table 6: Overhead relevant parameters for AODV and multipath routing packets.

Type	Value	Description
S_{RREQ}	24 Byte	Size of route request packet
S_{RREP}	20 Byte	Size of route reply packet
S_{RERR}	12 Byte	Size of route error packet
S_{RDER}	16 Byte	Size of route discovery error packet
S_{RRCM}	4 Byte	Size of route confirmation message
ΔS	4 Byte	Additional overhead
η	0.2	Proportional RDER packet forward ratio
τ	1250	Number of data packets over a set of path
M	3	Number of discovered multipaths
m	2.5	Number of used multipaths

Simulations in the previous section show additional backup paths improve the overall PLT. However, the extent of improvements from utilizing four or five paths instead of three paths is not significant. Therefore, the number of discovered multipaths M is limited to three. As described in 3.6.1, some algorithms are able to detect already broken back-up routes and remove them from their list of available paths. Therewith, they employ only a subset of these M paths. In order to regard this behavior within the following analysis, the number of used path m is set to 2.5.

The inventors of AODVM do not provide simulation results for the frequency of unsuccessful establishments of backup routes. Therewith, the number of emitted RDER packets and the average hop distance between the emitting node and the destination remains unknown. For the following consideration, it is assumed that 40% of all routes show errors during setup and require the emission of RDER packets. Additionally it can be estimated that each individual RDER packet requires half the hops of the average path length to return to the destination. Therewith, the evaluation utilizes η_M with 0.4 and η_L with 0.5. The multiplication of both parameters leads to a value of 0.2 for the proportional RDER packet forward ratio η.

All following overhead considerations illustrate the byte and packet overheads caused by a single route request-reply-error cycle with respect to different network sizes and node densities. As described, the algorithms AODV-FL and AODV-NDM cause equal overheads. Without the loss of generality, all figures depict only the AODV-NDM graphs, but results are valid for AODV-FL as well.

Figure 53 depicts the packet overhead for all examined routing protocols under various network conditions. The left figure depicts the overhead for different network sizes and a constant network density of ten neighbors on average. Obviously, the overhead heavily depends on the network size. The initial route request ring search causes most of this routing overhead. Due to the fact that the number of routing packet forwardings F is proportional to N, the packet overhead rises at least proportional to N as well. As expected, AODV requires the least number of routing packets under all kinds of network conditions. It requires about 9% less packets than the other protocols considered in this work. AODVM requires slightly more packets, but the difference to the other two remaining algorithms is insignificant.

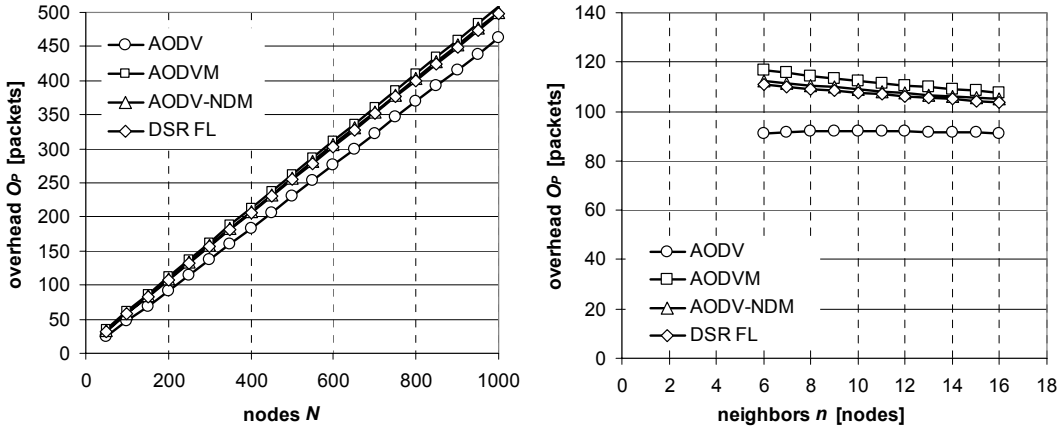

Figure 53: Left: Packet overhead for varying network sizes with ten neighbors on average. Right: packet overhead for varying node densities with a network size of 200 nodes.

The right picture of Figure 53 shows the results for a constant network size of 200 nodes and varying node densities. It depicts that the packet overhead does not significantly change with increasing node densities. The overhead of AODV remains almost constant for all kinds of network densities, while the overhead of the other four algorithms decrease with an increasing number of neighbors. The overhead reduction of AODV is in the range of 7% to 20% compared to the other algorithms. Additionally, the worse performance of AODVM is more obvious than in the left picture. Its RDER and RRCM messages require about 3% more packets than AODV-NDM and DSR-FL. However, the overhead difference between AODV-NDM and DSR-FL is irrelevant.

Figure 54 presents the byte overhead for the same network conditions as in the previous figure. The left side again depicts the overhead for increasing network sizes. Obviously, the overhead rises again almost proportional to the number of nodes within the network. In contrast to the packet overhead consideration, the flooding extension of DSR performs significantly worse than the other protocols. Its overhead increases with a factor of three, compared to all other algorithms. Because of the source routing approach, not only the routing packets but also data packets cause byte overhead. Therewith, the increasing number of bytes to store source routes combined with the enlarged number of necessary forwardings between source and destination cause this distinction.

As mentioned above, during the evaluation of the packet overheads, AODVM produces slightly more overhead in comparison to the other table-driven routing algorithms. Again, the additional RRCM and RDER routing packets cause this growth in byte overhead. The reason therefore is that increasing network sizes lengthen the average route length and therewith the effort to acknowledge found routes. The right part shows results for a constant node density of ten neighbors on average. Obviously, DSR-FL again performs worst for all node densities. However, the byte overhead is inverse proportional to increasing node densities. The decreasing average path lengths reduce the additional source routing overhead as well. The overheads of all other algorithms are almost independent from the node density. They mainly depend on the effort to perform route requests. As it is illustrated in (67), the ring search overhead F is proportional to the network size N, and almost independent from the neighbor node density n. The difference in byte overhead between the remaining three algorithms is insignificant.

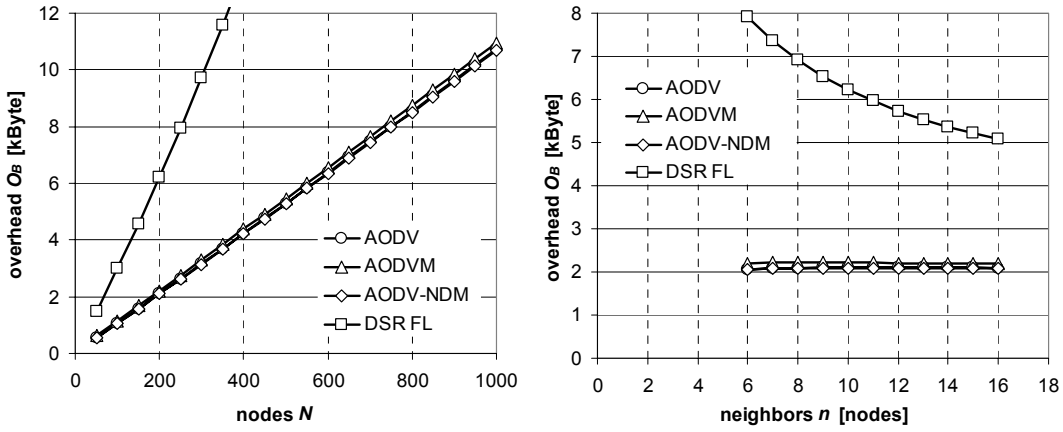

Figure 54: Left: Byte overhead for varying network sizes with 10 neighbors on average. Right: Byte overhead for varying node densities with a network size of 200 nodes.

Concluding, the non-disjoint multipath algorithm AODV-NDM improves the overall network performance in comparison to all other routing strategies. It causes a maximal multipath diversity overhead of 27%, but simultaneously increases the average PLT of more than 50% under most circumstances. The achieved benefit exceeds the necessary overhead. The usage of fully disjoint paths does not offer comparable improvements. It requires slightly more information exchange between nodes and therefore causes more overhead. Additionally, it degrades the performance in comparison to NDM algorithms. The requirement of disjoint path forces the algorithm to select backup paths with more hops than necessary and therewith leads to shorter PLTs.

3.7 Summary

The achievable PLT drops considerably with increasing path length. Simulations as well as the theoretical analysis in section 3.4 depict the negative behavior of PLTs with respect to the number of hops. Therefore, the utilization of a time-to-live field within RREQ packets is preferable. This TTL extension limits the maximal number of forwarding of RREQ packets and therefore restricts the maximal path length as well. A single user would notice the disability of his terminals to connect to distant nodes. However, otherwise discovered connections with numerous hops would create frequent route breaks and therefore cause an unacceptable user experience anyway. With a TTL extension, the overall network performance would improve while users with shorter connections would not be affected by frequent route request floodings. As expected, the number of multipath for reasonable networking is limited. One or two backup routes show best results, while more routes degrade the performance again. Utilizing more backup routes causes more overhead while not improving the PLT considerably. The additionally induced overhead exceeds the achievable improvement in PLT.

Results in section 3.5 illustrate that the non-disjoint multipath routing strategy allows an average PLT improvement of at least 50%. The DMP as well as the FL strategy improve the PLTs in comparison to the SP as well. However, the NDM routing strategy performs best under most networking conditions. Especially within large networks or within networks with a low mobility profile, multipath algorithms show significantly better results than the SP algorithm. In contrast to that, in highly dynamic networks, the SP performs almost as good as its counterparts. The achievable improvement with backup paths is not as significant. In these cases, backup paths often break before the primary paths shows permanent errors. Consequently, multipath algorithms must switch to another subsequent path and packet delivery is even more delayed and additional overhead is induced. Flooding based multipath algorithms do not discover optimal backup paths, but they still improve the PLT while not inducing considerably more overhead. Additionally, the utilization of FL algorithms does not require the usage of completely novel algorithms. As presented in 3.1.2, with some minor modifications AODV is able to enhance its performance with backup path based on the FL strategy.

The theoretical overhead analysis in section 3.6 shows that multipath algorithms based on AODV do not degrade the overall network performance due to increased overheads. The additional packet overhead ranges from 9% for 1000 nodes networks to 27% for 50 nodes networks. The additional byte overhead is almost neglectable. The major overhead is induced during the RREQ flooding. Subsequent route reply and route error packets do not affect the behavior of the overhead significantly. In contrast, the overhead heavily

depends on the network size. AODV causes least overhead in comparison to the other algorithms. This is independent of the network size or the node density. Obviously, the discovery of backup paths causes additional overhead. Although the packet overhead of source routing algorithms is as high as the packet overhead of distance vector protocols, the byte overhead is significantly increased. The inclusion of the complete route within all routing and data packets increases the size of all packets. In contrast to DSR-FL, the caused overhead variations between different AODV based multipath algorithms are neglectable.

Summarizing, the NDM routing strategy achieves 50% longer PLTs with 27% more routing overhead at most. Therefore, its performance exceeds the performance of existing SP strategies within most environments. In contrast to that, the DMP algorithm causes slightly more routing overhead and its discovered paths break earlier in comparison to NDM. Therefore, it is not the optimal choice for most circumstances. Within very small networks, the induced multipath overhead is much higher in comparison to single path algorithms. The RREQ flooding does not induce as much overhead and the fraction of overhead for multipath diversity is more significant. Therefore, within these environments the usage of SP is more appropriate.

4 Evaluation of Ad Hoc Networks in Urban Environments

After the analysis of path lifetimes in MANETs, the following chapter focuses on a further aspect in ad hoc performance evaluation. Existing performance evaluations [73, 77, 78] mainly consider the impact of routing algorithms on the packet loss or the routing overhead. The important results are summarized in section 2.5 as well. Following this argumentation, evaluations usually alter the number of nodes in a network, their respective maximal node velocities, as well as the induced network load. However, the impact of different environments (e.g. LOS, NLOS, urban or rural) on the network performance is usually not considered. Most evaluations assume line-of-sight between all nodes. Therefore, the packet reception mainly depends on the distance between transmitter and receiver. Although some simple initial approaches exist to analyze networks within non-line-of-sight environments, the published results do not allow an in-depth analysis. The environmental conditions are not yet recognized as one of the most important parameters for MANET performance. Therefore, this chapter introduces a reasonable simulation environment to allow in-depth evaluations of ad hoc networks within urban scenarios. It deals with all aspects of ad hoc performance and shows that different urban ground plans significantly impact the behavior of networks.

The remaining chapter is structured as follows: Section 4.1 gives a more general introduction into the ad hoc networking within urban scenarios. The following section 4.2 presents the existing ns-2 propagation models, and shows their inability to allow reasonable simulations within non-line-of-sight conditions. Some known network evaluations from the literature and their shortcomings to perform well within the proposed scenarios are discussed in section 4.3. Additionally, some reasonable node mobility models for urban environments are introduced in the same section. Existing path loss prediction models for various urban scenarios are presented in section 4.4. Additionally, the section examines the usefulness of the proposed models to achieve accurate predictions, while allowing short simulation periods. Section 4.5 presents a detailed description of the Walfisch-Ikegami propagation model, as it is the most suitable model. Moreover, a new urban mobility model is introduced to support characteristic node movements. Thereafter, the evaluation in section 4.6 considers all aspects of urban MANET conditions. The simulations include different urban scenarios as well as various known network parameters. Section 4.7 gives a conclusion and possible further work items.

4.1 Introduction

Ad hoc routing protocols are usually divided into proactive and reactive protocols, based on their respective route creation algorithm. As described in section 2.5.2, most existing

evaluations focus on the performance of ad hoc protocols in flat environments. They revealed the basic protocol behavior under certain network conditions. Proactive protocols are favorable for static networks or networks in which node mobility is low in comparison to the radio transmission range. In contrast, to that, reactive protocols usually outperform proactive algorithms in dynamic networks with frequent topology changes. Thereby, AODV achieves superior results under high load or in large networks. Evaluations predict that AODV scales well for networks with several hundreds of participants. In contrast to that, DSR is supposed to be optimal in small networks or with connections demanding lower packet rates. However, all evaluations were carried out with flat LOS-environments as simulation basis. This is obviously an artificial network setup and therefore depicts unlikely scenarios.

After years in which MANETs were primarily regarded as a research field, first real world applications emerge, either in military considerations, in vehicular scenarios or as coverage extension around access points. Certainly, these networks must cope with more challenging environmental conditions than those currently investigated. With the consideration of buildings and obstacles within the simulation area, algorithms must cope with higher network dynamics. Link breaks are more common, because the maximal radio range does not determine the link lifetime anymore. Obstacles may prematurely cause permanent link errors. Topology changes occur more frequently and therewith have negative effects on average path lifetimes. Consequently, routes must be shorter or the maintainable networks smaller. Additionally, the used routing algorithm must reestablish the paths which cause more routing overhead. The maximal packet rate requirements are more stringent, otherwise the networks becomes overloaded.

The novel consideration of building arrangements should clarify whether reasonable ad hoc networking is possible within urban environments. If so, it gives new insights into the behavior of algorithms and allows the development of improved protocols. The evaluations reveal whether different building arrangements have a varying impact on the overall network performance. Additionally, as open research issue, it has to be clarified which network parameters, like network size or maximal node velocity, behave differently in urban and unobstructed environments. In this case, the limits of the network have to be determined again, beyond which networking is not possible anymore.

4.2 Existing ns-2 propagation models

After the illustration of the particular challenges of urban ad hoc networking, the following section focuses on the existing possibilities to setup wireless simulations with the ns-2 network simulator. It solely concentrates on the channel and propagation model, a detailed introduction into the ns-2 is given in section 2.5. As described, the internal class structure of the ns-2 follows the ISO/OSI reference model. The channel class is responsible to forward transmissions to every node within the simulation area. An instance of the PHY-class within each node determines the receiving power P_r for every transmission, depending on the given propagation model and further simulation parameters. Therefore, the used propagation model has great impact on wireless network performance.

Obviously, simulations are unable to exactly reproduce real world environments. They must simplify, while still creating reasonable conditions. Especially the propagation

models must reduce the complexity and lessen the required computation to a minimum. This is necessary to accomplish reasonable simulations within acceptable time periods. They generally depend on various parameters. Some are directly determinable within simulations, like the distance between sender and receiver or the utilized frequency. But others must be represented as random functions or constant factors, like interference or fading effects. Currently, the network simulator ns-2 contains three different propagation models [88] to simulate wireless ad hoc networks, the free space (FS) model, the two ray ground (TRG) model and the shadowing model (SM).

The underlying transmission and reception model in ns-2 is simple. With the help of the transmitted power P_t, the simulator calculates the receiving power P_r for every transmission between two nodes with the chosen propagation model. The channel model distinguishes primarily between three cases. In case P_r is greater than the receiving threshold RX_{Thresh}, the transmission has enough power to allow proper reception at the receiver side. Figure 55 illustrates this context. Other simultaneous transmissions with reasonable transmission powers may certainly interfere with the current transmission and make a valid reception impossible. If P_r is below RX_{Thresh} but greater than the carrier sense threshold CS_{Thresh}, the receiving node is unable to correctly reconstruct the information and drops the packet. However, the receiving power of this transmission is still strong enough to corrupt other simultaneous transmissions. Another threshold determines whether simultaneous transmitted packets interfere. In case the receiving power of another packet is a factor of CP_{Thresh} greater than the power of interfering packets, the valid reception is possible. Usually, CP_{Thresh} is set to 10dB. Otherwise, both interfered packets are damaged and nodes drop them. Transmissions with receiving powers P_r smaller than CS_{Thresh} do not even interfere with other simultaneous transmissions at the same node. This case is only considered to improve simulation performance and to simplify packet processing during simulation. As the channel model of the ns-2 forwards all transmissions to all nodes, this is the most probable case.

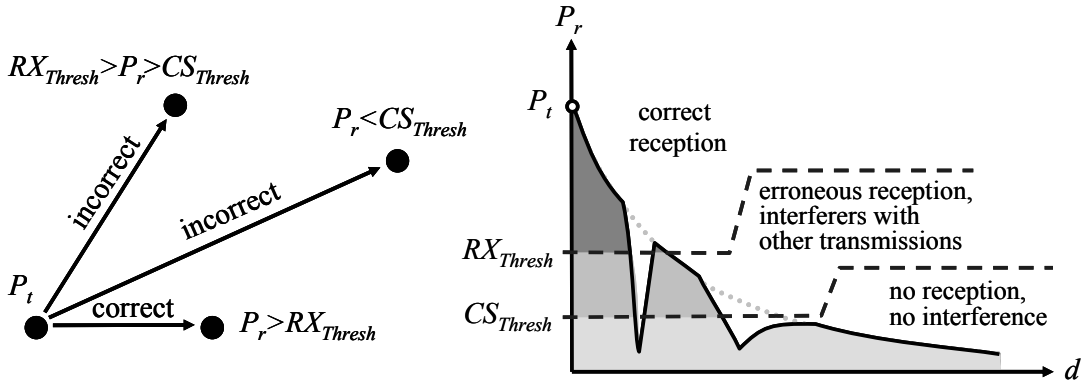

Figure 55: Physical layer model of the ns-2 network simulator.

4.2.1 Free space model

The FS model is the simplest implemented model. It is based on the considerations of Friis in [146] and only assumes the direct line-of-sight (LOS) path between transmitter *t* and receiver *r*. The receiving power P_r depends on the transmitted power P_t, the gain of the receiver and transmitter antenna (G_t, G_r), the wavelength λ, the Euclidian distance *d* between both nodes, and a system loss coefficient *L* with $L \geq 1$. All parameters, except the distance *d*, are system wide constant parameters and set at simulation start-up.

Usually, simulations assume omni-directional antennas with a gain of one and without additional system loss ($L = 1$). During simulation runs, the receiving power P_r only changes with the distance between sender and receiver. As both receiving parameters RX_{Thresh} and CS_{Thresh} are constant throughout simulations as well, receiving nodes are inside a perfect disc. Otherwise, they are unable to collect packets properly. The free space model calculates the receiving power as described in section (78).

$$P_{r,FS} = \frac{P_t \cdot G_t \cdot G_r \cdot \lambda^2}{(4\pi \cdot d)^2 \cdot L} \tag{78}$$

4.2.2 Two ray ground model

The TRG model extends the FS model. Besides the direct ray between sender and receiver, it considers the ground reflection as well. Figure 56 contains a schematic illustration. The considerations in [147] illustrate that this model gives more accurate prediction than the FS for long distances. As with the FS model, both nodes are assumed to be in LOS. The heights of both antennas over the ground are depicted with h_t and h_r. They are set at start-up and thereafter they are constant throughout simulations. However, the TRG model does not achieve acceptable results for short distances due to the oscillation caused by the constructive and destructive combination of both rays. Instead, the TRG model is equivalent to the FS model up to the crossover distance $d_{Thresh} = 4\pi \cdot h_t \cdot h_r / \lambda$. Beyond this distance, the ground reflection destructively interferes with the direct ray at the receiver and further reduces the field strength. The receiving signal strength is then inverse proportional to d^4. Just like the FS model, TRG contains only the distance between sender and receiver as the variable parameter. Consequently, the TRG model calculates the receiving power as

$$P_{r,TRG} = \begin{cases} \dfrac{P_t \cdot G_t \cdot G_r \cdot \lambda^2}{(4\pi \cdot d)^2 \cdot L} & d < d_{Thresh} \\ \dfrac{P_t \cdot G_t \cdot G_r \cdot h_t^2 \cdot h_r^2}{d^4 \cdot L} & d \geq d_{Thresh} \end{cases} \tag{79}$$

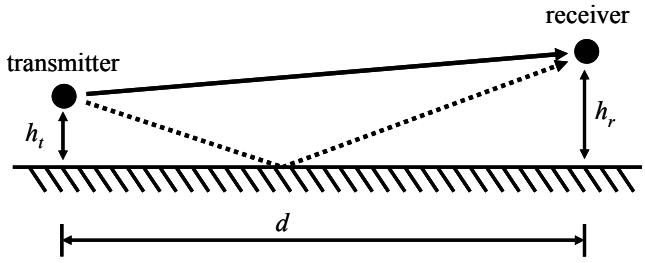

Figure 56: Two ray ground propagation model with its direct ray and the reflection.

4.2.3 Shadowing model

As described, throughout simulations, the sender-receiver distance is the only variable parameter for both described models. Therefore, the coverage area forms a perfect disc around sending nodes with exactly determinable range limits. Beyond this range, further receptions are not possible. To overcome this shortcoming, the shadowing model (SM) introduces a random component. The model contains two parts. The first comprises the

path loss while the second consists of a function with a random component. The first part contains the average receiving power $P_r(d)$ as a function of the distance d and the path loss exponent β.

$$P_r(d) = P_{r,FS}(d_0) \cdot \left(\frac{d}{d_0} \right)^{\beta} \tag{80}$$

The term $P_{r,FS}(d_0)$ calculates the FS signal strength for the reference distance d_0, which is usually set to 1 m. The path loss exponent β depends on the simulated environment and is constant throughout simulations. Table 7 presents some reasonable values taken from [88].

In order to achieve a random characteristic, the SM model introduces the random variable X. The probability distribution $\xi(x)$ of the random variable X is normal distributed with $N(\mu, \sigma^2)$. It has an average μ of zero. The standard deviation σ is called the shadow deviation and is invariant throughout simulations. Characteristic values for various environments are again depicted in Table 7. The values for the path loss exponent β as well as for σ were empirically determined.

Table 7: Typical values of path loss exponent β and the shadowing deviation σ [88].

Environment		β	Environment	σ in [dB]
outdoor	LOS	2	outdoor	4…12
	shadowed urban area	2.7…5	office, hard partition	7
			office, soft partition	9.6
indoor	LOS	1.6…1.8	factory, LOS	3…6
	obstructed	4…6	factory, obstructed	6.8

The receiving power of the shadowing model is calculated as

$$P_{r,SH} = P_{r,FS}(d_0) \left(\frac{d}{d_0} \right)^{-\beta} \cdot 10^{X}$$
$$with\ X(x): \{ x \in [-\infty, \infty] \mid \xi(x) = N(0, \sigma^2) \} \tag{81}$$

The model calculates the receiving power for every packet reception independently. Therewith, the shadowing model introduces some kind of unpredictability for data transmissions. Correct receptions are guaranteed for close proximities. However, they are impossible over long distances, whereas receptions over medium distances are unpredictable. Nevertheless, the statistical coverage area still forms a disc when considering numerous transmissions.

The unpredictability is also the great disadvantage of this model. The signal strength variations are not direction-dependent and possible errors can occur during every transmission. It varies significantly between consecutive transmissions and even differs for the reception of the same transmission at different receivers. This might force ad hoc routing algorithms to establish new routes, even if packet losses are one-time events and following packets would be successfully received.

As shown, the receiver signal strength of all currently implemented propagation models for the ns-2 solely depend on the distance between sender and receiver as variable

parameter. All other parameters are constant throughout simulations. Only the SM model is able to approximate real environmental conditions. But it sets the parameters already at startup, and therefore, the behavior within the complete simulation is predefined again. Its introduced variations in radio transmission ranges do not rely on current environmental and network conditions.

With those simplified models, reasonable simulations of the behavior of ad hoc routing algorithms in urban areas are impossible. Within metropolitan environments, radio transmissions are heavily direction dependent. Correct packet receptions do not only rely on the distance between sender and receiver but also on the surrounding conditions. A LOS situation has significantly longer transmission ranges than a path screened by obstacles, like buildings. Additionally, urban simulations may contain transmissions through walls, if one or both participants are indoor. A reasonable propagation model must distinguish between all those cases on a per packet basis and not globally at startup.

4.3 Related work

4.3.1 Ad hoc simulations in non-flat environments

A comprehensive number of papers exist, focusing on the performance of ad hoc routing protocols. Section 2.5 presented the main contributions and gave insights into the behavior of ad hoc routing algorithms under various network conditions. However, the major drawback of these simulations is their simulation setup. Although application scenarios for MANETs are mostly situated in urban areas or indoor, they assume flat simulation areas. All nodes are within line of sight (LOS) and correct packet reception depends only on the distance between sender and receiver. Therefore, this section presents existing work focusing on performance measurements within non-flat environments.

The earliest work considering non-flat environments for ad hoc network simulations is from Johansson in [77]. Besides common performance considerations with the TRG propagation model, it carries out simulations with a simple LOS model. Waves are unable to penetrate walls or reflect at them. Therefore, communication is completely prevented in case obstacles separate both communication partners. The authors carried out simulations within three scenarios, mostly differing in the grade of node mobility. In general, AODV as well as DSR are able to cope with all three scenarios. With the first two low mobility profiles, both algorithms show very low packet loss, only marginal differing from previous results within flat environments. The high mobility case causes frequent link breaks and requires continuous route reestablishments. Both algorithms perform almost equal and show significant packet loss rates of about 40%. However, results do not reveal new insights, because mainly the higher network dynamic causes higher packet losses. It remains unknown how MANETs perform within different urban environments.

Seet et al. in [148] for example, uses the same simple LOS model with no wall penetration within a Manhattan grid scenario. The authors choose a grid distance between street intersections of 400 m. Common WLAN 802.11b usually does not support transmissions over those distances. When assuming only outdoor connections, the model is reasonable. However, it is again insufficient for arbitrary building arrangements.

Another model to simulate urban environments is based on the possibility to directly vary links between adjacent nodes. Therewith, the approach is related to the already existing ns-2 shadowing model. However, instead of varying the link quality on a per transmission basis, the new model alters the existence of links between nodes on a temporal basis. The ns-2 shadowing model does not allow statements about the reception probability of two consecutive transmissions. On the contrary, an ON/OFF process on link basis usually achieves a high reception probability of packet transmissions under the assumption that the previous transmissions were successful as well. The authors in [149] use a static network setup and model the mobility of nodes as well as terrain conditions with the link up-time fluctuations. They use an independent Markov-modulated Poisson process to model the link up-times. A more advanced model could utilize a random node mobility model and then in addition the link up-time fluctuations to model the terrain. However, as with the ns-2 shadowing model, the Markov process is unable to simulate certain local environmental conditions, but only global situations. Therefore, it is again unsuitable for MANET simulations in urban environments.

To our knowledge, no further publication exist, considering the impact of more advanced propagation models on wireless ad hoc network performance.

4.3.2 Urban mobility models

Section 2.5 explains the commonly utilized random direction mobility model for flat environments. This model assumes open, unobstructed areas, in which nodes move only in accordance to the characteristics of the mobility model. Besides, a few models are available to model node mobility in city environments.

The city section mobility model in [134] is closely related to the Manhattan model in [150]. Both simulate streets within a city or on a campus. The ground plan is usually a building grid and node movements between buildings are limited to a single dimension. Only at intersections, nodes are able to change directions. A random process assigns each node a starting position and afterwards they move towards their next destination on the shortest path. When assuming a vehicular network, nodes are able to follow safe driving characteristics, like speed limits and a minimum distance between any two nodes. However, the model does not support random city ground plans, but only Manhattan grid style scenarios.

An extension to the above models is the M-grid presented in [151]. It utilizes cars and buses as two different types of vehicular nodes and assigns both different characteristics. Buses show higher movement regularity and a lower mobility compared to cars. They accomplish predefined round trips whereas cars move randomly. The model assumes that all streets utilized by buses are main streets. Cars give a higher preference to turn at intersections into main streets than into regular streets. The model tries to precisely imitate city traffic. However, new simulations with varying ground plans cannot be created automatically, because the bus routes have to be predetermined.

The authors in [152] present the obstacle mobility (OM) model, which supports arbitrary ground plans. It is based on the assumption that pathways lie “halfway-in-between” adjacent buildings. The graph in Figure 57a) illustrates these pathways for an urban environment with three buildings. The vector based Voronoi diagram of the two-dimensional ground plan is able to exactly describe these pathways. The corners of all obstacles are the basis to determine the movement graph. The Voronoi diagram has the characteristic that its calculated line segments are equidistant to two corners and

intersections are equidistant to three corners. The OM model uses the line segments as possible pathways for node movements. It randomly assigns each node an initial position and a destination. Nodes move towards their destinations utilizing the shortest path between both locations. The edge weight of each line segment is its Euclidian length. Additionally, nodes are able to move through buildings to reach their destinations. Upon reaching these, nodes pause for some rest period before they get assigned a new movement destination. The great advantage of the OM model is certainly its ability to cope with any arrangement of buildings. However, the calculated Voronoi graph limits the node movements again to a single dimension, whereas turns are only possible at intersections.

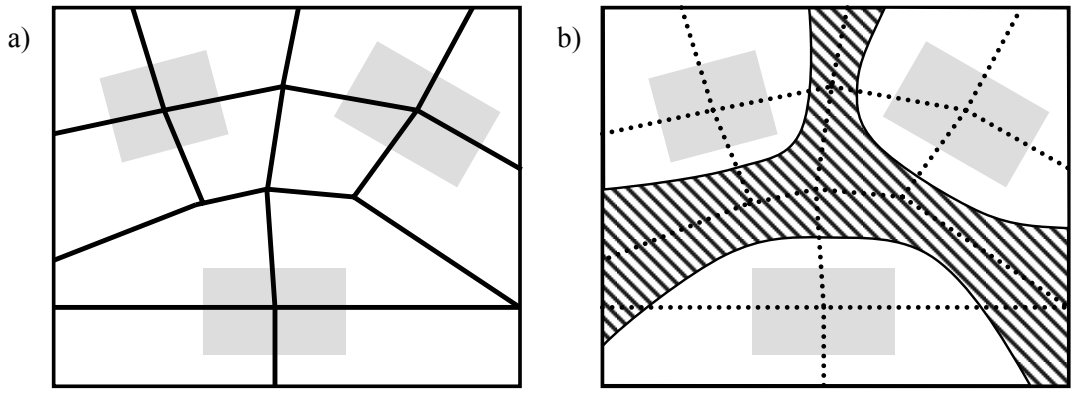

Figure 57: Left: Obstacle Mobile Model, Right: Voronoi Mobility Model.

In [17], an improved mobility model based on the OM model is presented. Instead of using the corners of obstacles as basis points it uses the building walls as basis for the Voronoi graph. Therewith, the resulting graph does not contain line segments but curves as pathways. These curves of the pathways together with the walls of buildings are again used as basis for an additional second order Voronoi diagram. The result is a corridor between the buildings, in which nodes can move according to e.g. the RD model with bounce back at the corridor boundaries. Therewith, the resulting mobility model is simple, as it only assigns each node an initial direction, velocity, and movement duration. The graph in Figure 57b) presents the mobility corridor of the improved Voronoi mobility model. The depicted results are based on the same building arrangement as in the left graph. Nodes are able to move into buildings with a certain transition probability and back again. The model combines the advantages of the OM model with the possibility of unrestricted node movements from flat node mobility models. However, it was just recently invented and therefore could not be used as mobility model for the ad hoc network simulations described in the following section.

4.4 Existing path loss prediction models

As depicted, neither the ns-2 integrated propagation models nor the propagation models used for previous non-flat ad hoc network simulations are able to allow reasonable statements concerning the performance of MANETs within urban environments. Therefore, the following section should give an insight into existing path loss prediction models and evaluate them for further usage.

Within the literature, numerous path loss models for various applications are introduced. As described, the environment invariant models like the free space [146], the one-slope [153] or the dual-slope model [154] are not applicable. Therefore, only those models are useful as basis for an integrated model, which predict the field strength with respect to a certain ground plan. Besides this adaptability, a chosen model must meet several requirements. It has to be fairly accurate, allows a simple computation to achieve short simulation times, and is deterministic to permit the reproduction of certain simulation results. The existing models can be divided into methods for three different scenarios. The first set is suitable to predict the coverage in small macro or micro cells in dense urban areas, the second is solely utilizable for indoor propagation, while the third focuses on outdoor-to-indoor propagation. The authors in [155] and [156] give a comprehensive overview over several existing propagation models for all three scenario setups.

4.4.1 Outdoor prediction methods

The propagation prediction in small macro-cells can be subdivided into vertical plane models (VPM), multipath models (MPM) and horizontal plane models (HPM).

For VPM models, the most important propagation effects are along streets in the LOS case, and the diffraction over roof-tops in case of NLOS situations. The Walfisch Ikegami model (WIM) [157, 158] can be classified as VPM method. It was especially designed for cellular network propagation prediction and achieves small deviations compared to real-world measurements. It is applicable when buildings are approximately of equal heights and the mobile nodes are located within street canyons (below roof-top). It assumes LOS propagation and over-roof diffractions for the NLOS situation. Figure 58 depicts the propagation in the NLOS case. It requires only few calculations to achieve sufficient accuracy for signal strength predictions in urban areas [159]. The 'European Co-Operation in the field of Scientific and Technical research' (COST) developed the model and the International Telecommunication Union (ITU) accepted it as propagation model for cellular networks. The COST 231 project [160] enhanced it for predictions in urban cells with small to medium radio transmission ranges.

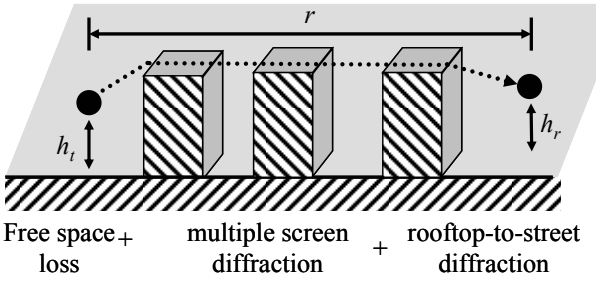

Figure 58: Walfisch Ikegami over-roof propagation.

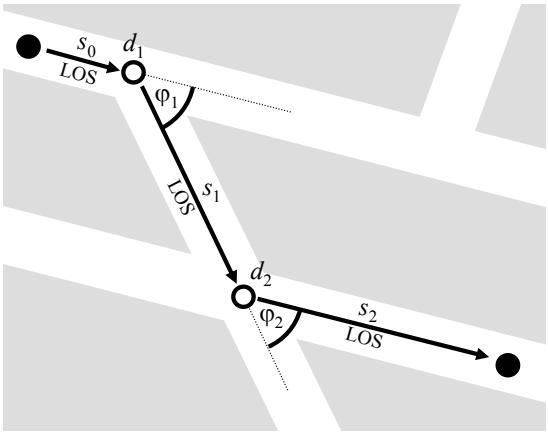

Figure 59: Berg horizontal plane propagation model.

If the heights of the diffracting edges are not homogenously distributed, knife-edge models [161] are more appropriate. They utilize the WIM model for LOS situations, but use another heuristic model for NLOS scenarios. Additionally, the impact of vegetation on the wave propagation can be modeled with a single-knife-edge diffraction over the tree next to the receiving node. However, the calculation of these more exact path loss values becomes more complex as well.

MPM models are mainly utilized for scenarios where the locations of transmitter and receiver are in close proximity. Multipath signals are dominant and contribute significantly to the received energy. Therefore, only buildings within a certain diameter around the sender are of certain relevance. However, the model fails to forecast path losses for single-path transmissions, e.g. in case two nodes are separated by a building, and the transmission paths are multiple diffracted.

Unlike the WIM model, the Berg model [162] is especially suitable to predict path losses in urban micro cells. It assumes building heights, which are considerably higher than the utilized antennas, and therefore are always below roof-top. It solely assumes wave propagations within the horizontal plane and therewith is disjoint to VPM approaches. Additionally, it does not consider propagations through buildings. The predicted wave propagation always follows the street canyons (see Figure 59). Along a street with LOS between sender and receiver, the Berg model utilizes the free space or the dual slope model. In an NLOS situation, the model divides the path into several LOS sections s_i and diffractions d_i at the transitions. Transition points are always in the center of the respective street intersections. Each LOS section again predicts the additional path loss with the help of a LOS model. To determine the diffraction loss, the model uses a function of the opening angle φ_i between streets. Therewith, the model recursively determines the imagined distance between sender and receiver. Figure 59 shows an example with two diffractions and three LOS sections. The berg model especially allows accurate predictions in the 5 GHz band, in which over-roof propagations with multiple screen diffractions are neglectable. Besides the WIM model, the ETSI standardization institute proposes it as model for UMTS test scenarios [163].

The simplest urban propagation model considers only LOS propagation. Therefore, in NLOS situations with obstacles between sender and receiver, the model assumes field strengths of zero at the receiver side. Johansson utilizes this model in [77] for indoor as well as for outdoor simulations. Section 4.3.1 presents further details. However, the model is only utilizable for high frequencies beyond 5 GHz or very low power transmissions, in which wall penetration, diffractions, and even reflections are neglectable. Broadband wireless access scenarios in urban pico cells usually assume very short transmission ranges and therefore a LOS communication. Therefore, the model only has to distinguish between LOS and NLOS cases, and is able to quickly predict path loss values. However, the proposed scenario is an urban ad hoc network with WLAN 802.11b as MAC protocol. With its 2.4 GHz frequency, not only the LOS propagation contributes to the signal strength. Consequently, this model is oversimplified and therefore insufficient.

4.4.2 Outdoor-to-indoor prediction

The second type of models focuses on the prediction of indoor coverage from outdoor transmitters. Two different empirical approaches are known [164]. The first approach uses a model in which indoor path losses are calculated as function of all surrounding

outdoor path loss values. An empirical determined penetration loss factor describes the signal strength difference between adjacent outdoor and indoor positions. The received indoor energy level varies from floor to floor due to outdoor diffraction effects. An also empirically developed height-gain model forecasts the variations between the path loss at ground level and the path loss at a certain building heights. The height-gain model in [164] reports a gain of about 2dB for regular floor heights and 7dB for floor heights of up to 7 m. Therewith, a set of heuristics calculate the predictable path loss for all positions within buildings. The challenge is the determination of the relevant outdoor path loss values to achieve suitable results. The second semi-empirical method extends the previous model. In case of a LOS situation between sender and the building, it considers the angular incidence of a wave to predict the indoor propagation.

4.4.3 Indoor prediction

Indoor radio propagation models assume both, sender as well as receiver, within the same building. They usually do not consider certain ground plans but only take the penetrated number of walls and floors into account. The multi-wall (MW) model [164] is especially suitable for these situations. It calculates the indoor path loss as free space loss and adds losses for penetrated walls and floors. A value of 37 dB as minimal path loss results from an empirical evaluation for UMTS indoor environments within a single floor. A similar model is the Keenan-Motley approach which utilizes the one-slope model with additional penetration losses for supplementary floors. However they suggest a minimal path loss of 21dB. The multi-wall-and-floor model (MWF) extends the MW model. It considers that each preceding penetration alters the penetration loss. The authors in [165] introduce reasonable parameters for 1.9 GHz indoor environments.

The theoretical analysis in [166] presents a uniform mathematical model of diffraction for complex indoor radio environments for WLAN systems. Additionally, [167] characterizes the channel for a broadband 17 GHz indoor WLAN scenario. However, both introduced models are too complex and too specific to be utilized in a combined indoor-outdoor propagation model for ad hoc simulations.

4.4.4 Ray Tracing

Ray tracing models are special among the suitable models, because they are solely based on geometric considerations. They are applicable to predict wave propagations within all scenarios. The only limitation is that the utilized wavelength must be magnitudes smaller than the geometric objects within the area. This is necessary to allow the simplified characterization of electromagnetic waves as simple rays. For each sender position, rays are homogenously transmitted from a unit sphere with the sender at its center. Rays are followed until they reach an object and there they get diffracted or reflected. The field strengths and the new directions of rays are determined by the objects' geometries and electrical and magnetic characteristics. Rays approaching a thought detection sphere around the emitter account to the received signal. Obviously, the predicted propagation is highly accurate, as 3D ray tracing methods exactly reconstructs the wave propagation. However, as shown in [168], the computational effort is significantly higher, as each tuple of sender and receiver position requires an individual complex calculation to determine the path loss. Additionally, the accurate determination of the geometric and electro-magnetic characteristics of objects is crucial, and not always possible. The computational effort can only be reduced when considering only a limited number of reflections and diffractions, but then the field strength predictions again lose accuracy.

4.4.5 Rating and validation

As described in the previous section, the ray tracing approach allows the most accurate results, in case exact environmental information is available. In case geometric and physical information is imprecise or even incorrect, this approach shows severe shortcomings. Then, more generalizing models achieve superior results.

The Berg model is especially suitable for high frequency transmissions in street canyons. However, the model requires the determination of the dominant wave propagation path. Obviously, this is difficult to determine, and especially in dense urban environments several different paths may exist. The correct path determination is only achievable in homogenous environments, but almost impossible for arbitrary scenario profiles. In these cases, the computational complexity certainly is beyond an acceptable range. As last negative item, the path loss predictions to and from nodes are usually not equivalent. Two nodes may encounter a different connectivity status, and therefore routing protocols supporting unidirectional paths are necessary. AODV as well as DSR do not support those paths.

The recommendation of the ITU-R shows the common acceptance of the WIM model. However, the final COST report notes that the model is most appropriate for transmitters above roofs and in case of a regular building deployment. If the transmitter is below the roof of buildings, wave propagation within street canyons becomes more dominant and therefore, the prediction errors of VPM models increase. However, the modeling does not consider these errors. Nevertheless, although the WIM model was originally developed to forecast field energies in urban cellular networks, this model is also reasonable to predict signal strengths in MANET simulations. The few and simple calculations allow fast simulations even in large network environments. And in contrast to the previous Berg model, the path loss table is symmetric, additional considerations for unidirectional links are unnecessary. The knife-edge models achieve more accurate results compared to the WIM model. However, the additionally achieved accuracy does not justify the significantly increased computational complexity. As conclusion, the WIM model is elected as most suitable outdoor propagation model, although it has some shortcomings as well.

All previous models are focusing on outdoor predictions, and therefore are unsuitable to predict indoor or outdoor-to-indoor coverage. However, the proposed outdoor-to-indoor models simply use the outdoor field energy to determine the indoor field strength. The complexity lies in the determination of the dominant wave propagation path. Indoor propagation models usually predict the field strength with respect to the penetrated indoor walls and floors. More advanced models do not exist. For the sake of simplicity and to achieve fast simulations, both cases are neglected. The penetration loss for building walls are simplified with a constant value and indoor propagations are assumed to be always LOS.

4.5 WIM model and urban mobility model

As described, the WIM model is most appropriate to simulate ad hoc networks within urban environments. The following section gives a short introduction into the basic calculations and necessary model parameters. The subsequent section presents improvements to achieve reasonable simulation periods with acceptable path loss

prediction accuracy. Section 4.5.3 introduces a new urban node mobility model to allow simulations with reasonable network dynamic.

4.5.1 Walfisch-Ikegami path loss model

The European research project COST 231 combined the path loss models of Walfisch and Ikegami. The WIM model was initially developed to predict the path loss of GSM mobile communication infrastructure within micro and mini cells in urban environments. It is based on the assumption that the over-roof propagation is dominant within urban environments. Therefore, the model creates a vertical cut through the ground plan. It solely considers the buildings between sender and receiver for the path loss prediction. The original WIM model for GSM base stations supports sender positions over and below roof-top. The building profile of the vertical cut determines the values of the heights of building roofs h_{roof}, the average street width w, and the buildings separation b. The antenna heights of the transmitter h_t and the receiver h_r and their respective distance d are necessary as well. The model assumes a uniform building development and equal building heights. Figure 60 visualizes all parameters.

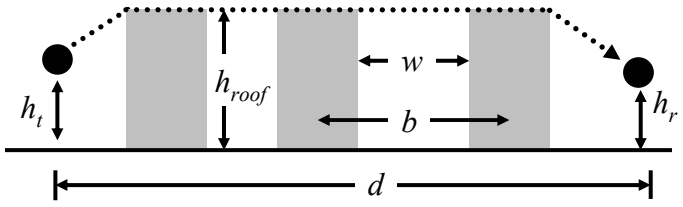

Figure 60: Necessary parameters for the WIM propagation model.

The path loss for the LOS case within street canyons is determined as

$$L_{LOS} = -35.4 + 26 \cdot \lg(d_{[m]}) + 20 \cdot \lg(f_{[MHz]}) \tag{82}$$

For the NLOS case, the WIM model only considers diffraction effects for the wave propagation over roofs. The approach of Walfisch and Bertoni [157] determines the path loss of the multiple screen diffraction (see Figure 58), while the model of Ikegami [158] determines the diffraction loss at the edge of the last building's roof.

Consequently, the NLOS path loss L_{NLOS} between transmitter and receiver superposes three independent path loss terms. It contains the free space loss L_{FS} between the transmitter and the first roof, the multiple screen diffraction loss L_{MSD}, and the final roof-to-street diffraction loss L_{RTS}

$$L_{NLOS} = \begin{cases} L_{FS} + L_{MSD} + L_{RTS} & for\ L_{MSD} + L_{RTS} > 0 \\ L_{FS} & for\ L_{MSD} + L_{RTS} \le 0 \end{cases} \tag{83}$$

with

$$L_{FS} = -27.6 + 20 \cdot \lg(d_{[m]}) + 20 \cdot \lg(f_{[MHz]}) \tag{84}$$

$$L_{RTS} = -16.9 - 10 \cdot \lg(w) + 10 \cdot \lg(f_{[MHz]}) + 20 \cdot \lg(h_{roof} - h_r) \tag{85}$$

$$L_{MSD} = K_A + K_D \cdot \left(\lg(d_{[m]}) - 3\right) + K_F \cdot \lg(f_{[MHz]}) - 9 \cdot \lg(b) \tag{86}$$

The expression K_A in (86) is responsible for a lower loss prediction, in case the transmitter antenna is above roof top. For varying distances r and frequencies f, the

terms K_D and K_F adjust the behavior of the path loss prediction over roofs. However, as the proposed ad hoc network evaluation contains only vehicular or pedestrian nodes, the common antenna heights for h_r and h_t is 1.5 m and therefore certainly below roof-top. Additionally, the field energy of ad hoc node transmissions in distances beyond 500 m is minimal, and hence the parameterization of K_D is unnecessary as well. Consequently, for MANET evaluations in urban environments, different adjustments for all three parameters are unnecessary. Therefore, the following formulas contain only the parameters for ad hoc network specific path loss predictions

$$K_A = 54 - \frac{1}{625} \cdot r_{[m]} \cdot (h_{roof} - h_r) \tag{87}$$

$$K_D = 3 - 15 \cdot \frac{h_t}{h_{roof}} \tag{88}$$

$$K_F = -5.5 + \frac{1.5}{925} f_{[MHz]} \tag{89}$$

Following the COST 231 evaluation, the predictions obtained by the WIM model are only valid for certain parameter ranges. Table 8 depicts these parameters.

Table 8: Valid parameter ranges for the WIM model.

Parameter	Range
frequency f	0.8…2 GHz
distance r	20…5000 m
transmitter heights h_t	4…50 m
receiver heights h_r	1…3 m

Obviously, the frequency range and the suitable transmitter heights are not within their valid parameter ranges. The common transmission technique IEEE 8021.11b for ad hoc communication utilizes the 2.4 GHz band. Hence, it is higher than the proposed 2000 MHz maximal frequency. As described, the transmitter heights of ad hoc nodes are usually 1.5 m, while the minimal transmitter height for optimal signal strength predictions is 4 m. The WIM model was originally developed to predict the signal strength of cellular base station and hence the COST project only verified reasonable parameter ranges for those networks. The accomplished evaluations never comprised ad hoc scenarios. However, both parameters are only slightly out-of-bounds. Therefore, the calculation errors are within reasonable ranges and the predictions are still suitable for ad hoc simulations.

4.5.2 Improvements

The usage of a modifiable building data base allows the evaluation of different urban scenarios without the need to recompile the ns-2 program suite each time. The exact structure of the data file containing the building profiles can be found in [169, 170].

It contains a sequence of an arbitrary number of buildings. Each set includes the number of edges followed by the X- and Y- coordinates of each edge separated by commas. As described in section 4.2, the ns-2 PHY-class calculates the field energy separately for each transmission and transmitter-receiver combination. Therefore, the computation effort is too high for a simultaneous calculation during ns-2 simulations, although the

WIM model allows fast and efficient path loss predictions. However, with the knowledge about the exact positions of buildings, the calculation of path loss values is separable from following simulation runs. This allows the preceding computation of the path loss within the simulation field. As a pre-computation could not cover an infinite number of possible combinations of transmitter-receiver positions, the program must divide the simulation field into tiles. All tiles have an equal rectangular size. A multiple of the edge lengths of the tiles defines the size of the simulation field. Additionally, the center of each tile determines the path loss of the complete tile. Hence the computational complexity only depends on the number of tiles within the simulation area. The program must calculate the path loss value between any two grid elements. With δ_x the number of tiles in x-direction and δ_y in y-direction, the number of necessary loss computations C is

$$C = \delta_x \cdot \delta_y \cdot \frac{(\delta_x \cdot \delta_y + 1)}{2} \approx \frac{\left(\delta_x \cdot \delta_y\right)^2}{2} \tag{90}$$

With $\delta_x = \delta_y = \delta$, the computational effort increases with $O(\delta^4)$. The bisectioning of edge lengths requires eight times more path loss calculations. To allow fast provisioning of the necessary tables, to minimize the size of output files, and to save memory during ns-2 simulation runs, δ must be as small as possible. Obviously, the reduction of the number of tiles has some disadvantages as well. In case a tile is partly covered by a building, the program must either recognize the tile as indoor or outdoor. Figure 61 illustrates an example for different tile sizes. The center of a tile solely defines its path loss value. Therefore, fractions of tiles are outside of buildings but have loss values as they would be inside or vice versa. Smaller tile sizes reduce the error, but again increase the computational effort. To determine a reasonable grid element size, multiple simulations with 5 m, 10 m, and 20 m edge length were carried out in [169]. Packet loss and routing overhead of various ad hoc routing algorithms differ only by 2% for varying element sizes. However, large edge length occasionally causes NLOS conditions for clear LOS situations. This falsification obviously increases with increasing edge lengths, although these variations only slightly influence performance. Therefore, an edge length of 20 m seems too simplistic and inappropriate. On the other hand, a tile size of 5 m × 5 m already requires long computation periods while it does not achieve a significant increase in accuracy. Therefore, simulations suggest an optimal tile size of 10 m × 10 m.

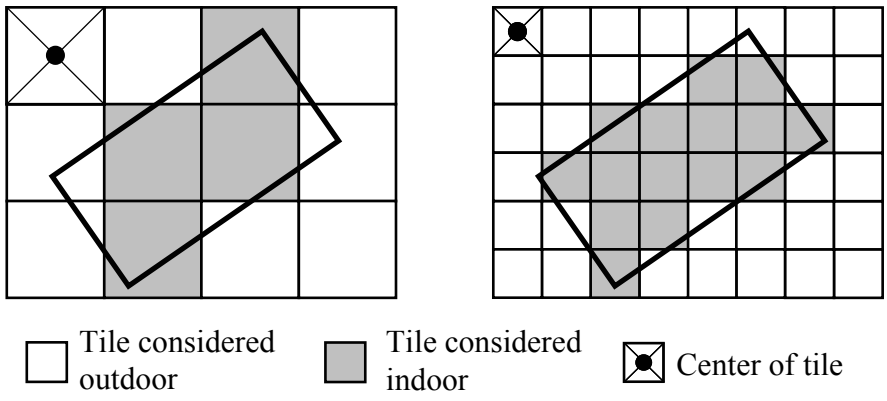

Figure 61: Impact of tile sizes on the accuracy.

As described, the WIM model only predicts the accurate path loss for transmitters and receivers outside of buildings. In order to allow simulations with nodes inside of buildings, some further computational methods are added. The program distinguishes between six different scenarios. The first is the WIM-LOS case, while the second covers the regular WIN-NLOS case. For loss calculations between positions inside buildings and locations outside, the program assumes LOS but adds an additional 20dB path loss to the result to consider the wall penetration. This assumption is a simplified method of the outdoor-to-indoor models presented in section 4.4.2. The propagation model does not consider walls and floors within buildings, and therefore two nodes within the same building are in LOS. For nodes in different buildings, the model again assumes LOS, but adds two times 20dB for both walls. In case one node is within a building and the other is outdoors with the direct path cut by an additional building, the model assumes NLOS with an additional 20dB for the necessary wall penetration. The last two cases usually lead to path losses of more than 100dB, and therefore surely do not allow packet reception.

The computation program saves all obtained path loss values Δ_{pl} in [dB] in a separate data file. As described, the number of computations depends on the dimension of the simulation and the size of the tiles. Therefore, the size of the data file depends on the same parameters as well. However, with the help of subsequent optimizations, memory and hard disk requirements could be reduced significantly. Evaluations in [170] illustrate that both are limited to less than 10% of the original consumption.

During simulation, the ns-2 program completely caches the path loss data base within its memory. Thereafter it is able to determine the path loss between any two positions without the need to recalculate it. During a transmission, the program detects the positions of both participants and selects the respective loss value Δ_{pl} from the data base. Subsequent to the calculation of the loss factor f_{loss} in (91), ns-2 computes the resulting receiver energy P_r. The labeling of variables in (92) follow the same nomenclature as presented in section 4.2. Thereafter, the receiving power is further computed to determine the ability of the node to properly obtain the transmission.

$$f_{loss} = 10^{\Delta_{pl}/10} \tag{91}$$

$$P_r = \frac{P_t \cdot G_t \cdot G_r}{L \cdot f_{loss}} \tag{92}$$

Figure 62 illustrates the direction dependent transmission range, as well as the connectivity of an arbitrary network. For illustrative reasons, the maximal LOS transmission range is set to 300 m. Both plots depict the Munich city center with a 500 m × 500 m simulation area. The left figure presents the radio coverage ($P_r > RX_{Thresh}$) around node 57. Obviously, it is able to reach more distant nodes in case they are within LOS, whereas it only achieves short ranges in NLOS cases or to indoor nodes. The right figure depicts the connectivity plot for the same network. The node connectivity is high for outdoor nodes within main streets, while the order of connectivity is low for nodes in side roads or indoor. Node 89 is even disconnected, as it is positioned indoor and afar from other nodes.

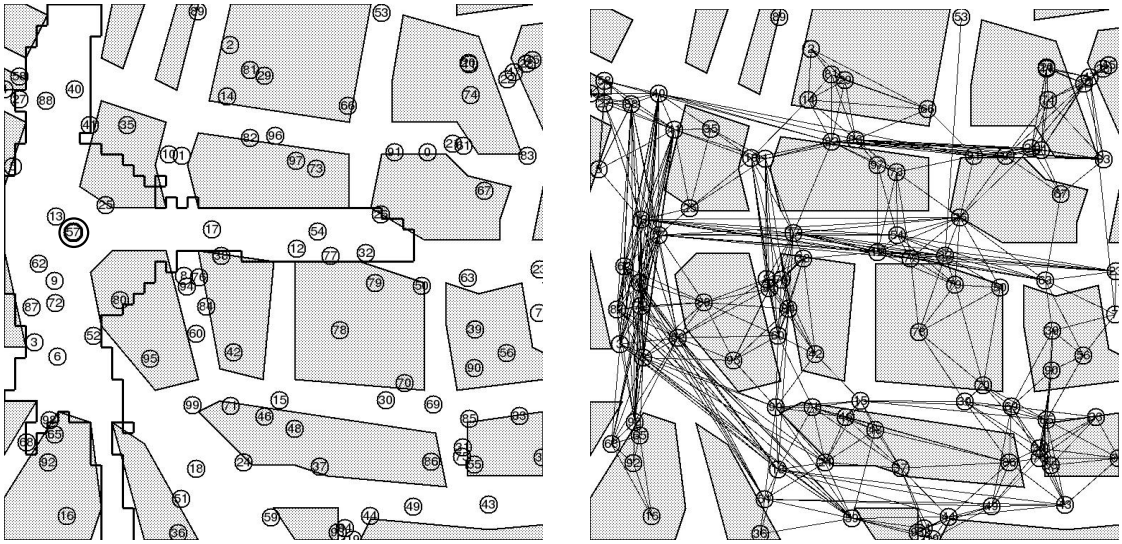

Figure 62: Munich scenario:
left: Transmission range of node 57; right: Connectivity plot of the network.

4.5.3 Urban Mobility Model

As described in section 4.3.2, only a few urban mobility models exist, but all show some serious shortcomings. On the other hand, several mobility models with different characteristics exist for flat environments. Section 2.5.1 presents the most important ones. These movement patterns work well for flat simulation areas, but show unrealistic behaviors in urban environments. Participants would not interfere with their surrounding environment. The model allows them to enter and leave buildings at any place and participants not try to sidestep obstacles. In order to generate characteristic node movements, the original City Motion (oCM) mobility model was introduced in [9]. It is based on the RWP mobility model presented in [89]. As described, the behavior of the RWP model is disadvantageous, and therefore [10] presents an improved version of the City Motion (CM) model, which is based on the RD mobility model. This implies that CM controlled nodes randomly select movement directions, velocities and movement times rather than destination points. As border behavior, the CM model uses the bounce back algorithm to allow uniform node distributions.

Figure 63 illustrates a possible movement path of a node. At point *A*, the random process stops the current movement and assigns a new mobility vector to the node. Thereafter the node starts moving towards this direction with its assigned velocity. In these cases, the CM model behaves equivalent to the RD model. However, it behaves different in case nodes approach obstacles. A random process decides with the probability p_{in} whether it bounces back (point *B*) or enters the building (point *C*). In case nodes bounce, they maintain their movement velocity and angle of arrival α, but change their directions. When entering buildings, velocities and movement directions stay constant. The combined movement time for an individual movement episode never changes because of e.g. bouncing events. If nodes are already inside of buildings and touch walls, they leave the building with a probability p_{out} or stay inside with a probability $1\text{-}p_{out}$. Besides the maintenance of an almost uniform node distribution within the simulation area, the bouncing characteristic at walls has another advantage. It allows nodes to travel through even very narrow street canyons. They are able to identify a path back to the main street and therefore are never trapped within streets. In order to force nodes to stay

mainly outside $p_{out} > p_{in}$ and $p_{in} < 0.5$ is recommendable. For all following simulation scenarios, the variables are set to $p_{in} = 0.2$ and $p_{out} = 0.8$. Therewith, nodes enter or stay within buildings with 20% probability.

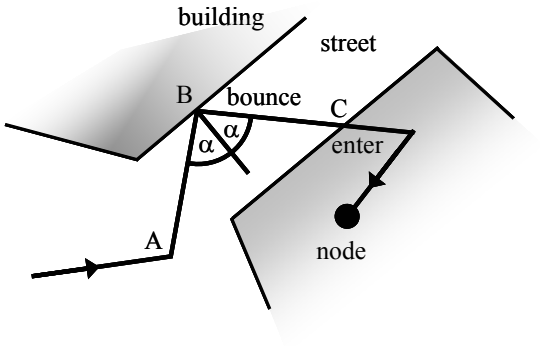

Figure 63: Behavior of CM controlled nodes at walls of buildings.

Figure 64 depicts a movement trace of an arbitrary node within the Munich city environment. Obviously, nodes show realistic movement behaviors. Therewith, CM is especially suitable as mobility model to test and classify ad hoc routing algorithms in urban environments. Whereas, for flat environments, CM mobility traces are equivalent to movement traces of the RD model.

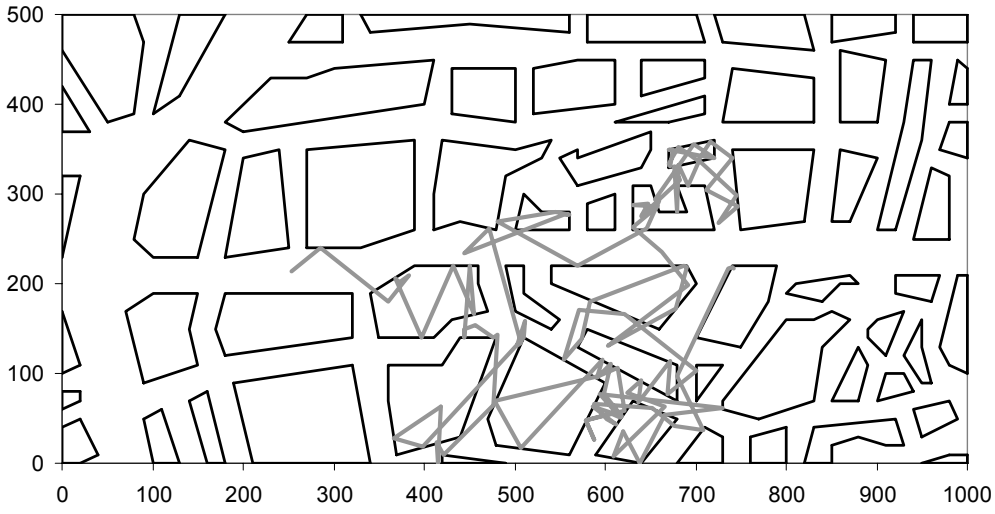

Figure 64: Exemplary node movement within the Munich scenario.

4.6 Simulations

The previous section described a realistic and suitable path loss prediction model and additionally presented a novel mobility model to create characteristic node movements in urban environments. Therefore, all necessary parts are introduced to accomplish reasonable ad hoc network evaluations under NLOS conditions.

4.6.1 Urban environments

To verify the impact various ground plans have on the performance of ad hoc routing protocols, simulations consider four different building deployments. The Manhattan scenario as proposed in [171] has a dense building deployment. Each building has an edge length of 80 m and a street width of 20 m separates every two buildings. As the name implies, this scenario imitates the regular block structure of New York City (see Figure 65 left). Due to its structure, connections favor paths in either horizontal or vertical directions. It leaves little space between buildings and therefore most connections require more hops than within a free space scenario.

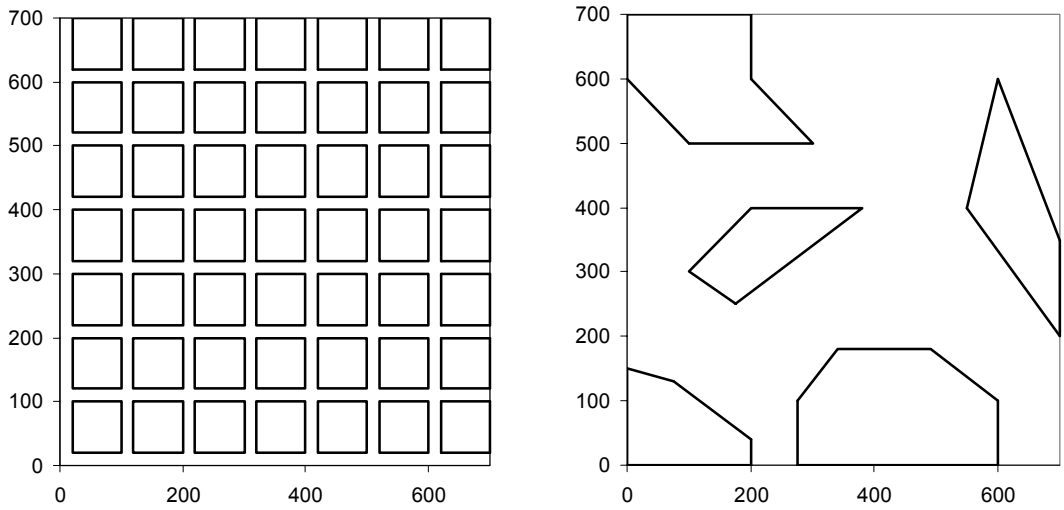

Figure 65: Deployment of buildings within the Manhattan and the Italy scenario.

The right picture of Figure 65 presents the Italy scenario. It has only a few and sparsely deployed buildings and therewith numerous connections allow LOS links. The scenario has an unstructured building deployment with large space between each. Connections should require fewer hops on average than e.g. within the Manhattan scenario. As a result, routing protocols should achieve a better performance within this scenario than in other scenarios. The third scenario exactly reproduces the pedestrian-only area of downtown Munich (see Figure 64). The scenario has numerous, patchy organized buildings with little space between each of them. In contrast to the Manhattan scenario, it does not favor any direction. Most connections have NLOS links, and therewith require many hops compared to free space scenarios. The last environment used, is the free space scenario. As the name implies, it does not contain buildings. All nodes are within LOS to each other and the transmission coverage area again approximately forms a disc around the transmitter.

The transmitting power P_t is determined so that the maximal radio transmission range is about 150 m in all scenarios. Depending on the transmission range, evaluations based on the free-space scenario are comparable with previous results from the literature.

Summarizing the properties of the above scenarios, the number of adjacent neighbors heavily depends on the building ground plan and on the current position of the transmitting node. The average node density is equivalent in all scenarios. Usually, one terminal is within each $50\,\text{m} \times 50\,\text{m}$ square. Therefore, the edge length l of the

simulation area depends on the number of nodes within the simulation. With the help of the number of nodes N, the edge length is calculated as $l = 50m \cdot \sqrt{N}$.

Nevertheless, the average node connectivity heavily depends on the building profile. Although the initial transmission power P_t is kept constant within all scenarios, the respective radio transmission range varies. While nodes within flat environments commonly have 28 neighbors, they have 17 neighbors in the Italy scenario. The Manhattan environment achieves an average node connectivity of 13 neighbors and the Munich scenario averagely permits only 11 neighbors with otherwise equal parameters.

4.6.2 Simulation parameters

Evaluations in chapter 3 and previous considerations in [78] depict that AODV outperforms other single-path ad hoc routing algorithms like DSR in highly dynamic scenarios. Preliminary simulations and evaluations in [9] and [10] confirm the superior performance of AODV in urban environments. Additionally, it is the only yet standardized ad hoc routing protocol. Therefore, the following evaluation solely utilizes AODV as routing algorithm and tests how it copes with urban environments. The main mobility model is the CM mobility model and nodes do not pause between consecutive movements. To minimize start-up effects, the initial start-up phase is set to 20s. In accordance to the description in appendix C, the program run is terminated whenever results show sufficient confidence levels. The results illustrated below present the average out of ten independent simulations with varying traffic and mobility traces. Additionally, the figures contain the respective 95% confidence intervals. Again following previous publications, the evaluation utilizes the existing WLAN 802.11b implementation of the ns-2 2.1b9a version with a maximal throughput of 11 MBit/s. As already mentioned, the maximal radio transmission range obviously varies with the environment, but LOS communication covers about 150 m. Additionally, the AODV-UU ns-2 extension [172] from the University of Uppsala is used, as it is fully compliant to the RFC. As described in section 4.5.2, a tile size of 10 m × 10 m is optimal. It allows the best tradeoff between simulation accuracy and memory constraints.

During simulation, 20 sources simultaneously generate packets with a payload of 1024 Byte. If not otherwise stated, simulations contain 100 nodes and each source generates ten packets per second. Data traffic flows are constant bit rate (CBR) flows, packets are unacknowledged, and therefore sources never retransmit lost packets. During simulation, source-destination pairs alternate randomly with an average holding time of 50 seconds and a five second cool-down phase between consecutive connections. Therefore, the overall generated traffic of all sources is about 1450 kBit/s, when assuming common scenario parameters. The overall MAC traffic is much greater, as data communication mostly require multi-hop connections and implies routing overhead. Nevertheless, it is noticeably below the maximal throughput of WLAN 802.11b, thus, the network is never saturated. Additionally, simulations use the CM mobility model with 2 m/s maximal node velocity as common parameters.

The simulations evaluate the performance of AODV within various scenarios. The evaluation is based on three different metrics, which are briefly described in the following.

- *Packet loss*:
  The packet loss p is the ratio between lost data packets and generated data packets. The packet loss ratio is presented in [%]. Packet losses occur because of queue overflows in intermediate nodes or collisions on the wireless medium.
- *Throughput*:
  The throughput T is the overall received data rate of all destinations. The metric counts all packets equally, independent from the necessary number of hops and the additional induced overhead. Obviously, the throughput and the packet loss correlate, because higher packet loss leads to lower throughput and vice versa.
- *Routing overhead*:
  Routing packets are packets solely used to setup and maintain network connections, they do not contain data. Consequently, the routing packet overhead O_P is the ratio between transmitted routing packets and all transmitted packets. Multiple forwardings count multiple times and the overhead O_P is in the range between zero and one.

4.6.3 Simulation results

Figure 66 depicts the impact of varying urban environments and network sizes on the performance of ad hoc routing algorithms. As expected, the packet loss is lowest within the free space scenario. Moreover, it perfectly scales with increasing node numbers N. The throughput results impressively illustrate that, because they remain almost constant and close to the optimum of 1450 kBit/s. In contrast to that, AODV has some difficulties with urban environments. All three scenarios cause significantly higher packet losses in comparison to the free space simulations. While the Italy scenario only causes 5% packet loss with 50 nodes and reaches 22% for 200 nodes, while the Munich scenario causes even higher loss rates in all cases.

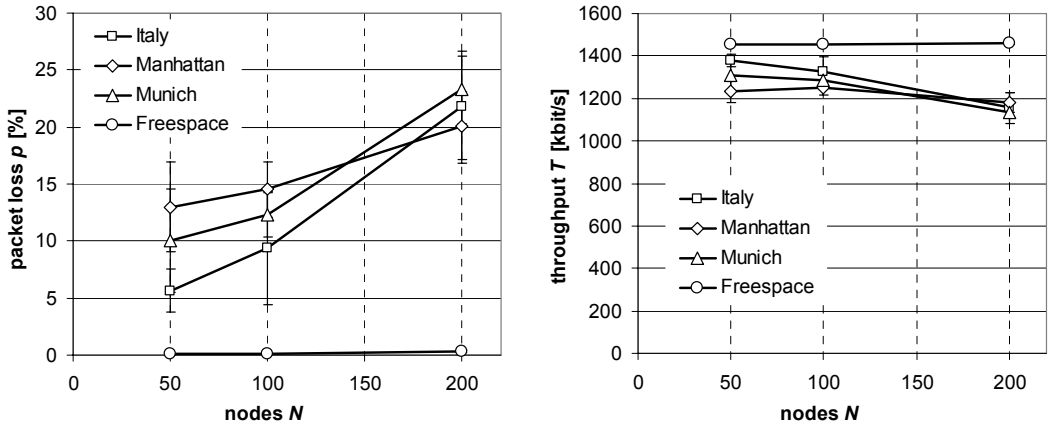

Figure 66: Packet loss and throughput for varying network sizes and scenarios.

The Manhattan scenario causes the highest loss within the 50 node scenario, while it achieves a smaller loss with 200 nodes. The increasing number of nodes does not cause a significant impact on the packet loss. The regular block structure of Manhattan allows radio transmissions along street canyons to span large distances. Therewith, it is irrelevant, whether the destination is just around the block, or a couple of blocks away. The average path length between two nodes does not alter considerably with the network

size and therefore minimizes the impact of increasing node numbers. The throughput graphs emphasize the accuracy of this evaluation. When considering 200 nodes in urban environments, AODV achieves the highest throughput within the Manhattan scenario.

In contrast to that, AODV causes an increasing packet loss above average within the Munich and Italy scenario. AODV is only able to maintain small networks with few hops between sender and receiver. With increasing network sizes, the hop distance between both communication endpoints increases as well. However, connections do not exist as long as in flat environments and sources must frequently reinitiate route requests. This in turn increases the routing overhead and congests the network, which raises the possibility of collisions on the physical medium. Consequently the network load increases again. As expected, ad hoc networking is much more difficult within urban environments. Routing protocols must cope with unfavorable conditions more often. The self-organized network approach comes closer to its limits or even passes them.

As mentioned in chapter 3, the limitation of the maximal length of routes seems promising. Although it prevents certain connections, it greatly improves overall network performance. The described introduction of a time-to-live field within route requests (RREQ) decreases the average occurring number of route errors per time. Therewith, the necessary number of route reestablishments is reduced as well and minimizes the additionally induced network load. Relaying nodes can focus more on data forwarding rather than RREQ broadcasting, and consequently the remaining connections within the network achieve higher throughputs.

The packet rate depicts the frequency of generated packets of each individual source. Obviously, it varies the overall induced data traffic load. Figure 67 illustrates its impact on the network performance. With a twice as high packet rate in comparison to the common parameters, the overall generated data traffic raises to about 2900 kBit/s. Comparable to the previous free space evaluations, AODV is able to keep the packet loss close to zero for all kinds of network loads. Consequently, the throughput rises almost linearly. In contrast to free space scenarios, AODV experiences difficulties to keep the loss rate minimal in all other scenarios. The loss rate within all these scenarios is worsened by a factor of 2.5 when doubling the packet rate from 10 to 20 packets per second. Obviously, the network is overloaded and the loss rate increases above average. Among the urban environments, the Italy scenario achieves lowest packet losses with all kinds of network loads, whereas AODV performs worst in the Manhattan environment. However, with 20 data packets per second, the achievable throughout of AODV already reaches saturation. With even higher packet rates, the throughput does not increase anymore, only the packet loss ratio asymptotically approaches 100%.

With these environmental conditions, reasonable networking is only possible with packet generation rates of ten packets per second or less. The other possibility to reduce the overall network load is the limitation of the number of sources. However, this would require the introduction of some type of call admission control, which obviously violates the ad hoc paradigm of an unrestricted and fair network access.

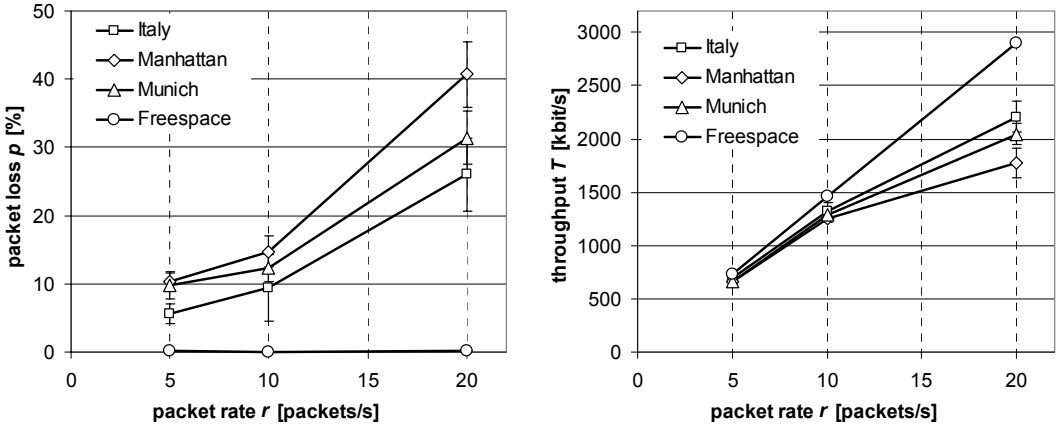

Figure 67: Packet loss and throughput for varying network loads and scenarios.

As expected, the underlying mobility model and the maximal node velocity significantly impact the performance of ADOV. Figure 68 illustrates the packet loss rates of AODV with respect to the maximal node velocity. It presents results for scenarios with 100 nodes within the Manhattan and Munich scenario. Additionally, the figure distinguishes between the random direction (RD) and the city motion (CM) mobility model. It is obvious that the packet loss rises with increasing maximal node velocities due to an increasing network dynamic. But it is unexpected that the CM model has packet loss rates well below the values for the RD model. This behavior is independent from other parameters.

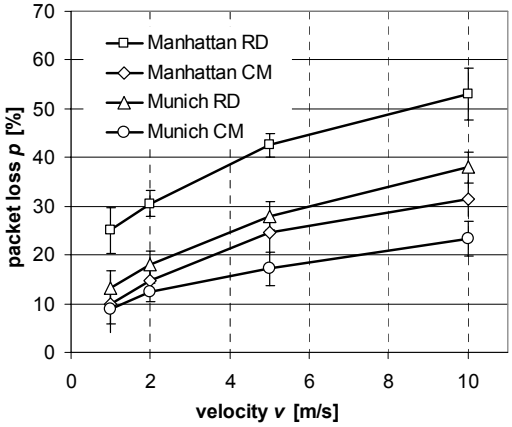

Figure 68: Impact of the node velocity and the mobility model on the network performance of AODV.

A detailed analysis shows that packet loss often occurs in case nodes move inside of buildings. This significantly alters the propagation loss between transmitter and receiver. Routes frequently break during these events and thereafter require new route setups. This generates additional routing overhead and network load, and consequently worsens the overall performance. As the CM model prevents frequent transitions between positions outside and positions inside of buildings, it achieves better results than the RD model.

For maximal node velocities beyond 5 m/s, the packet loss is above 20%, even for CM as mobility model. With those loss rates, reasonable communication turns impossible. This again limits the usability of MANETs within urban environments to scenarios with a pedestrian like average node mobility.

As final evaluation, Figure 69 depicts the impact of varying network sizes and packet rates on the induced overhead O_P. Obviously, with increasing network sizes the induced overhead increases as well. The routing overhead rises proportional to the average route length, and the route length depends on the network size. Therefore, the overhead rises as well linearly with the number of nodes. The behavior of the AODV packet overhead within the free space scenario impressively emphasizes this characteristic. The overhead of AODV within the three urban environments is significantly higher. With only 50 nodes, the Italy scenario causes the fewest overhead, while it is highest for 200 nodes. Both other scenarios induce almost equivalent overheads. The induced overhead is already beyond a certain limit. Due to the large number of routing packets, the wireless medium is overloaded, and numerous collisions occur. Nodes try to solve the congestion by delaying packet transmissions, which causes packet drops in queues. In case routing packets get lost, the routing algorithms do not receive vital information and may choose unfavorable routes, which again worsens the performance.

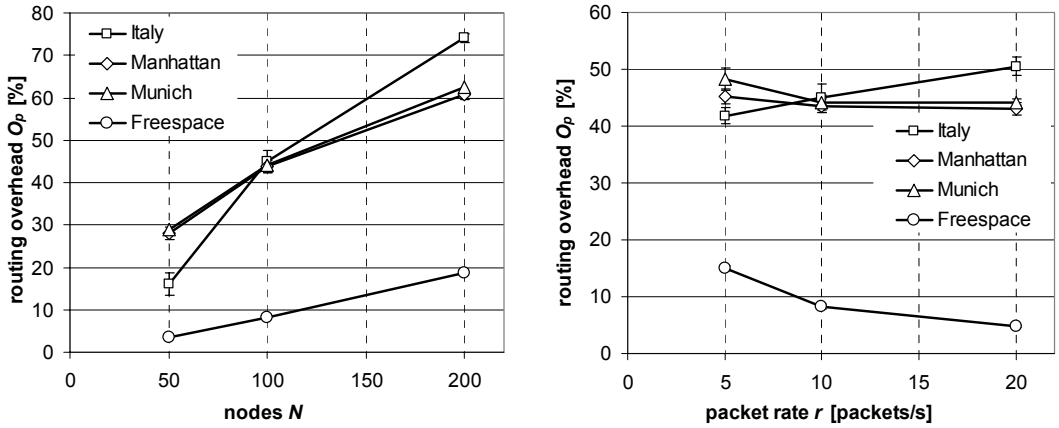

Figure 69: Impact of network size and packet rate on the AODV routing overhead O_P.

The routing overhead O_P is the ratio between routing packets and all transmitted packets while the number of routing packets per second is unaffected by the data packet rate. Therefore, O_P must be inverse proportional to the packet rate, when neglecting all other influences. Because AODV is easily able to achieve loss rates close to zero within the free space scenario, its overhead exactly pictures this behavior. The AODV routing overhead shows a different performance in urban environments. Instead of a decreasing routing packet ratio, it remains constant or even increases with increasing packet rates. To explain this behavior, previous observations are necessary. As already mentioned, networks in urban environments are usually fully overloaded at high packet rates. Figure 67 illustrates that the packet loss rises above average, while the number of received data packets (the throughput) does not increase with the same proportionality as the packet rate. Consequently, the number of routing packets must increase, so that the maximal achievable network capacity is again fully occupied. The number of generated data packets and the number of routing packets rises, hence, the ratio O_P between both remains constant in urban environments.

4.7 Conclusion

AODV is able to perform perfectly well within free space scenarios. It scales independent from the network size and the network load. It causes only marginal packet losses, while the throughput is optimal and close to the achievable maximal. However, AODV encounters difficulties to perform its tasks in NLOS urban environments. Within large networks, the packet loss is significantly higher, while the throughput degrades. Additionally, the achieved throughput in urban scenarios declines in comparison to the free space scenario for high packet loads. Even higher packet rates do not increase the throughput any further, only the packet loss asymptotically approaches 100%. As a result, the urban scenario significantly reduces the threshold, after which a network is overloaded.

To circumvent the negative impacts of overloaded networks, certain extensions are possible. Network layer protocols could limit the maximal packet rate. However, this would also limit the maximal throughput within free space scenarios. Consequently, the network layer protocol must consider the environment to determine the maximal throughput, which is difficult. Another possibility is the introduction of a jointly used call admission control. However, this requires a distributed knowledge about all flows within the network and violates the MANET paradigm of unlimited and free network access.

As described in chapter 3, the usage of routes with numerous hops does not allow reasonable communication between network participants. The results in this chapter confirm that this is especially true for urban environments with frequently changing network topologies. The limitation of route length with a time-to-live field circumvents the shortcomings, but again unnecessarily limits the usage of ad hoc networks within non-urban environments. The utilization of the encountered packet loss as indicator for the quality of a route might be possible. In case the packet loss is beyond a certain threshold, packet destinations independently limit their connections in order to allow concurrent network participants to experience an improved network service. However, such a threshold certainly depends on the network conditions and environments. It would be vital and fundamental for the overall network performance, and therefore must be set carefully. The determination of the correct threshold for arbitrary network conditions would be an open research issue.

Increasing maximal node velocities have a negative impact on the packet loss and the achievable throughput. However, as depicted in the previous section, urban scenarios worsen this relation. They limit node velocities to pedestrian speeds. Even the otherwise suitable statement that an increasing radio transmission range can cope with higher node velocities is not valid anymore. The path loss through buildings or around them limits the radio range so rigorously that an increasing transmission power does not significantly enlarge the radio coverage. Hence, network layer protocols must again limit the number of hops or identify static or slow moving nodes as favorable connection relays.

As a conclusion, urban scenarios dramatically alter the network conditions and certainly reduce the usability of ad hoc networks. Consequently, reasonable network connections require shorter paths and must limit their maximal packet rate. However, even with those limitations, MANETs within urban environments still allow reasonable data exchange between distant nodes and communication remains possible.

5 MAC Layer Extension Providing Fair Throughput

After the consideration of urban environments, and the determination of their impact on the performance of ad hoc networks, the following chapter focuses on the fairness within networks as another possible source for user dissatisfaction.

Following the ISO/OSI reference model, the task to establish favorable paths through networks is within the responsibility of routing protocols. In contrast to that, the medium access control (MAC) protocol has the obligation for direct node-to-node packet transmissions. Therefore, the quality of the MAC protocol also has severe impact on the overall network performance.

Within distributed ad hoc environments, the network load is evenly spread. All nodes statistically uniformly act as source, destination or as forwarding node for connections. However, scenarios without fixed infrastructure are uncommon in future considerations. It is unlikely that access points (AP) with gateway functionality to the Internet are unavailable. Scenarios with mobile networks grouped around an AP are much more probable. The gateway allows connection to the Internet, while the MANET enables distant nodes to communicate with this gateway. Possible scenarios are airports, train stations, university campuses, or downtown areas. Thus, single APs can provide Internet access to tens or even hundreds of mobile nodes.

Networks containing an AP as in the right illustration of Figure 70 show radial symmetric load distributions and increasing loads towards the central AP. Therefore, algorithms performing well in pure ad hoc networks may fail within these networks. The distributed coordination function (DCF) of the IEEE WLAN 802.11 [173] standard suite only considers pure ad hoc networks. It allows balanced per-node fairness, but is unable to adapt to networks with different load distributions. The round robin medium access is especially suitable for distributed network loads. The point coordination function (PCF) is an extension of the DCF and permits an AP to control the network. However, it solely supports single hop connections and therefore is inappropriate for multi-hop ad hoc networks around APs.

However, WLAN standards as well as routing algorithm developments put emphasis on distributed multi-hop ad hoc networks without any infrastructure. Up to now, research did not consider different environments. Therefore, it is necessary to investigate the behavior of IEEE WLAN 802.11 in networks with central APs and optimize its performance. Simulations illustrate that overall network performance is insufficient within these scenarios. The new Fair-MAC protocol extension developed in this thesis in cooperation with existing routing algorithms outperforms the above mentioned combination of legacy WLAN and routing algorithm.

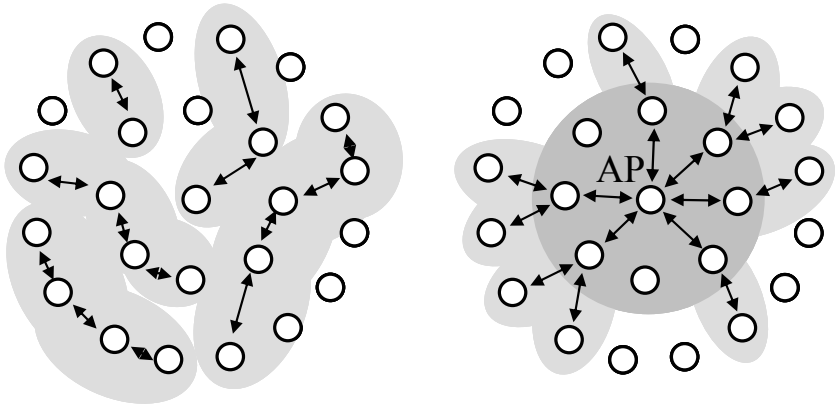

Figure 70: Load distribution within a distributed ad hoc network and a network with central gateway.

Section 5.1 gives a general introduction to the 802.11 standard. It generally introduces the distributed shared medium access scheme. The following section 5.2 describes the shortcomings of existing protocols to perform well within a network with unevenly spread load distribution. It identifies two independent reasons for the observed performance degradations. Some known approaches from the literature and their shortcomings are discussed in section 5.3. Section 5.4 depicts approaches to circumvent both reasons of the performance degradations. A novel fair-MAC protocol extension is introduced in section 5.5. It combines the above mentioned solutions and allows perfect flow fairness within radial symmetric networks. Section 5.6 presents simulation results in order to verify the ability of the proposed protocol to perform well under LOS and even NLOS conditions. The chapter sums up with a conclusion in section 5.7.

5.1 Introduction

The wireless local area network (WLAN) standard IEEE 802.11 became the most widely used protocol for wireless connections between mobile terminals. Most sold notebooks as well as an increasing number of desktop PCs have an in-built WLAN protocol stack. The IEEE as standardization organization continuously increases the achievable data rate. The first standard version was released 1996 [42] and allowed a maximal bit rate of 2 MBit/s. Since then, the bit rate increased to 11 MBit/s [174] and currently allows 54 MBit/s with the most recently released versions 802.11a/g. The next substandard 802.11n should achieve bit rates beyond 100 MBit/s [175, 176]. This maximal bit rate is available only for a single hop connection with exclusive medium access. The rate for each node degrades, in case APs serve several different connections or data exchange requires multihop transmissions. However, the usage of multihop connections provides the possibility to increase the serving area of an AP, without the need to establish additional APs with gateway functionality. A decentralized ad hoc network utilizes nodes in the proximity of the AP for data relaying of more distant sources. It presents a cost-saving probability to connect a larger number of users to a single AP.

As described in section 2.3, the other technique to wirelessly connect nodes to an AP is HiperLAN/2 [177-179]. However, the current HiperLAN/2 standard does not include a protocol for a decentralized medium access. The literature proposes several extensions to

allow distributed ad hoc networking with HiperLAN/2 [180, 181], but a standard is not yet defined. Therefore, and because 802.11 has a significantly broader installed basis, the following considerations focus on the shortcomings of 802.11.

Section 2.3 only gave a brief introduction into the functionalities of IEEE 802.11 and therefore the following section introduces it in more detail. The original standard was released in 1996 and used direct sequence spread spectrum (DSSS) or frequency hopping spread spectrum (FHSS) as access scheme to achieve a bit rate of up to 2 MBit/s. Besides the physical layer, the standard defined the MAC protocol as well. It contains the mandatory DCF function [173] as well as the optional point-coordination function (PCF). Nodes using the DCF as access mechanism independently compete for the carrier access. Detailed analysis and simulative results can be found e.g. in [182-184].

In contrast to the DCF, the PCF allows an AP to control its proximity. The AP divides timeframes in two parts. The first allows a synchronized access; the second uses again the distributed DCF function. The ratio between both parts is variable and the AP determines it. While the PCF is active, every node registers its transmission attempt at the AP, and the AP grants exclusive medium access for every demanding station in a round robin fashion. However, as the DCF function is especially designed for distributed multi-hop ad hoc networks, it will be discussed further.

Within uncontrolled networks, nodes try to access the wireless carrier simultaneously. The MAC protocol on each node must prevent these uncoordinated accesses. Otherwise data loss occurs, because terminals receive interfered and corrupted signals from multiple senders. Therefore, the DCF function uses the carrier sense multiple access with collision avoidance (CSMA/CA) [185] as access scheme. It is based on the previously invented CSMA with collision detection (CD) method [186], which is mainly utilized in Ethernet 802.3 networks. With only a single antenna and transceiver, nodes cannot detect simultaneous transmissions during ongoing own sending attempts. Hence, the WLAN standard must use the collision avoidance method to reduce the necessary number of retransmissions. Figure 71 depicts the details of the CSMA/CA protocol.

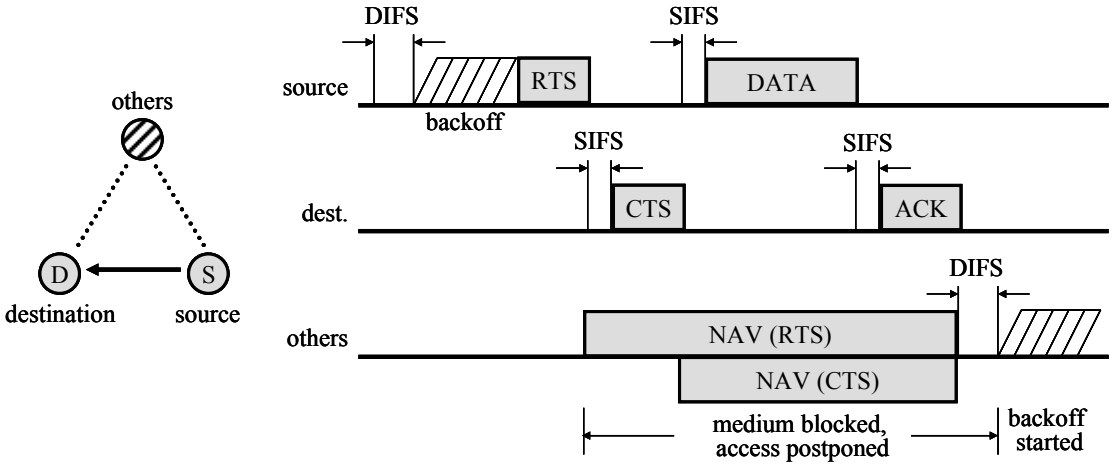

Figure 71: CSMA/CA with RTS/CTS of the 802.11 MAC protocol.

Before a node activates its antenna for a sending operation, it senses the carrier. The sensing period contains the constant DCF interframe space (DIFS) and an additional backoff period. Every node randomly calculates its own backoff period. The contention

window (CW) defines the maximal possible backoff. This random delay reduces the probability that two neighboring nodes start transmitting at the same time. Whenever a packet collision occurs, both transmitting nodes double their CW to reduce the possibility of recurring errors. In case the medium is unoccupied till the backoff timer expires, the node transmits its data packet immediately. If the node overhears transmissions from other nodes while the backoff timer is running, the protocol postpones the timer expiration and therewith the transmission.

Transmitting nodes expect an acknowledgement (ACK) from the receiver after the error-free transmission of the data packet. In order to detain other nodes from starting transmissions between the DATA and ACK packet, the receiving node waits only a short interframe space (SIFS) period, which is shorter than the DIFS period.

In a CSMA/CA system occur two primary phenomena that reduce the achievable overall throughput. Both of these exist due to the fact that unlike wire-line networks, terminals in a region may not see the same state of the medium. Only if all stations are within each others radio range, the CSMA/CA method operates error-free. The first problem happens when a node senses the carrier as unoccupied and starts its transmission, while the receiver already gets a packet from a different station. Usually, both sending stations are on opposite sides of the receiving node and not within each others radio range. The literature refers to this as hidden terminal problem [187].

To circumvent the hidden terminal, the 802.11 standard optionally allows the usage of a ready-to-send/clear-to-send (RTS/CTS) method [188, 189]. Before the sending node forwards its data, it transmits an RTS packet. The RTS packet contains the proposed receiver of the data packet as well as the necessary transmission period. All nodes within the proximity overhear this packet and calculate the network allocation vector (NAV). With the help of the NAV vector, uninvolved nodes postpone all their scheduled transmissions. The proposed receiver answers with the CTS packet. Neighboring nodes overhear the packet and adapt their NAV vector again. Thereafter, the sending node transmits its data. Figure 71 depicts all necessary packet transmissions for a complete sequence in a timely manner.

The usage of RTS/CTS packets is especially useful for large data packets. The additional overhead induced by the RTS/CTS packets is acceptable in comparison to the high possibility of packet errors and necessary following retransmission. While the RTS/CTS mechanism prevents the occurrence of hidden terminals, it introduces the second dilemma. The literature refers to it as exposed node problem [190]. As all neighboring nodes of the sender as well as of the receiver are blocked, they are unable to transmit data, even if the network configuration would allow an error-free transmission. Figure 72 depicts the transmission sequence as well as the blocking status. However, the RTS/CTS still improves the overall performance in distributed networks compared to protocols without a hidden terminal avoidance algorithm.

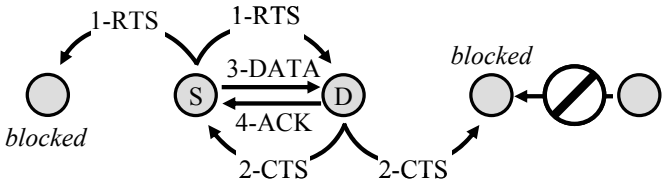

Figure 72: RTS-CTS-DATA-ACK mechanism of WLAN 802.11.

5.1.1 The 802.11 standard suite

The 802.11 standard does not define a single protocol, but comprises several different parts. Besides protocols achieving superior data rates it contains improvements for higher ISO/OSI layers and regulatory issues as well. Parts of the standard are differentiated by subsequent characters. The IEEE 802.11 working group still develops new sub-standards. The following list should give a short overview over the currently existing standards and their scope. It does not consider currently discussed draft versions of upcoming parts. Figure 73 presents existing sub-standards and contains a classification into the ISO/OSI reference model.

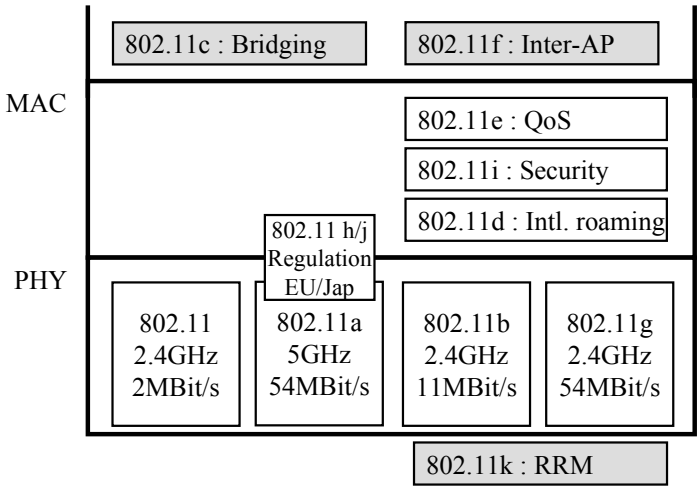

Figure 73: Overview over the 802.11 standard suite.

Physical layer (PHY)

- *802.11 legacy*: It is the original standard and achieves up to 2 MBit/s in the 2.4 GHz Industrial-Science-Medical (ISM) band.
- *802.11a:* This high speed WLAN standard for the 5 GHz ISM band utilizes the orthogonal frequency division multiplexing (OFDM) access scheme to achieve gross bit rates of up to 54 MBit/s.
- *802.11b*: Improves the DSSS modulation scheme and allows a maximal bit rate of 11 MBit/s with complete downward compatibility to the 802.11 legacy standard. It is the most widely utilized and supported standard.
- *802.11g*: Extends the achievable bit rate within the 2.4 GHz band to 54 MBit/s via the usage of OFDM carriers. It generally has a broader application area in comparison to 802.11a, because the lower frequency band allows better propagation characteristics.
- *802.11k*: This extension enables higher layers to perform radio and resource measurements (RRM).

Medium access control layer (MAC)

- *802.11d:* This standard is responsible for the compliance of 802.11 protocols with regulatory specifications. It mainly limits the maximal transmission power in certain countries to achieve this compliance. It allows a worldwide usage of 802.11 hardware.

- *802.11e:* It extends the 802.11 MAC protocol to allow networks to forward traffic of real-time critical applications. It uses multiple MAC queues with different priorities to achieve certain quality of service (QoS) requirements.
- *802.11i:* It closes the existing security loopholes of the originally developed wired equivalent privacy (WEP) protocol [191]. It circumvents the shortcomings of the authentication and allows the usage of Wi-Fi protected access (WPA). Additionally it substitutes former encryption methods with 802.1x functionality, the temporal key integrity protocol (TKIP) and the advanced encryption standard (AES).
- *802.11h:* It defines the spectrum management to use 802.11 compliant hardware within the 5 GHz band in Europe.
- *802.11j:* Defines a similar spectrum management for the Asia-Pacific region.

Higher layers

- *802.11c*: This substandard allows APs to retrieve information about bridging operations. Correct bridging is necessary to extend the AP with Internet gateway functionalities.
- *802.11f*: The Inter-AP-to-AP protocol defines the necessary communication between neighboring APs. Therewith, it allows a simple AP-to-AP handover of mobile nodes.

5.1.2 The 802.11e QoS extension

Voice over IP (VoIP) like Skype [192] or streaming are new real time critical applications. They require more challenging QoS parameters than existing best effort applications like web browsing or file downloads. Therefore, the IEEE defined the QoS extension 802.11e [193, 194]. It allows the definition of various priority classes to achieve certain QoS levels.

The standard modifies the legacy PCF function to become the hybrid coordination function (HCF) and the DCF to the enhanced DCF (EDCF). Again, only the usage of the EDCF is valuable in distributed MANETs. It introduces eight priority classes for various traffic types. Each class utilizes its own protocol instance with MAC queue and backoff parameters. All eight protocol instances compete for exclusive medium access. The backoff parameters are independently modifiable and allow different possibilities for medium access. The delay period before a protocol instance gets exclusive access comprises the random backoff as well as the new arbitration interframe space (AIFS). Dependend on the traffic classes (TC), the AIFS varies, but it is as least as large as the original DIFS. High priority packets wait short AIFS periods, while best effort traffic packets experience longer AIFS delays. Figure 74 depicts the competing protocol instances and illustrates the different AIFS times after a successful packet transmission. Additionally, the 802.11e standard modifies the initial minimal CW values with respect to the priority classes, while the binary exponential backoff behavior is unchanged. Therefore, 802.11e hardware is fully backward compatible with previous standards and networks with a mixed usage of protocol versions are possible.

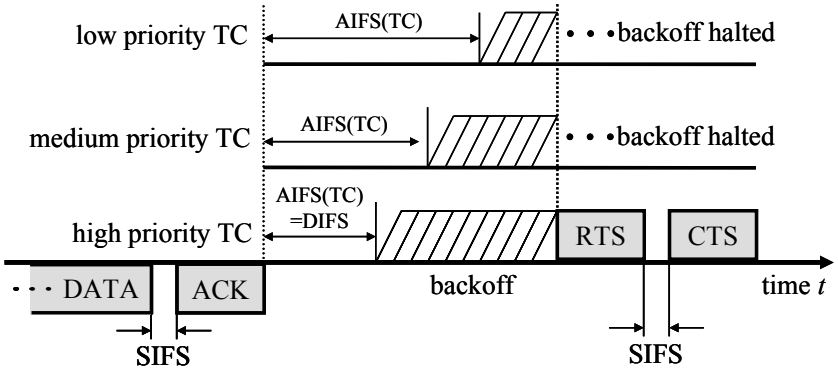

Figure 74: 802.11e enables different priorities with the help of variable medium access timers (from [194]).

The protocol instances work independently from other instances, and therefore behave equivalent to protocol stacks within two competing terminals. In case the protocol grants a priority instance exclusive access, all other instances postpone their transmission attempts. A subsequent scheduler prevents internal collisions, in case two instances start their transmission at the exact same time. The 802.11e protocol stack obtains the packet priority from the yet unused type-of-service (ToS) field within the IP header. According to their priority, the protocol enqueues packets in certain queues. Therefore, the extension forwards packets with high priority more often than those packets with best-effort characteristics. It usually achieves the expected QoS requirements without providing hard guarantees.

The 802.11e extension was mainly developed to improve typical WLAN scenarios, in which all nodes are within the transmission range of an AP. For these scenarios, 802.11e enabled terminals show significant improved performance characteristics in comparison to legacy 802.11 terminals. It outperforms the original standard especially with respect to delay and jitter constraints [195, 196]. Unfortunately, the literature does not discuss the performance impact on distributed MANETs up to now. Theoretical or simulative results are not available yet. However, the protocol extension does not address the fairness among nodes and additionally it is fully backward compliant to the original 802.11. Therefore, it is also not able to perform well in networks with non-uniform load distributions.

5.1.3 Ad hoc networks with central APs

The PCF function enables an AP to control the network, but it supports only single hop connections. Therefore, it is inappropriate for multi-hop ad hoc networks around an AP. In contrast to that, the DCF function considers pure ad hoc networks, but scenarios without any fixed infrastructure are unusual. It is much more likely that mobile nodes are grouped around an AP. The combination of an AP with a distributed MANET allows the connection of all participating nodes to the Internet. Possible scenarios are airports, train stations, university campuses, or downtown areas. The advantage is that the AP serves not only its close proximity as Internet gateway, but also nodes several hops away. Thus, single APs can provide Internet access to tens or even hundreds of mobile nodes. Therefore, connections to APs within these scenarios usually require multi-hop routes.

However, the ad hoc network paradigm expects network configuration without the need of central entities. This includes the maintenance of frequently changing network topologies because of node movements or terminal switch-offs. Therefore, the medium access and routing must be controlled in a decentralized way. The AP is unable to obtain all information about the network wide radio resource management.

The routing protocol can be any existing or future ad hoc protocol. The determination of the optimal routing protocol depends on the proposed scenario. As presented in section 2.4, proactive protocols like DSDV or OLSR are advantageous in small and stable network environments. On the other hand, reactive protocols like AODV or DSR are most useful in larger and more dynamic networks.

Although the physical topology is uncoordinated, the solely reasonable logical topology for the network around the AP is a tree structure with the AP as root node. The disadvantage of this topology is that interior nodes must forward all traffic from and to exterior nodes, and therewith usually have much higher loads to handle than exterior nodes (see Figure 70). Consequently, the network load has a radial symmetric distribution. However, WLAN standards as well as routing algorithm development put emphasize on distributed multi-hop ad hoc networks without any infrastructure. Simulations illustrate that the overall network performance is insufficient within networks with central APs utilizing WLAN 802.11. It is necessary to investigate the behavior of WLAN within these scenarios, and to optimize its performance. The following section indicates shortcomings of WLAN 802.11, especially occurring in radial symmetric networks.

5.2 Shortcomings of the conventional WLAN 802.11 standard

As presented, the DCF function of the IEEE 802.11 legacy protocol contains a distributed shared medium access method for MANETs. However, in case data flows mostly traverse to and from an AP, the overall performance of the network and the fairness between different sources degrades significantly with 802.11. Based on these observations, the following section presents and illustrates the shortcomings of the DCF function. The performance degradations result from shortcomings of the packet queuing mechanism within the link layer and from an unbalanced medium access scheme. The theoretical analysis assumes any of the currently used WLAN standard 802.11a/b/g as MAC protocol, but also considers the QoS extension 802.11e for certain cases.

5.2.1 Shortcomings of the queuing algorithm

As described in section 5.1.1, the WLAN 802.11 standard follows the ISO/OSI reference model. The logical link layer control (LLC) is on top of the MAC layer. Between both layers resides a forwarding queue (MAC queue) to cache packets before they are transmitted. In case the MAC protocol successfully transmits a packet, it dequeues a new packet. The queue within the MAC layer certainly has a limited size, and therefore drops packets whenever the queue is overflowing. The dropping algorithm is not standardized, and can follow any scheme like drop tail [197] or random early detection (RED) [198]. More sophisticated algorithms may prioritize important routing packets by placing them on top of the queue. All ad hoc scenarios require that intermediate nodes act as relays

and forward packets from other sources towards distant destinations. Obviously, MAC queues must cache these packets and therewith contain forwarding packets as well as packets from their own applications. Both types of packets are stored within the queues in a "first-come-first-serve" order.

In the following, a simple example illustrates the unfairness of the existing queuing algorithm. Figure 75 depicts a small ad hoc network with sources and an AP. The exterior node creates packets with a data rate D_2 of 50% of the maximal access rate while the interior node has a variable data rate.

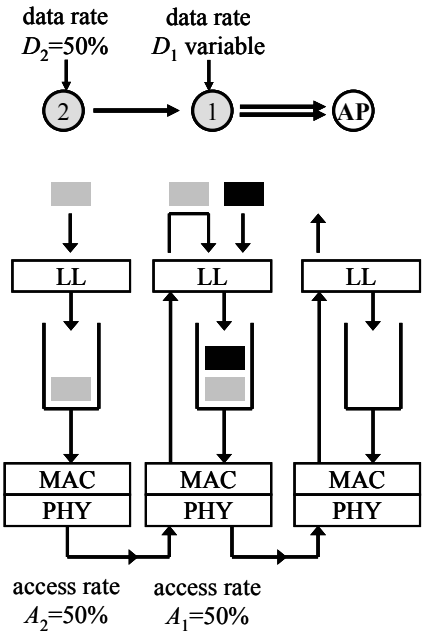

Figure 75: Network consisting of two nodes, illustrating MAC 802.11 queue unfairness.

Node n_1 and n_2 simultaneously transmit packets to the AP. Because only these two nodes transmit packets and both nodes have equal probabilities to access the channel, the available medium access time is fairly divided. The very balanced per-node fairness of the MAC DCF function prevents unequal access rates between competing nodes. Obviously, the medium access rate of both nodes is 50% of the theoretical maximal throughput (TMT). As described, the source data rate of node n_2 is limited to 50% of the TMT. This certainly prevents packet drops within the MAC queue Q_2 of node n_2. A detailed description of the TMT is given in [199].

Figure 76 illustrates the queue occupancy of node n_1 for various data rates D_1. The queue occupancy depicts the relative share of packets from n_1 within its own queue. The figure contains the results from theoretical analysis, simulation, and the fair queue occupancy under the assumption that the queue contains sufficient but not an infinite number of slots. Without packets from node n_1 ($D_1 = 0$) the networks operates with the optimal packet throughput. Node n_1 receives packets 50% of the time, and with the remaining 50%, it forwards these packets to the AP. The queue in n_1 never drops packets. However, with even a small source data rate D_1, the combined data rate of both nodes exceeds the available TMT. Therefore, node n_1 contains an always-filled queue Q_1 and in addition, the queue must drop packets. While this example seems quite constructed,

with more sources and longer forwarding paths, queue overflows occur even at low data rates.

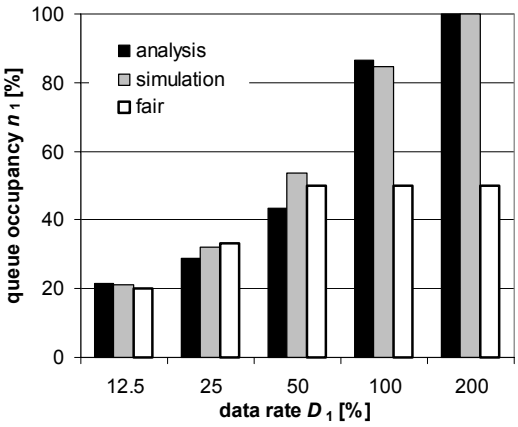

Figure 76: Ratio of packets from source node n_1 within queue Q_1 for different source data rates D_1.

Therefore, (93) depicts the relative shares R_1^{fair} of packets from n_1 within Q_1 in order to achieve fair packet throughput for different data rates D_1 and D_2. Figure 77 visualizes the different queue filling rates for a fair ratio of packets R_1^{fair}. This does not prevent queue overflows, but achieves fair queue fill ratios and therewith an equal probability that all flows experience comparable packet loss rates.

$$R_1^{fair} = \begin{cases} \dfrac{D_1}{D_1 + D_2} & for\ D_1 + D_2 < 1 \\ 1 - D_2 & for\ D_1 + D_2 \geq 1 \cap D_2 < 0.5 \\ D_1 & for\ D_1 + D_2 \geq 1 \cap D_1 < 0.5 \\ 0.5 & otherwise \end{cases} \tag{93}$$

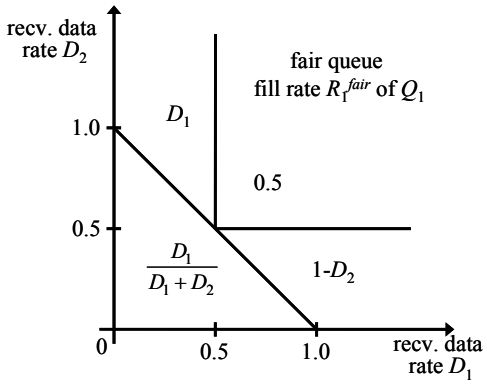

Figure 77: Fair queue fill rate R_1^{fair} of packets from n_1 within Q_1.

Following the formula in (93), the queue occupancy is fair as long as the data rate of node n_1 is below 50%. However, with increasing data rates, the occupancy becomes unfair. Above data rates D_1 of 50%, Q_1 should be equally filled with packets from both nodes. However, packets from n_1 prevail. If node n_1 requests exclusive utilization of the access rate (100%), it stops communication of exterior nodes to the AP. The queue

contains only 17% of packets of node n_2. A data rate twice as large as the access rate is obviously without practical usage. It causes considerable packet drops of own packets within Q_1. However, WLAN 802.11 does not prevent this behavior. Allowing the network layer to enqueue packets at such high rates completely suppresses any kind of data forwarding of exterior nodes. The queue Q_1 does not contain any packets of subsequent nodes. Hence, malicious nodes close to APs are capable to prevent communication by simply generating infinite numbers of UDP packets. The example clearly depicts the ability of an uncooperative or malicious node n_1 to prevent any exterior node to connect to the AP.

To determine the queue occupancy theoretically, we assume constant data rates. Further more, the queue of n_1 is already filled. All slots contain packets and the enqueuing algorithm drops every subsequent packet. Within the example environment above, the queue reaches this state after a short period, depending on the size of the queue. Thereafter, the algorithm only enqueues new packets, after it had dequeued a packet. Therefore, node n_2 and the upper layer of n_1 compete for this currently unoccupied slot. Before the MAC algorithm of node n_2 is able to forward a packet, it must probe the medium, probably requires some back off, and must cope with collisions. Therefore, the forwarding of packets requires significantly more time than the basic physical transmission. In fact, the transmission time is the least significant part of the forwarding time. Average forwarding times between two nodes for various data rates are given in [199].

The network layer of node n_1 generates packets with a constant rate. After the MAC algorithm dequeued a packet, the network layer statistically requires half this generation period before requesting to enqueue its packet. With known data rates, the average queue arrival times t^q for packets from both nodes can be calculated. The probability that the algorithm enqueues a packet from its own network layer rather than a forwarded packet from n_2 is the queue fill rate R_1. It is the quotient between ${t_1}^q$ and ${t_2}^q$ with

$$R_1 = \frac{t_1^q}{t_2^q} \tag{94}$$

Figure 78 shows a schematic example for the queue filling probability. The legacy 802.11b standard achieves 11 MBit/s gross data rate and a net-rate or TMT of 6.5 MBit/s. Assuming packet sizes of 1024 Byte and a designated data rate D_1 of 100% of the TMT, the upper layer protocol of n_1 generates every ${t_1}^q = 5.2$ ms a new packet and requests enqueuing within Q_1. With the results from [199], n_2 is able to transmit every ${t_2}^q = 6$ ms a packet towards n_1. The relative queue fill rate R_1 is 5.2 ms / 6.0 ms and therewith 86.6%. Packets from n_2 are enqueued with a probability of 13.4%. This obviously confirms the additionally obtained simulation results. It prevents a fair allocation of the available resources between n_1 and n_2.

As explained in section 5.1.2, the QoS extension 802.11e augments the basic DCF algorithm with additional queues for different priorities. These queues improve the fairness among packets with different priorities. For packets of the same priority class, the unfairness problem remains unsolved. All packets are again cached in the same way as with the single queue of the legacy 802.11 protocol.

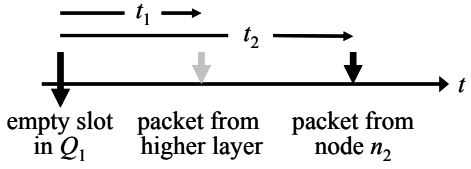

Figure 78: Schematic example for the queue filling probability.

This clearly depicts that high traffic load scenarios causes forwarding queues of intermediate nodes to drop packets and the common DCF/EDCF algorithms to fill up queues unfairly. The existing MAC protocols favor nodes with direct links to APs. The number of own packets often exceed the number of relayed packets and therefore relayed flows experience higher packet drop probabilities. In particular cases, interior nodes have the possibility to completely prevent communication of exterior nodes with the AP. Simulations as well as theoretical considerations prove this observation.

5.2.2 Shortcomings of the distributed medium access scheme

The current implementation of IEEE WLAN 802.11 allows very balanced per-node fairness. As described, the DCF with its CSMA/CA access method guarantees a round-robin access to the physical medium. It is the optimal access scheme for distributed networks with fair load distributions between all nodes in the network. However, MANETs with APs at their centers have radial symmetric distributions of network load and require different access schemes for optimal performance. The load constantly increases towards the AP, as all nodes must communicate via this central gateway (see Figure 70). Thereby it is of no consequence if packets traverse the network to the AP or away from the AP towards mobile nodes. Interior nodes always forward packets for the communication of leaf nodes. The per-node fairness of the DCF function allows leaf nodes within a centralized network to access the medium over proportionally. They are unable to detect the inability of interior nodes to provide the expected throughput. Consequently, interior nodes frequently drop packets of exterior nodes.

Figure 79 depicts a typical scenario with a couple of nodes grouped around an AP. Without loosing generality and to simplify the example, it only considers flows towards the AP. The numbers below the nodes depict the access rate with the RTS/CTS collision prevention mechanism activated. The arrows depict the number of individual flows from different sources forwarded at any node. As RTS/CTS mechanism blocks medium access attempts within a two-hop radius around transmitters, the access rate of interior nodes is lower than the rate of edge nodes. Consequently, leaf nodes are more seldomly blocked than interior nodes and therefore have the highest medium access rate. However, interior nodes must forward their own traffic and the traffic of all exterior nodes, but achieve the lowest access rates. This constellation leads to queue overflows and high packet loss probabilities. It is obvious that this configuration is not able to allow optimal performance. As a result, leaf nodes must reduce their data rate in order to stabilize the network.

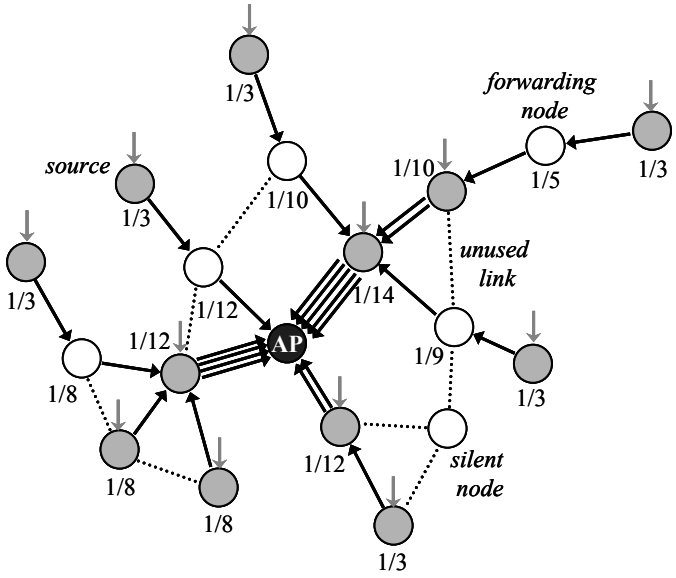

Figure 79: Flows through nodes and their individual access rates when utilizing the RTS/CTS mechanism.

To allow a more detailed analysis, Figure 80 presents a simpler example scenario with four nodes and an AP. Each node acts as source node and in addition nodes n_3 and n_4 must forward the traffic of the other nodes as well. All nodes generate packets with the maximal possible data rate of 100%. Utilizing the RTS/CTS mechanism within the depicted scenario implies that all other nodes block node n_1. As four nodes compete simultaneously for exclusive medium access, n_1 achieves an average medium access rate of 25%. Reusing the above assumptions, it turns out that all four nodes have the same equal access rate of 25%. The round-robin access mechanism of 802.11 does not consider the necessity that n_3 and n_4 must forward packets for the other nodes. As a consequence, especially n_1 is unable to access the medium as often as required, and therefore its forwarding queue overflows. The considerations in section 5.2.1 depict that single queue implementations favor own packets rather than forwarded packets, and therefore node n_4 drops most of the packets from other nodes.

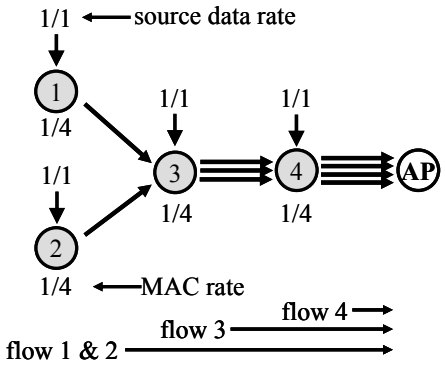

Figure 80: Scenario with four uncooperative source nodes and equal MAC access rates.

In order to verify the above made assumptions and statements, simulations with the network simulator ns-2 are carried out. Figure 81 depicts the results and confirms the statements. To prevent simulator internal shortcomings, each flow starts one second after

the previous one. Thereafter, the induced data rates remain constant till the end of the simulation. The throughput of flow f_1, f_2 and f_3 drops to zero, while node n_4 is able to transmit almost all packets of its own flow f_4 to the AP. On the one hand, this behavior is caused due to the unfairness of single queue caching methods. However, the unfairness occurs especially when nodes receive more packets than they are able to forward. They must cope with excessive load. The exterior nodes inject too many packets and hence overload the complete network. The unfairness initially occurs due to this overload. The simulation results in Figure 81 confirm that the access rates of all nodes are between 20% and 40% (average about 25%) and therewith verify the above made derivations.

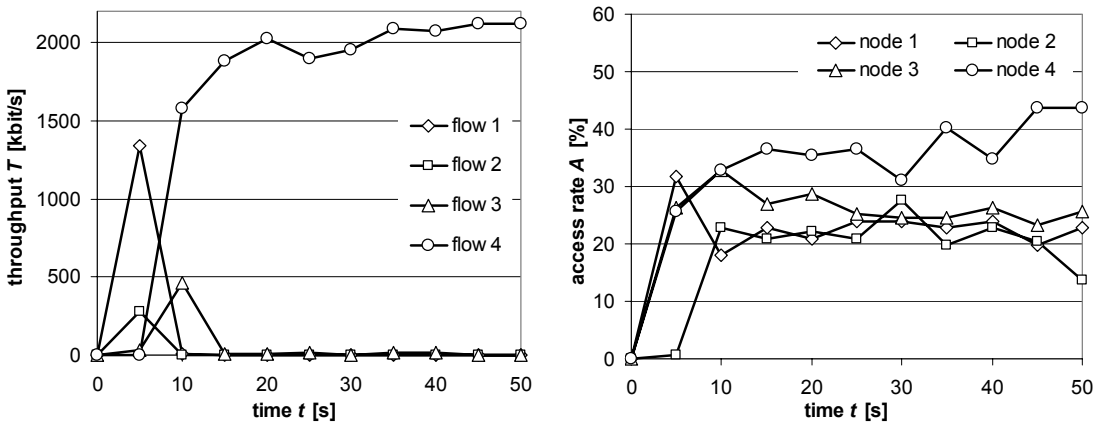

Figure 81: Throughputs and access rates for the scenario in Figure 80.

The simulation results for this simple scenario illustrate that the unfairness with the common DCF access method is already detectable within small and static network configurations. The impact increases with increasing network sizes and packet loads and the missing per-flow fairness becomes severe.

In order to circumvent the fairness challenge of the DCF function, an advanced protocol must limit the MAC access rates of exterior nodes. As the network topology within mobile ad hoc networks constantly changes, a control mechanism is necessary which is able to quickly adapt to new network conditions. An improvement like the QoS extension 802.11e does not circumvent the behavior as it basically utilizes the same distributed access scheme as the legacy 802.11.

Therefore, fair queuing does not guarantee low packet loss. Even if the protocol fills queues perfectly fair, packet losses still occur because of overloaded networks. Only a mechanism circumventing both misbehaviors prevents queue overflows and unnecessary collisions on the physical medium.

The following section discusses protocols and algorithms from the literature to prevent the performance degradations of 802.11 in radial symmetric networks.

5.3 Related Work

The throughput improvement of the WLAN 802.11 legacy protocols is a widely discussed research issue in recent years. Solutions for the fairness dilemma in ad hoc networks based on WLAN are analyzed as well (see [200, 201]). However, the proposed

optimizations are generally designed for distributed multi-hop MANETs without the availability of an AP. However, the above mentioned challenges of networks with AP are not covered by these publications.

Xu and Saadawi [44] originally examined the unfairness of the IEEE 802.11 within multihop ad hoc networks. They simply simulated a network comprising seven static nodes forming a string topology. All nodes use TCP Reno as transmission control protocol. An initially created multihop connection achieved constant throughputs, before a second single hop connection is setup. Shortly thereafter, the multihop connection fails to achieve any throughput at all. A subsequent analysis reveals that an intermediate node of the multihop connection drops an ACK packet. Thereafter, the flow control of the source reduces the data rate, while the other TCP connection continuously increases its data rate. This effect amplifies with every additionally dropped packet, till the single hop connections utilize the entire available data rate. As conclusion, the shortcomings of WLAN 802.11 occur independent from the utilized transport protocol. TCP as well as UDP connections are affected.

The authors in [202] emulate a self-clocked fair queuing scheme in a distributed manner. The algorithm is comparable to the legacy WLAN 802.11, it only modifies back-off timers, and intervals for the shared medium access challenge. Hence, nodes transmit packets with highest priority tags first. However, the proposed protocol is unsuitable for networks with APs, since it assumes uniformly distributed network load. Further, the protocol does not support multiple flows as it does not support per-flow fairness. Consequently, it is unsuitable to achieve variable access rates for leaf nodes with single flows and intermediate nodes with many simultaneous flows.

The literature refers to multi-hop scenarios with an AP often as wireless mesh networks. This term is usually used for network environments with static nodes or at least very slow moving nodes. Section 2.3 already discussed that the IEEE originally developed the 802.16 [40] standard for fixed broadband wireless access systems. The extension 802.16e allows for node mobility, but it solely supports single hop connections. Hence, it is unsuitable for multi-hop connections. On the other hand, the proposed 802.16f extension deals with multi-hop connections, but again focuses primarily on static network topologies. It does not rely on continuously retrieved network information to optimize the overall data throughput. Both extensions do not cover all characteristics of a mobile wireless multihop ad hoc network with AP. Therefore, both extensions are unable to allow fair multihop connections within these environments.

The publications [203-206] introduce the challenges inherent for scenarios with dynamic topology changes and present some interesting new approaches. The authors in [203] propose a new protocol to improve QoS in mobile meshed networks. It is based on the QoS extension WLAN 802.11e [196] and supports networks with an AP within its center. Nodes constantly classify themselves to a certain priority class. The decision depends on the node's current throughput, its current class and its throughput bound. This indicates that an optimized scheduling together with a prioritization of certain data packets improves the overall throughput and reduces the packet delay. The algorithm demonstrates good results within small scenarios with up to 20 nodes and a low mobility profile. Simulations illustrate that the overall packet delivery ratio does not increase, but the algorithm improves the in-time delivery of delay constrained packets. The publication does not consider high packet loads as additional challenge. Further on, the authors admit that the protocol may not converge in all cases, especially within heavily

loaded networks. Simulations only use scenarios with few nodes and connections require only single-hop transmissions. Additionally, nodes are static or continue along certain repetitive paths. The paper does not make performance statements for large networks with highly dynamic nodes and frequently changing topologies. It remains questionable, whether this algorithm is able to maintain larger networks with faster moving nodes.

The challenge of fair individual throughput among all participating nodes within MANETs with AP is explained in [204]. The network layer individually prioritizes each data stream and exclusive queues for each flow in the MAC layer allow perfect fairness among different flows. However, in real world environments, the maintenance of individual queues in all participating nodes for any data stream is unrealistic. Some P2P clients for example open as many TCP connections as possible.

The authors in [205] indicate the shortcomings of per-node fairness strategies. As depicted in section 5.2.2, the particular structure of MANETs with AP reduces the overall available throughput rate of the network. The advantages of spatial reuse mechanisms do not improve the capacity of the network as in distributed ad hoc networks. The authors theoretically prove that most parts of the data communication to and from the AP must use the same few links. Consequently, these links create a bottle neck for the overall throughput. They show that optimally calibrated network parameters reduce packet loss and delay and improve the overall throughput in comparison to the legacy 802.11 protocol. However, the calculation of optimal parameters require a global knowledge about the current network topology and therefore network nodes with only a local view are unable to set the parameters correctly. Simulations only demonstrate the ability of the protocol to improve the network performance in static and small networks. The mobility of nodes is not considered as a simulation parameter.

The work in [206] uses a different approach to improve the performance in networks with AP: The extension requires multiple transmission channels for signaling and data communication. Therefore, it is especially suitable for Orthogonal Frequency Division Multiple Access (OFDMA) systems like WLAN 802.11a. The RTS-CTS mechanism necessary for shared medium accesses is substituted with a four-way handshake. The initial signaling requires a common channel to negotiate a dedicated data channel. Therewith, the subsequent data transmission is able to use a channel exclusively. The dynamic channel selection allows a significant performance improvement within small and static networks. Again, with larger and dynamic scenarios, the results severely degrade.

The proposed MAC algorithm in [207] allows a distributed packet scheduling to meet delay bounds. The RTS/CTS as well as DATA/ACK packets contain additional information to optimize scheduling of differently prioritized MAC packets. Nodes overhear communication of neighboring nodes to optimize their own scheduling. The authors illustrate that their algorithm reduces delay by 50% for high loaded network. However, all simulations run again with static nodes, therefore, statements about performance in mobile environments are impossible. Especially the impact of frequent route breaks and reestablishments on the behavior of the algorithm remain unanswered.

The analysis of the shortcomings in section 5.2 indicates that the legacy-MAC protocol does not allow flow fair throughputs within networks with APs. Especially the queuing within the MAC layer and the shared medium access shows insufficiencies. The literature introduces some interesting approaches, but fails to give a commonly utilizable solution. Neither of the above extensions and protocols is able to improve the

performance in the mentioned scenarios. Therefore, the following sections indicate optimizations and new protocols to circumvent the above mentioned performance degradations. They generally optimize the overall throughput and allow very balanced per-flow fairness.

5.4 Enhancements achieving fair throughput distributions

5.4.1 Optimization of the MAC queue

Section 5.2.1 describes the inability of the current MAC protocol to fairly fill up its queue. Nodes favor own packets rather than forwarded packets and as a result, packet ratios in queues are unbalanced. Especially under high traffic loads, forwarding flows experience high packet loss probabilities. To prevent the unfair enqueue ratio, the separation of both types of packets seems promising. Instead of using multiple queues for different priority classes, as 802.11e does, the differentiation criterion is the origin of packets. The primary queue is for own packets while the secondary queue is the relaying queue, containing only packets from other nodes. The two queues do not distinguish between different priority classes yet. However, an optional extension with this functionality would require only minor changes. Before separating packets in different priority queues, the protocol must distinguish between own and forwarded packets. Following the 802.11e standard, the new MAC layer protocol contains two independent protocol instances to access the medium. Comparable to 802.11e, a subsequent scheduler behind the queues prevents internal packet collisions. Both instances dequeue packets only from their assigned queue and maintain their own backoff timer and contention window. Figure 82 presents the functional blocks for the legacy-MAC as well as the new MAC with separated queues and protocol stack.

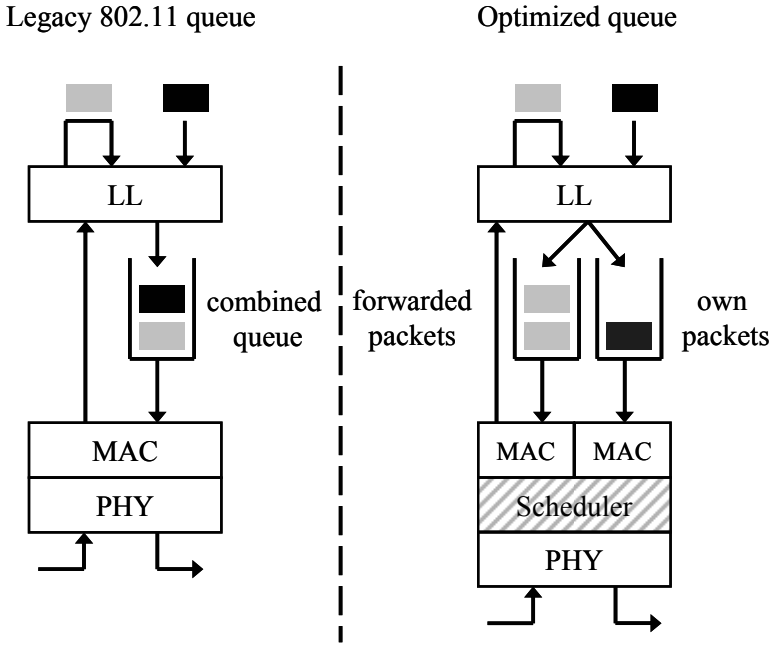

Figure 82: Queues and protocol stacks for the legacy 802.11 and for the fair-MAC extension.

In case both instances use the same timer settings for the CW and for the DIFS value, they are equivalent. The utilization of two queues within a node is transparent for other nodes, as both queues block each other and therewith the overall medium access rate for the node remains unchanged. Both instances have an access probability of 50%, when assuming that a combined stack has an access rate of 100%. The probabilities are variable when modifying the CW and DIFS parameters. In case one queue is empty, the other queue is able to achieve full access.

Figure 83 depicts a simple static network scenario with three nodes and an AP. Two sources send with 100% data rate and the third node forwards both flows. The queuing algorithm is able to distinguish between marked packets from node n_1 and n_2. This scenario already verifies the ability of two independent queues to improve the fairness among different flows. Following previous considerations, all nodes have the RTS/CTS mechanism turned on, and therefore should achieve an overall access rate of 33%, independent whether they utilize a single protocol instance or two. However, the two-queue implementation divides the available access rate in all nodes different. Simulations in [208] verify that the assumptions are valid and the combined access rates are 33%. The queues in each node are self-controlled and therefore can block themselves.

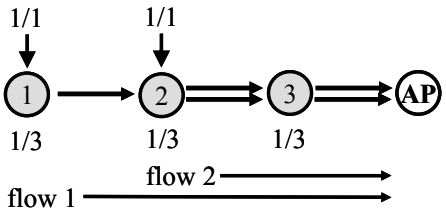

Figure 83: Static network scenario with two sources and a forwarding node.

The results in Figure 84 depict averages for simulations over 50 seconds with the above introduced example network. The left figure depicts the overall achievable throughput for both flows f_1 and f_2 utilizing the single queue stack of 802.11 and the new optimized two-queue solution. With respect to the queuing solutions, both flows experience different achievable throughputs. As described in section 5.2.1, with the single queue solution, node n_2 favors own packets and mostly drops packets of flow f_1. Consequently, the throughput of f_1 drops to 100 kBit/s while flow f_2 achieves 1800 kBit/s. With the two-queue solution, both flows achieve almost the same throughput. Interestingly, the combined throughput is equal in both cases. The realized fairness does not degrade the maximal combined throughput. It clearly depicts the ability of the optimized queue to fairly divide the available access rate among both flows.

The right graph in Figure 84 confirms this behavior. It illustrates the transmitted packets per second of node n_2 with respect to their origin. With a single queue, n_2 transmits own packets in the majority of cases. The ratio is about 25:1, this means it forwards 25 own packets for each packet of flow f_1. With the two-queue solution, node n_2 forwards almost as much packets from f_1 as from f_2. The introduced second queue is solely responsible for forwarding packets and therewith prevents the unfairness of the legacy-MAC stack. However, this simple example scenario only allows fair results, because the number of own packets and forwarding packets are equivalent in node n_2. The two-queue solution achieves perfect fairness among both flows, whenever two flows have equal data rates in intermediate nodes. In case multiple flows must be forwarded, the queue ratio of own

and forwarded packets turn again unfair. Therefore, the algorithm is not yet utilizable in general. Both queues compete for the access, and the forwarding part of the protocol stack is unable to gain access with the correct rate. With only a second queue, the protocol stack fails within networks where forwarding traffic exceeds the traffic from upper layers.

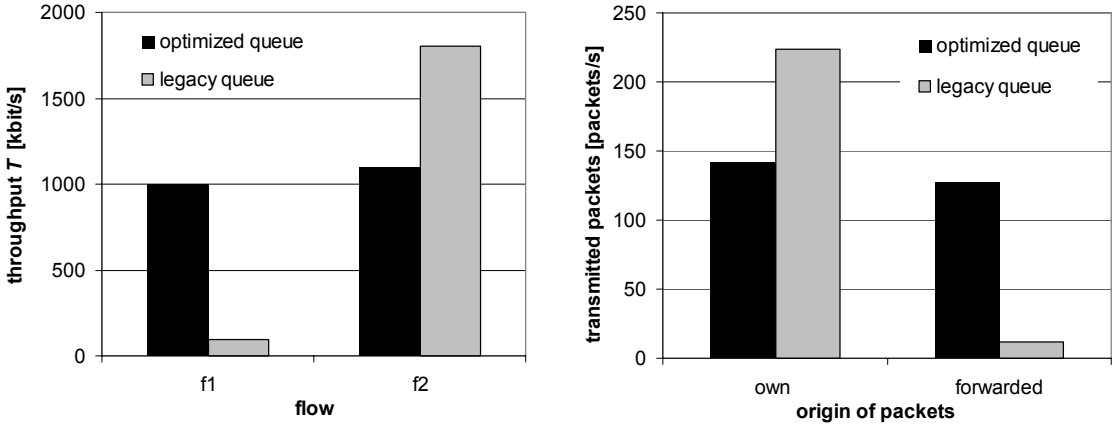

Figure 84: Throughput and packet ratio for a single queue and two independent queues.

However, as described the two-queue solution allows a fair distribution of medium access between own and forwarded packets. But it does not prevent overload situations due to excessive node data rates. Therefore, nodes do have to limit their data rates. The following section describes the basic calculation method to prevent these overload situations.

5.4.2 Prevention of excessive overload conditions

The limited radio ranges of ad hoc nodes require multi-hop connections, but allow the spatial reuse of the wireless channel. The medium is simultaneously utilizable, as long as concurrent transmissions do not interfere. The spatial reuse allows much greater network capacities. However, the advantages of spatial reuse are not usable in networks with a central AP. As all data flows traverse through the AP, the channel around the AP forms a bottleneck. Even if two distant nodes are able to transmit packets simultaneously, these packets reach the AP, and within its proximity the improvements of spatial reuse are not applicable. Therefore, the capacity of channel around the AP limits the overall achievable throughput.

As demonstrated, a medium access algorithm with per-node fairness is inappropriate for MANETs with a central AP. The utilization of a per-flow fair protocol is favorable, because each flow achieves the same throughput towards the AP. Therefore, the access mechanism must consider the forwarding traffic separately. Consequently, nodes with high forwarding traffic load must increase their access rates to minimize packet losses. Additionally, other nodes (mostly exterior nodes) must cooperatively reduce their rate, to prevent an overload situation within networks.

Figure 85 depicts the same network as in Figure 80, but indicates the optimal MAC access rates to achieve per-flow fair throughputs. Each node requires two information to determine the correct access rates: the overall number of accesses A_{comb} and the number of flows forwarded by itself. The number of combined accesses A_{comb} is determined as

the sum of all forwardings of all flows. Within the example, flow f_1 and f_2 requires three forwardings, while f_3 and f_4 require two respectively a single forwarding. Therefore, the combined number of accesses A_{comb} of all flows is nine. It defines the denominator for each individual access rate.

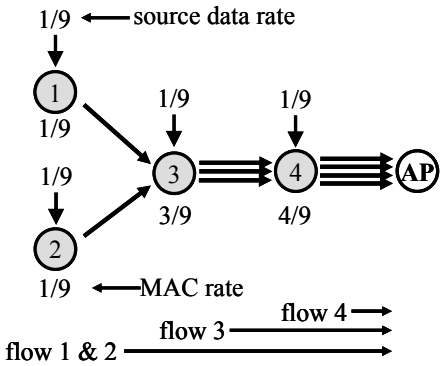

Figure 85: Cooperative nodes with optimal access rates within a per-flow fair network.

The number of flows forwarded by each node defines the enumerator for the access rate. As node n_1 only forwards its own flow, the enumerator of the access rate is one. Consequently its rate as well as the rate of node n_2 is 1/9. Node n_4 forwards four flows (three as relay and one own flow) and obtains an access rate of 4/9. The access rate of n_3 is calculated in the same manner. In order to prevent packet losses in queues, the data rates must be reduced as well. As the throughput should be shared uniformly among all flows and the AP is the bottle neck within the scenario, A_{comb} defines the data rate per flow as well. The data rate D_i for flow f_i is $1/ A_{comb}$. The optimal data and access rates grant each node and flow as much access and throughput as necessary to permit flow fair operation. It allows each flow to achieve equal throughputs and prevents packet losses in intermediate nodes.

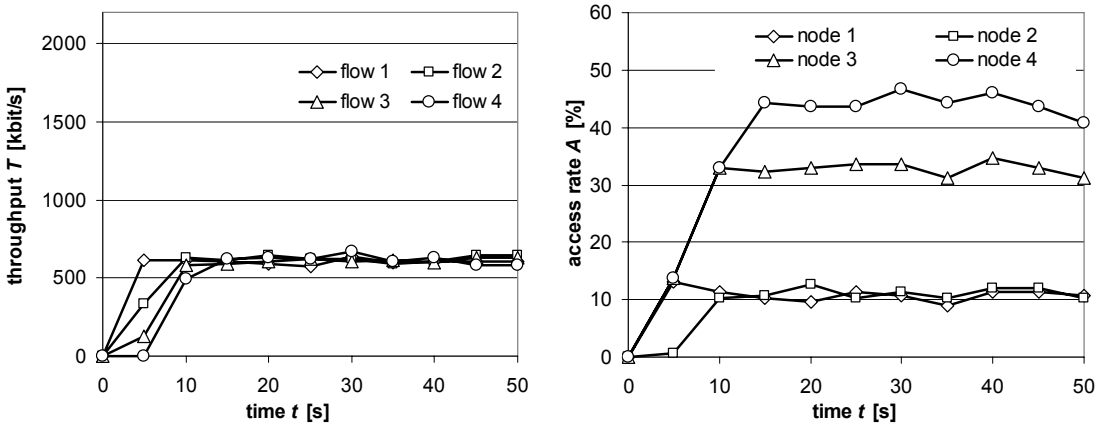

Figure 86: Simulation results with optimal data and access rates.

Therefore, the data rates for all flows are reduced to 1/9 of the maximal possible throughput. Figure 86 presents the simulation results for the optimal data and access rates. The results illustrate that the throughput for all flows is equivalent. Obviously, the access rate of each node varies with the number of forwarded flows. The optimization reduces the throughput fluctuation to an absolute minimum. Nodes are able to gain

access to the medium with the correct probability. The new access and data rates achieve perfect per-flow fairness. Therewith, the basic characteristic requirements are available to achieve fair throughput distributions within networks with APs.

As depicted, nodes and their individual protocol instances require different medium access rates. A first option to achieve these variations is the modification of backoff times of individual protocol instances, comparable to the approach of 802.11e. Protocol instances, requiring high access rates have shorter initial CWs than others with small rates. Simulations in [208] demonstrate that the modification of the CW do improve the fairness among different flows. However, the mechanism does not achieve optimal results and equal throughputs.

Another possibility is the limitation of the source data rates. The MAC protocol instance which is responsible for own packets, has knowledge about the optimal data rates. Therewith, it is able to restrict the number of self-injected packets. This prevents the appearance of overload situations directly at the basis. If necessary, it simply delays own packets before the dequeuing to achieve the optimal average induced rate. More sophisticated queuing schemes [209, 210] like token bucket [28] or leaky bucket [211] would be utilizable. But they are not discussed further, because the current simple solution already allows optimal performance. Modifications at the relaying queue are unnecessary, the utilization of the existing protocol parameters is sufficient. With the limitation of source data rates, the remaining data rate for packet relaying is adequate to prevent queue overflows.

Figure 87 depicts the results for the scenario in Figure 85. The graphs again present the throughputs for all four flows as well as the access rates of all nodes as average over 50 seconds. The simple limitation of source data packets is sufficient to achieve optimal medium access rates. The system is self-organizing. Queues do not drop packets, because they never congest. The graphs clearly depict that the data rate of all flows are exactly equivalent, because the access rate of all nodes are optimal.

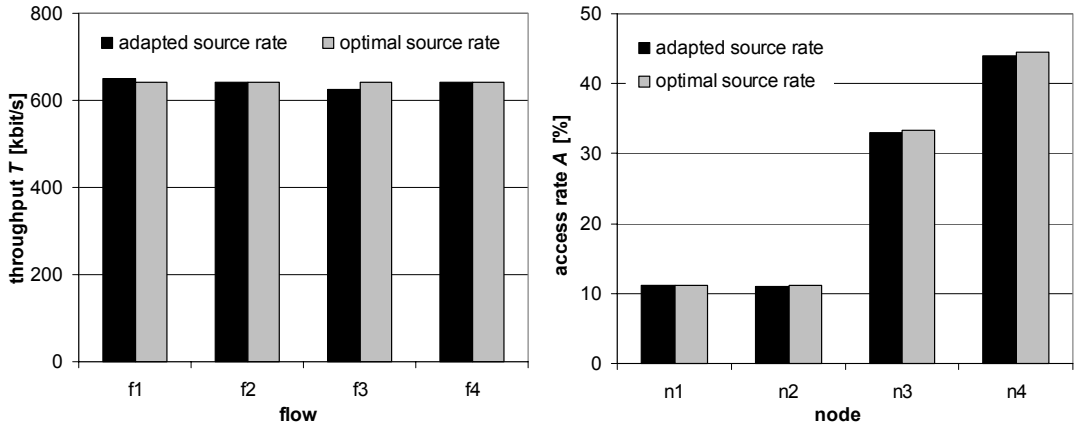

Figure 87: Throughput and access rate results with limited source data rates.

If upper layers request higher data rates, the delaying of source data packets leads to queue overflows. Obviously, this artificial limitation of data rate causes additional packet loss. If the MAC layer must reduce the throughput of source data rate, TCP may erroneously detect a network overload situation and initiates a slow-start. Or UDP constantly keeps its data rate unchanged and packet destinations realize an intolerable

high packet loss. A cross-layer exchange of information is able to prevent unnecessary performance degradations. The MAC protocol notifies higher layer about the changed maximal available data rates. Thereby, the higher protocol instances are able to instantly adapt to the new conditions and therewith avoid packet drops of own packets.

5.5 The fair-MAC protocol extension

The following section describes in detail the fair-MAC extension. It overcomes both mentioned shortcomings of the legacy 802.11 protocol within MANETs grouped around an AP. It utilizes the solutions presented in section 5.4. The fair-MAC extension is especially designed for networks with AP, but considers node-to-node communication as well. A previous version of the extension is described in [7]. Additional insights are given in the pending patents [22] and [23]. It is only capable to handle unidirectional traffic from ad hoc nodes towards the AP. The extension, proposed in the following allows the utilization of bidirectional traffic, and therefore the usage of TCP traffic.

The extension utilizes all basic algorithms and protocols as the legacy-MAC, like CSMA/CA, backoff and carrier sensing. It simply extends the link layer (LL) with additional functionalities to achieve flow-fairness. Most traffic traverses via the AP and therefore the AP is the most suitable entity to adapt the key components for the optimal access rates of nodes. The protocol extension is transparent for upper layers, therefore, modifications of routing algorithms and transport layer protocols are unnecessary.

The protocol assigns a single MAC-flow to every packet-generating source in the MANET, whether it initiates several IP flows to different destinations or not. Intermediate nodes classify IP packets based on the address of the ad hoc node, independent if the packet contains the address as source or destination. Depending on the direction of the packet, the algorithm must evaluate either the source or the destination IP-address. For flows towards the ad hoc node, the destination address is necessary, for packets traversing the other direction, the source address is important. The utilization of both addresses simultaneously to determine the number of flows is impossible. Ad hoc nodes might create multiple flows to different destinations and the fair-MAC algorithm would count them multiple times when considering the source as well as the destination addresses. As a result, nodes are again able to demand full access rates by increasing their own number of flows.

The LL is without memory of previous packets. It is unable to distinguish whether a packet received from the higher layer is a relayed or self-generated packet. Additionally, it cannot determine if a packet traverses the network towards the AP or away from it. After the reception of a relay packet, the LL removes the MAC header. Therewith, all MAC layer related information is erased. The LL forwards the packet to higher layers for further processing. Figure 88 schematically illustrates the processing. Afterwards, the network layer returns the packet to the LL, and it adds a new MAC header to the packet. Hence, additional information within the MAC header about the status of the packet is invalid. The usage of IP or TCP headers to tag packets is not possible either. It would require modifications at upper layer protocols.

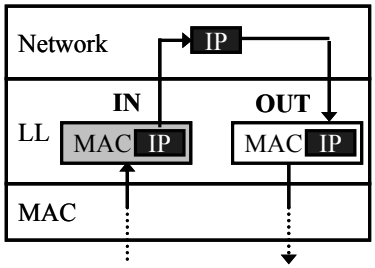

Figure 88: Schematic illustration of the memory less link layer.

In order to allow this differentiation, the AP adds the reverse-bit (R-bit) to all packets coming from the fixed network, indicating that this packet traverses towards nodes within the network. Additionally, it puts the destination IP address as packet ID into the MAC header (see Figure 89). The link layers of ad hoc node sources utilize their source IP address as ID and set the R-bit to zero. The extension to the regular MAC header contains the ID, the hop count *HC*, the optimal rate *OR* as well as the reverse bit. The extension requires six Bytes, while the common MAC header requires 30 Byte. Figure 89 illustrates the new structure of the MAC ACK headers as well. The MAC extension differentiates between ACK packets in the direction towards the AP and in the opposite direction. With together 15 Byte, the new ACK headers require one additional Byte. The extension does not require additional packets for information exchange. It is possible to communicate all changes in the flow topology with the already available packets.

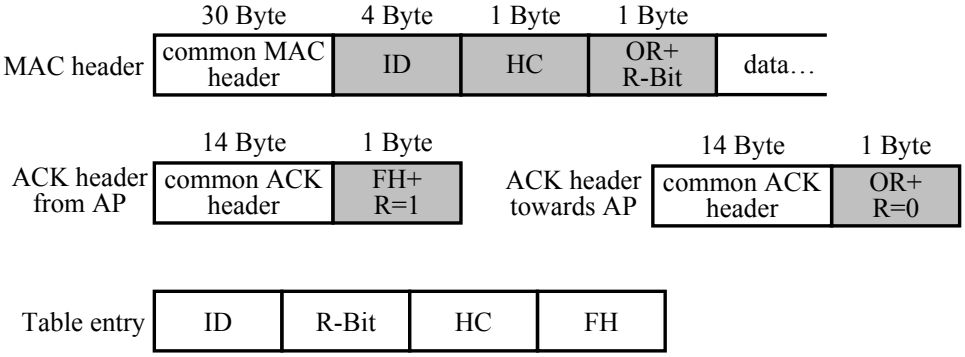

Figure 89: Table entry to store flow information, as well as MAC and ACK header of the fair-MAC protocol.

The LLs of intermediate nodes receive theses MAC packets and record the ID and the R-bit within their flow tables before they forward the packets to upper layers. Figure 89 shows such a table entry. When again receiving packets from the network layer, the LL checks whether the destination address is available within its flow table. If so, it sets the ID for the outgoing packet to the destination address and sets the R-bit to one. Otherwise, it checks for the source address within the list. If available, the LL uses this address as ID and sets the R-bit to zero. In both cases, the algorithm saves the corresponding *HC* and the *OR* as rate within the MAC header. The ID does not necessarily be the IP address of source or destination, there are also other node differentiations possible. However, this approach seems to be the most appropriate and practical solution. If both queries fail, the algorithm assumes a new flow from this node. The LL uses its own IP address (stated in the IP packet) as ID and sets the R-bit to zero.

Within the outgoing packet as well as within the ID-list, the algorithm sets the *HC* to zero.

Figure 90 presents a small network scenario example, which utilizes the functions of the fair-MAC extension. Node n_1 initiates a TCP connection with an Internet based destination. Therefore, the AP serves as additional source for packets within the MANET. The individual states illustrate the traversing packets, the current node access rates, as well as the table entries of node n_1 and the AP. In (b), n_1 initiates the new flow and the LL saves the packet information in the table. Each forwarding node increases the hop count *HC* by one. After reception of the packet, the AP knows the optimal rate *OR*. It replies ACK packets back to the source with an *OR* of three. Certainly, MAC ACK packets acknowledge only the correct reception of packets between link neighbors. Therefore, node n_1 does not receive the ACK with the correct *OR* immediately, but only after the transmission of multiple data packets. For the sake of simplicity, Figure 90 depicts these multiple ACKs as single step. In (d), the AP initiates its returning packet flow. Upon reception, node n_1 gets knowledge about the utilized number of hops *HC*. The ACK packets towards the AP in (e) contain the utilized number of hops of the flow between the AP and the ad hoc source as flow hops *FH*. Therewith, the AP is able to determine the optimal rate *OR* with six and subsequent ACK packets inform all network nodes about the new *OR*.

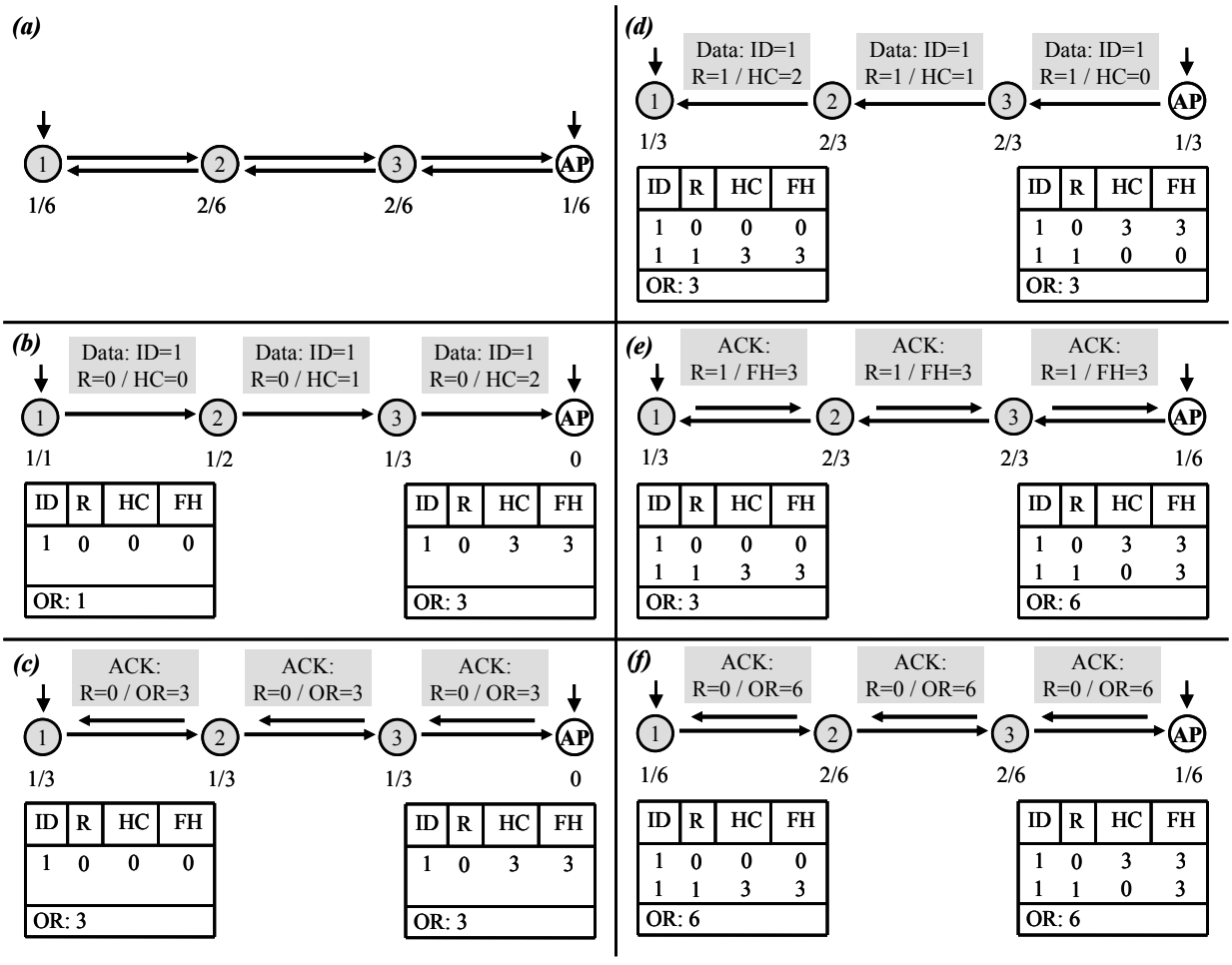

Figure 90: Example for the distribution of the optimal access rates with the fair-MAC extension.

Within all ACK packets traversing the network in the direction towards ad hoc nodes, the fair-MAC extension informs exterior nodes about the new *OR*. This information migrates towards the leaf nodes with every new data packet towards the AP. Ad hoc nodes, receiving new information about the *OR* adapt their own medium access rate in order to meet this rate. Additionally, whenever they begin relaying a yet unknown packet flow or start generating an own flow, they adapt their medium access rate as well. As stated in section 5.4, both MAC instances independently compete for the medium access. The fair-MAC algorithm within each node calculates the access rate of the source packet queue as $1/OR$ and the relaying queue as FF/OR, where FF is the number of forwarding flows.

In order to circumvent difficulties with flooding messages from ad hoc routing protocols, nodes use a threshold value of ten packets with corresponding IDs to indicate a new flow. If not utilized, flooding messages instantly create numerous new flows and the fair-MAC algorithm would adapt the medium access rates of nodes to almost zero. To prevent numerous table entries and to reduce memory requirements, the extension uses timers to remove aged entries. Every reception of a packet belonging to a certain flow resets the flow timer. In case the timer expires, the protocol removes the specific table entry. The expiration period depends on the granted data rate and is dynamically adaptable. As described in section 5.4.1, nodes are able to enqueue own packets within the source queue at any rate. However, the fair-MAC extension informs the upper layers about the current maximal access rate for source packets and therefore allows a quick adaptation.

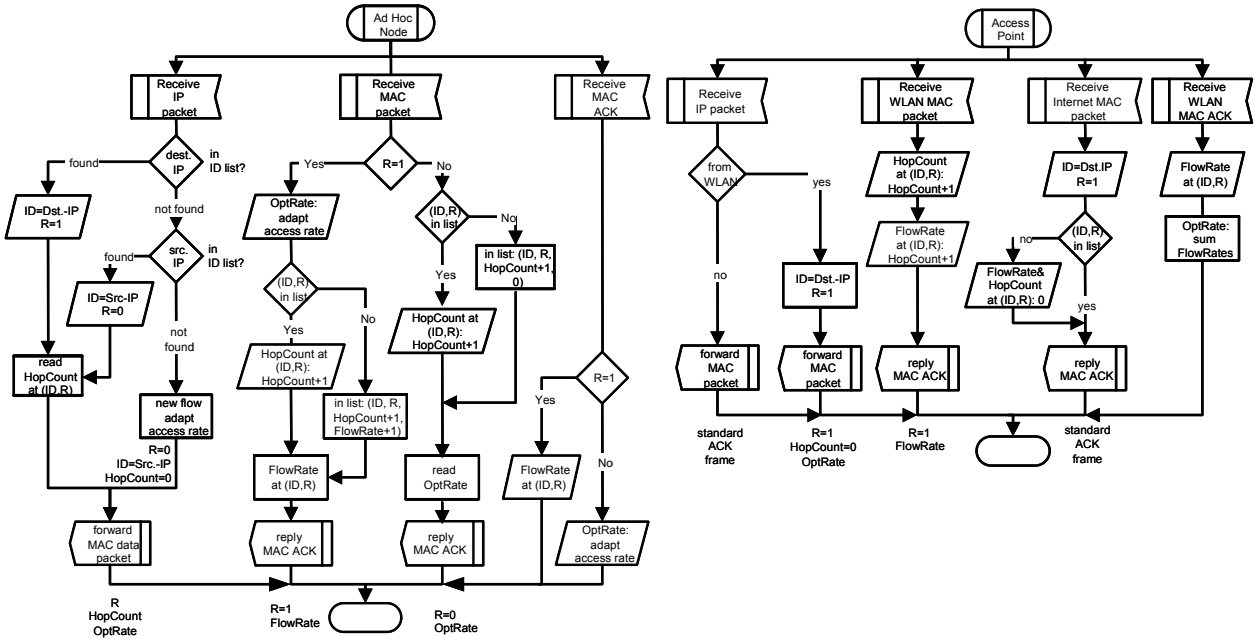

Figure 91: SDL diagram of the fair-MAC extension.

The distributed access rates of nodes are temporarily not optimal, because node mobility, flow modifications, route breaks and establishments constantly alter the network conditions. However, the fair-MAC implementation is able to cope with frequent route breaks and re-setups. The network is only transiently out of its optimal condition and the MAC queue responsible for relaying packets must drop packets rarely. As new routes often have comparable number of hops, the old flow rates are good approximations till

new values reach the AP. Figure 91 depicts the SDL diagrams for the fair-MAC extension for nodes as well as for APs. To reduce the complexity of the diagram, it does not include timer events to remove table entries as well as the threshold values to detect new flows.

5.6 Evaluation

5.6.1 Simulation environment

In order to evaluate the proposed fair-MAC extension, simulations with the ns-2 were carried out. Section 2.5 describes the ns-2 in detail. The network layer protocol is not part of this evaluation and shall not affect the results. Therefore, simulations utilize the AODV algorithm, as it is the only available reactive routing protocol standardized as IETF RFC [74]. The AODV-UU [172] ns-2 extension from the University of Uppsala is compliant to the RFC and therefore used in the following.

The common simulation setup was already introduced in section 4.6.2. The evaluations utilize the WLAN 802.11b version coming with ns-2. Its overall bit rate is 11 MBit/s which permits a TMT of 6.5 MBit/s. As described, node movements follow the RD mobility model with bounce back at borders [92]. The maximal node velocity is scenario dependent and between zero and 10 m/s. The number of nodes is either 50, 100, or 200 while the node density remains constant. In contrast to the evaluations in chapter 4, the initially considered environments are flat and therefore nodes are always in line-of-sight to each other. Therefore, a reduced node density of ten neighbors on average guarantees fully connected networks and is in accordance to the examined optimal node density in [132]. It reliably prevents misrepresentation of results due to special topology conditions. With a maximal free-space radio transmission range of 150 m, the simulation area grows from $550\,\text{m} \times 550\,\text{m}$ to $1150\,\text{m} \times 1150\,\text{m}$. A bidirectional connection consists of two packet flows, from a mobile ad hoc node towards the AP and back. The ad hoc source initiates its flow and the AP follows with its flow one second later. This emulates bidirectional traffic like TCP, while it permits exact conditions during simulation runs. Each connection lasts between 50 and 150 seconds with an average of 100 seconds. Thereafter a new ad hoc node establishes a connection with the AP. The overall number of parallel connections throughout simulations is 20 and constant in all scenarios. The continuously changing connections emulate real network behaviors and present the worst case scenario with notable routing traffic. The packet rate of each connection is variable as well, and varies from 5 to 50 packets per second. The payload of each packet is 1024 Byte and constant within all simulations. If not otherwise stated, the common parameters are 100 nodes, 2 m/s maximal node velocity, 20 packets per second and 20 simultaneous connections (40 individual flows). Therefore, the overall generated traffic of all sources is about 6000 kBit/s. All figures show the average results out of ten independent simulations with the 95% confidence interval. As described in appendix C, the duration of individual simulations varies. It is controlled by the CNCL class library. If the deviation from the average packet loss lies within a 95% confidence interval, it finally terminates the simulation.

The simulation results comprise five different parameters as output metrics. A description for the packet loss, the throughput, and the routing overhead metric is given in chapter 4.6.2. Due to different behaviors of fair and legacy-MAC, the evaluation treats packet losses within source queues differently. Queue overflows in source nodes

do not count as packet loss in case the legacy-MAC protocol is utilized, because it is unable to adapt the packet generation rate. However, they count with the fair-MAC extension.

Both other metrics are defined as follows:

- *Fairness*:
  The fairness index F is the metric to show the ability of the algorithm to share the available data rate uniformly among the present connections. The worst case (most unfair) is, if a single flow takes up the full throughput, while all other flows do not achieve any throughput. Following the definition of the standard deviation σ, the resulting deviation σ_{max} for this case is

$$\sigma_{max}^2 = \frac{1}{n_c - 1} \sum_{n_c}^{i} \left(T_i - \frac{T}{n_c} \right)^2 = \frac{1}{n_c - 1} \left(\left(T - \frac{T}{n_c} \right)^2 + (n_c - 1) \frac{T^2}{n_c^2} \right) = \frac{T^2}{n_c} \tag{95}$$

  with T describing the conjointly achieved throughput, T_i the individual achieved throughput of flow i and n_c defining the number of simultaneous flows. With the additionally computed deviation σ for arbitrary cases, the fair-index F calculates as

$$F = 1 - \frac{\sigma}{\sigma_{max}} = \frac{\sigma_{max} - \sigma}{\sigma_{max}} \tag{96}$$

  Thereby, greater values of F represent a fairer shared throughput and vice versa.

- *Transient period*:
  The transient period ΔT_P defines the times after which all nodes have readjusted their access rates to the optimal rates. In general, it depicts the rate at which the fair-MAC algorithm is able to adapt to new topologies or connections. Results only contain values for the fair-MAC algorithm, as the legacy-MAC does not require readjustments to new environments.

5.6.2 Simulation results

Figure 92 shows the packet loss for both MAC protocols against the network size and the packet generation rate of sources. Both figures clearly depict the advantages of the fair-MAC protocol. Although legacy-MAC results do not count packet losses because of queue overflows in source nodes, the packet loss is still magnitudes larger. The packet loss of the fair-MAC is always below 2%, independent from the network size and the packet generation rate.

Source nodes utilizing the legacy-MAC protocol frequently transmit packets while intermediate nodes must drop these, due to insufficiently granted access rates. Therefore, the packet loss rises for increasing network sizes and packet rates. In contrast to that, the fair-MAC protocol limits the packet access rates directly at sources. Nodes only emit those packets which have a high chance of delivery at the destinations. Therefore, the packet loss is much smaller. The packet loss increases only slightly for increasing network sizes, because the greater simulation area allows a better spatial reuse of the wireless medium.

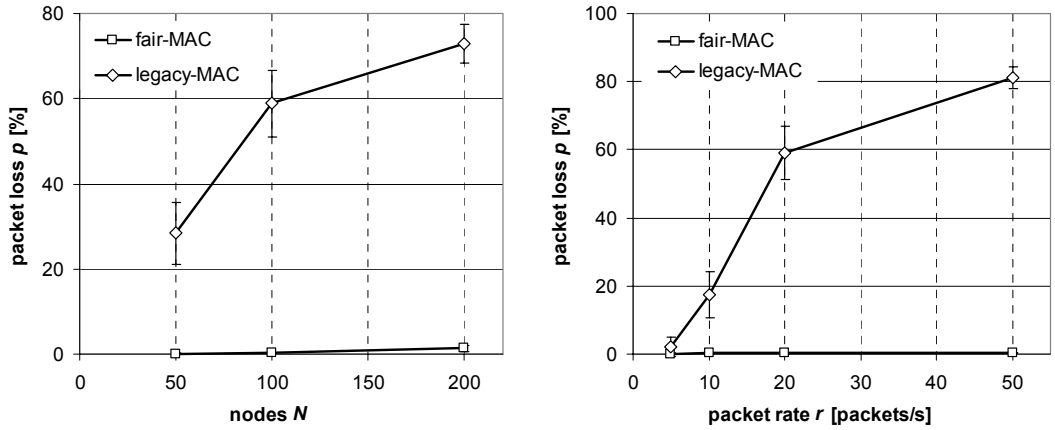

Figure 92: Packet loss of fair- and legacy-MAC against the network size and the packet generation rate of sources.

As described, during the transient period ΔT_P, the individual access rates of nodes are not optimal, and therefore packet losses may occur due to overloaded networks. However, the packet loss p with the fair-MAC extension increases linearly with increasing packet rates, because the number of emitted packets during ΔT_P increases as well. In contrast, the legacy-MAC again creates numerous packet losses. With packet rates greater than five, the loss rate is beyond an acceptable threshold. Only nodes with direct links to the AP are able to transmit their packets, all other packets are dropped by intermediate nodes.

The results for varying node velocities (Figure 93) are similar to those presented previously. The fair-MAC extension achieves significantly less packet loss in comparison to the existing legacy-MAC. As expected, the loss rate of the legacy-MAC increases with increasing maximal node velocities. The higher network dynamic causes more frequent route establishment attempts and therefore increases the overhead. This in turn leads to less available bit rate for data traffic and consequently increases the packet loss. In contrast to that, the increasing dynamic of the network does not lead to an increasing packet loss for the fair-MAC extension. It is again below 2%. The protocol extension within each node is always able to calculate the individual optimal access rates, independent from the maximal node velocity.

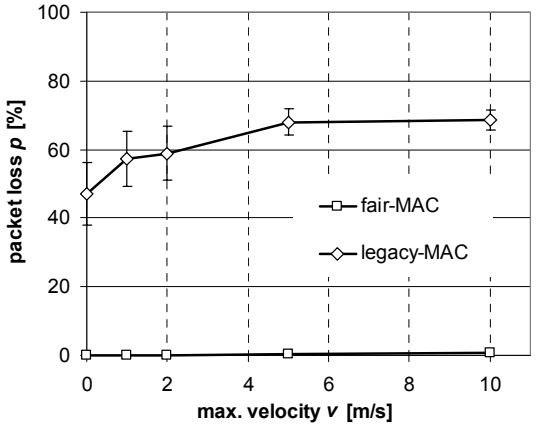

Figure 93: Packet loss p with respect to the maximal node velocity v.

Figure 94 shows the throughput on a flow basis. It shows the result for a single simulation, as the average out of multiple simulations again would equalize the throughput per flow. The scenario utilizes the previously mentioned standard simulation parameters, but only maintains 10 instead of 20 connections. The graphs obviously depict the inability of the legacy-MAC to share the available bandwidth fairly among the existing flows.

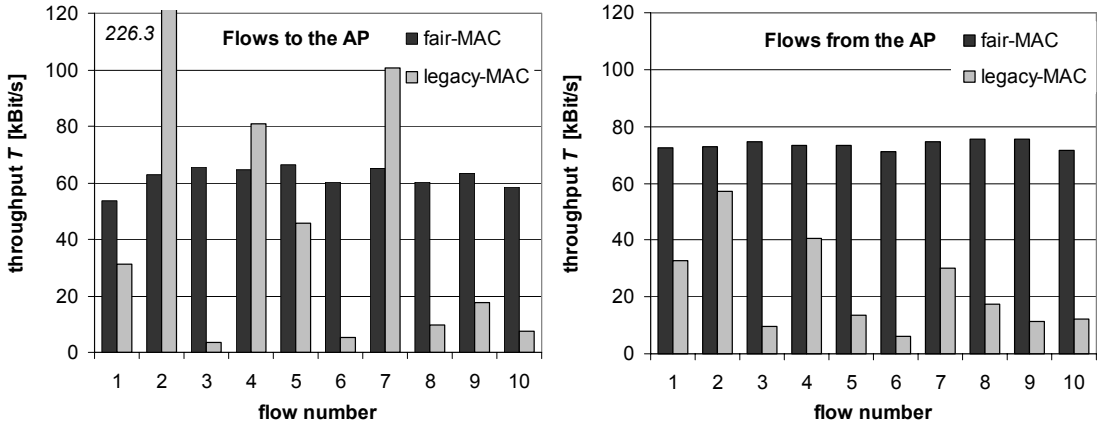

Figure 94: Throughput of fair- and legacy-MAC flows to and from the AP.

Utilizing the legacy-MAC, flow no. 2 uses three times more bandwidth than admissible, while other flows to the AP do not allocate any data rate at all. As the medium access of the legacy-MAC is node-fair, all flows from the AP to distant ad hoc nodes compete for the limited bandwidth of the AP. Therefore, they only get a fraction of their required data rate. In contrast to that, the fair-MAC algorithm achieves almost constant data rates for all incoming or outgoing flows. However, flows to the AP experience smaller throughput, as the medium access is always more difficult towards the AP. The packet loss probability in this direction is greater than in the opposite one.

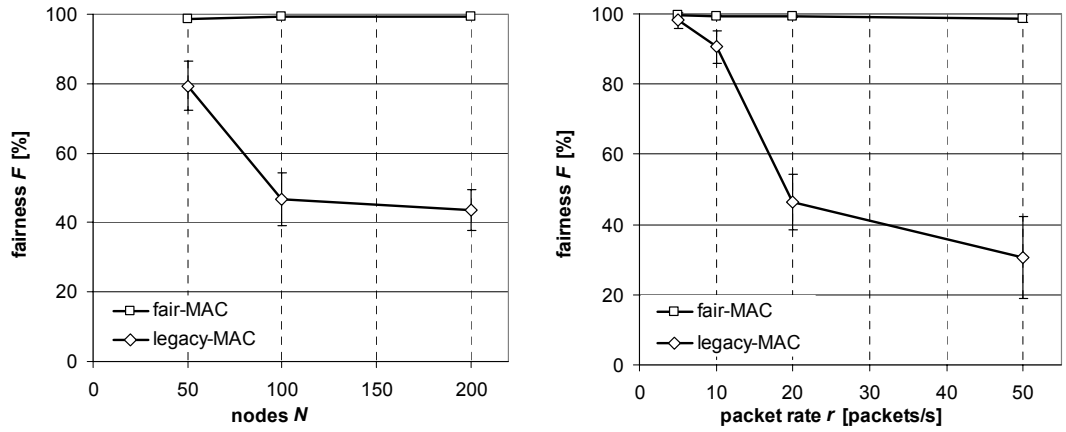

Figure 95: Fair index with respect to the network size and the packet generation rate.

In order to emphasize this observation, Figure 95 presents the fair index F for various network conditions. As described, large F illustrate the ability of a MAC protocol to share the bandwidth uniformly among all participating sources. Consequently, Figure 95 impressively depicts that the fair-MAC extension operates close to the optimal case. The

protocol extension grants all users within the network almost the same data rate, independent from the network size and the packet rate. As in the previous considerations, results only change slightly for different velocities. In contrast, the legacy-MAC achieves less than 50% fairness within networks of more than 50 nodes, or for high packet generation rates. This again indicates that some users obtain very high throughputs and almost exclusive medium access while others must deal with very low data rates.

Figure 96 illustrates the overall throughput for the fair-MAC and the legacy-MAC for varying network sizes and packet rates. With 50 nodes, the connections mostly contain only a single hop, and therefore the throughput is close to the maximal capacity. However, the achievable overall throughput degrades for increasing node numbers, independent from the utilized protocol. As the overall network capacity remains unchanged and the average route lengths increase, an increasing number of relays must share the available data rate. Therefore, each flow achieves less throughput. Obviously, the fair-MAC protocol does degrade the throughput in comparison to the legacy-MAC. However, the fairness of the legacy-MAC becomes significantly unfavorable for increasing node numbers. Only the direct AP neighbors obtain medium access and therefore keep the overall throughput high.

For only five and ten packets per second, the network scales perfectly, as the throughput increases linearly. For higher rates, the generated load exceeds the network capacity. Consequently, the throughput increases only steadily, if any. The fair-MAC throughput remains almost constant, but it keeps the packet loss minimal and remains a high fairness among different flows. In contrast, the legacy-MAC further increases its overall throughput at the expense of the fairness among individual flows. With 50 packets per second, the overall achieved throughput of the legacy-MAC exceeds the throughput of the fair-MAC by only 22%.

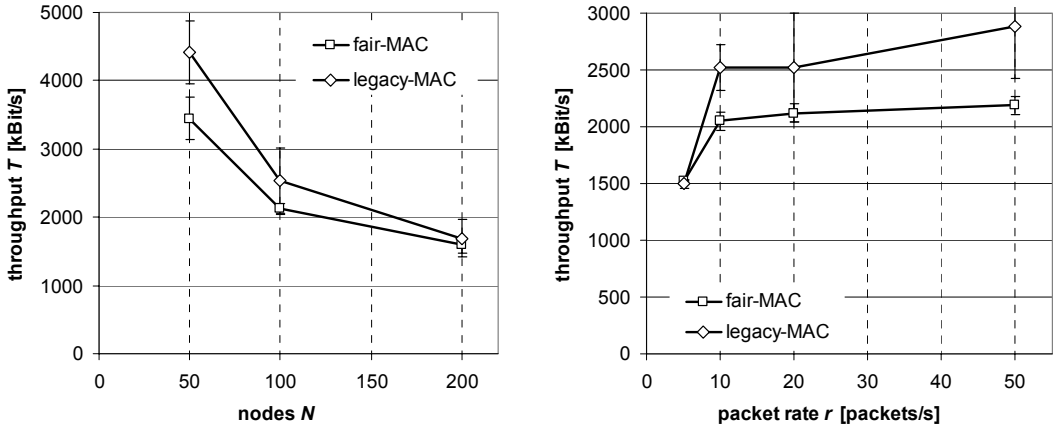

Figure 96: Overall data throughput for varying network sizes and packet rates.

The throughput results for varying maximal node velocities are not presented, because they do not reveal new insights. Certainly, the overall throughput decreases with increasing node velocities, independent from the utilized MAC protocol. For varying node velocities, the overall throughput of the legacy-MAC is at most 20% greater than the throughput of the fair-MAC, but the distribution of achieved throughputs is very inhomogeneous among flows.

As a summary, the fair-MAC protocol justly shares the existing data rate among all available flows, while the overall throughput remains constant for higher data rates. In contrast to that, the legacy-MAC extension achieves higher throughputs, but favors nodes close to the AP. Consequently, its fairness index drops to almost zero.

As expected, the transient period ΔT_P in Figure 97 increases with increasing node numbers. The length of routes between sources and destinations increases as well and therefore the protocol requires more time to inform all nodes about their new access rates.

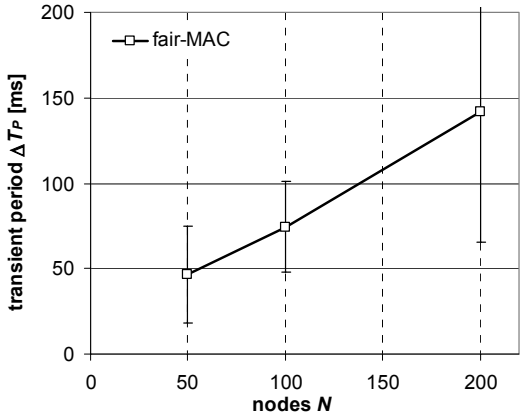

Figure 97: Transient period of the fair-MAC algorithm for increasing node numbers.

However, nodes are able to modify their access rates almost immediately after the network changed. During this settling time, intermediate values constantly alter till they reach the optimal values. The transient period ΔT_P is independent from the number of connections and the maximal node velocity, although the node velocity definitely impacts the number of necessary adaptations. The packet rate slightly alters ΔT_P, as very low packet rates also delay the spreading of information about individual access rates. However, only packet rates smaller than five packets per second shortens the transient period.

As described in section 5.5, the fair-MAC extension requires additional header fields to inform the network about new access rates. However, the extension does not generate additional signaling packets. The overhead results only from the extensions in the MAC and ACK headers. As depicted, the legacy-MAC header requires together with the ACK header 34 Byte. In contrast to that, the MAC extension requires 7 additional Byte, and therefore needs 20% more header size. However, when considering the maximal length of an Ethernet frame of 1518 Byte, the additional header parts are neglectable. They would require only 0.5% of the maximal packet length. When assuming the complete RTS/CTS/DATA/ACK cycle for a packet transmission, the additional caused time delay by the extension is significantly below 1%, independent of the payload size of the respective data packet.

5.6.3 Simulative results within urban environments

The following evaluation again focuses on the performance behavior of ad hoc networks with a central AP. However, in contrast to the previous analysis, it does not solely rely on flat environments, but considers urban surroundings as well. Therewith, it combines

the fair-MAC extension and the urban simulation environment of chapter 4. As previously illustrated, the MAC extension achieves superior performance within flat environments. However, it remains unknown, whether it is able to achieve similar results within non-line-of-sight surroundings.

Therefore, the following evaluation should reveal the behavior of the fair-MAC within urban environments, and whether it is able to cope with the significantly different characteristics.

Most of the simulation parameters for the following evaluations are equivalent to those used in the previous simulation. However, the urban environment presents more challenging conditions for ad hoc networks. Therefore, some parameters must be modified, so that the considered ad hoc networks are not completely overextended. The parameters, if not already introduced, are briefly described in the following.

The environmental scenarios have the same characteristics as those described in section 4.6.1. Consequently, the Walfisch-Ikegami model (WIM) is again used to predict the propagation of transmissions. Due to the non-line-of-sight characteristics of the urban scenarios, the average node density is raised to keep the network always connected. Therefore, the previously used simulation parameters from chapter 4.6.2 are again utilized, while the line-of-sight transmission range remains constant with 150 m.

Besides, the AODV-UU implementation of AODV together with the basic parameters of the IEEE WLAN 802.11b standard is utilized for the ns-2 simulations. Node mobility traces are again generated with the city motion (CM) model. The impact of node mobility on the performance of ad hoc networks and on the behavior of the fair-MAC extension is sufficiently examined. The following simulations do not examine the velocity impact and consequently nodes have a constant maximal node velocity of 2 m/s. Due to the challenging environment, only ten instead of twenty bidirectional connections generate packets. Additionally, the common packet rate is reduced to twenty packets per second, but the payload of packets remains constant with 1024 Byte. Therewith, the overall generated traffic rate of all sources is about 3200 kBit/s and only half as large as for the evaluations in section 5.6.2. The below illustrated results always present the averages out of ten independent simulations with varying traffic and mobility traces, additionally they depict the respective 95% confidence intervals. The simulations evaluate the performance of AODV within various scenarios. The evaluation is based on the already introduced metrics packet loss p, overall throughput T, and fairness indicator F. It does not evaluate routing overhead, because it does not reveal new insights.

The first evaluation in Figure 98 focuses on the impact of different urban scenarios on the performance of radial symmetric ad hoc networks. The figures depict results from simulations with 200 nodes and a common load of twenty packets per second and source. Obviously, the results for the free space environment show promising results. The fair-MAC packet loss is close to zero, but the throughput is about 15% below the throughput achieved by the legacy-MAC. The fairness index is not depicted, but for these conditions it is close to 100%. In contrast to that, the legacy-MAC generates about 15% packet loss and the fairness is lowered. The results change significantly, when considering the Munich and the Manhattan scenario. The legacy-MAC again achieves higher overall throughputs, but it causes 50% and more packet loss while the fairness drops to only 60%. Obviously, the fair-MAC extension again achieves fewer throughputs, but both other metrics remain close to their optimum. The Italy scenario is

unique in this way that the fair extension causes more than 20% packet loss, but achieves with 1050 kBit/s an equivalent throughput in comparison to the legacy-MAC.

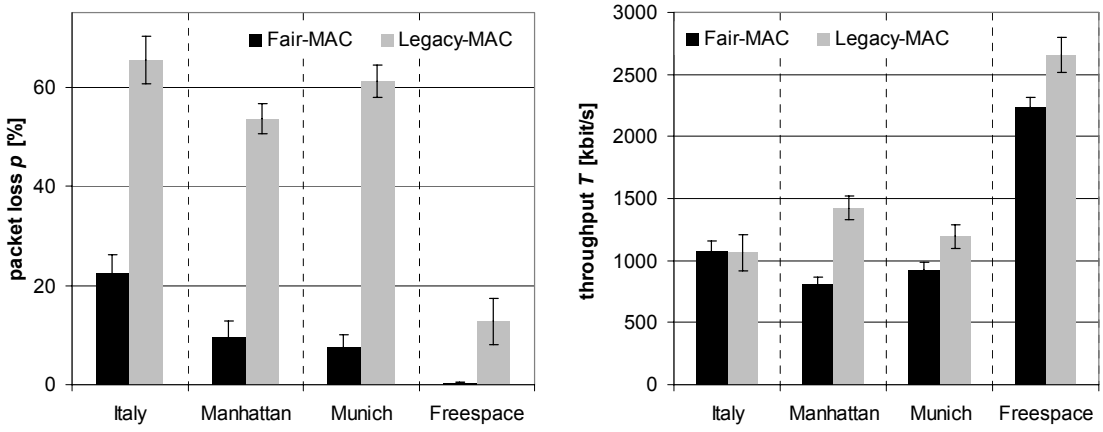

Figure 98: Packet loss and throughput for a 200 nodes network within varying urban environments.

These first results already impressively depict the ability of the fair-MAC algorithm to perform well within urban environments. At the same time, the existing MAC protocol shows severe inabilities to maintain the requested connections within non-flat environments.

Figure 99 presents packet loss and throughput results for varying network sizes. Assuming the free space scenario, the legacy-MAC achieves perfect results within networks with less than 100 nodes. The packet loss is minimal and the fairness is close to 100%. The achieved overall throughput emphasizes this rating. With a data rate greater than 3100 kBit/s, it is close to the offered load of 3200 kBit/s. However, for 200 nodes networks, the achieved throughput degrades to only 2700 kBit/s and the packet loss rises to 15%. With an average route length of 2.3 hops, the number of multi-hop connections within these networks is significantly higher than in the smaller network evaluations. Therefore, the network is unable to forward all packets, and consequently packets are dropped due to congestions. Within the Munich scenario, the achieved throughput only reaches 1600 kBit/s at most, and further degrades for increasing network sizes. Simultaneously, the packet loss reaches unacceptable regions.

The results are significantly different for simulations with the fair-MAC extension. While it achieves just 10-15% less throughput, the caused packet loss remains below 10%. Especially within the free space scenario, the loss rate is close to zero. This impressively illustrates the ability of the fair-MAC protocol to reliably predict the overall achievable throughput and to inform the individual sources about their maximal utilizable packet rates.

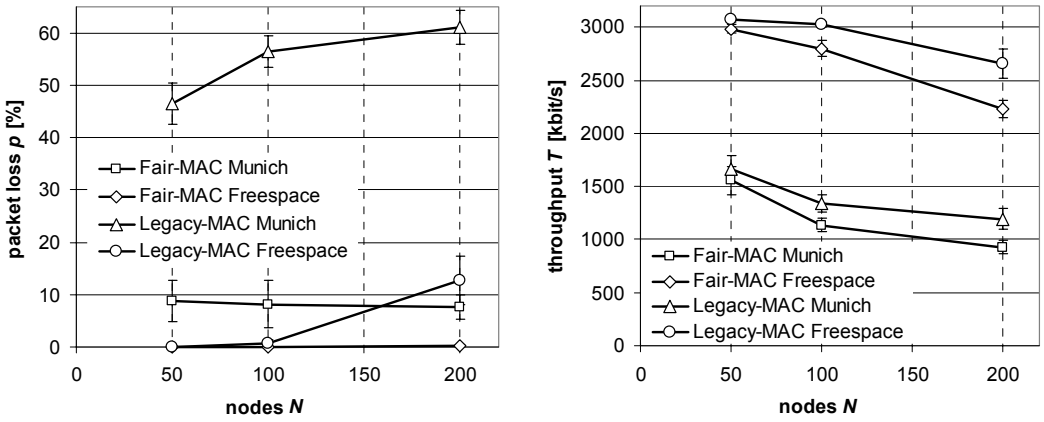

Figure 99: Packet loss and throughput for varying network sizes within two different environments.

The results in Figure 100 and Figure 101 depict the evaluations with varying source packet rates. The individual packet rates rise from only five packets per second to fifty, while the network size remains constant with 100 nodes. Figure 100 separately presents the packet loss results for both MAC protocols and for all four different urban scenarios. Due to the significantly varying results, the scaling on the y-axis differs in both figures. As expected, within the free space scenario, both MAC protocols perform well. At a rate of 5 and 10packets per second, they scale and achieve optimal results. With further increasing network loads, the packet losses rise as well. However, the high load condition causes only 3% packet loss with the fair-MAC extension, but almost 40% with the legacy-MAC protocol. Examining the simulations with the legacy-MAC in urban environments, results become worse. Only with rates of five packets per second, the network reaches reasonable packet loss ratios. For higher network loads, the protocol is unable to accomplish practicably usable packet loss results. This is independent from the underlying urban scenario.

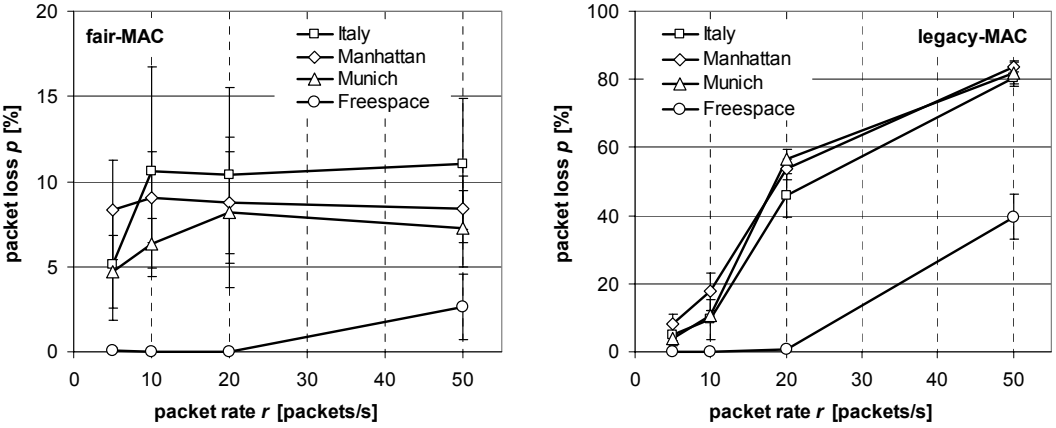

Figure 100: Packet loss results different packet rates for both MAC protocols in varying urban environments.

However, when analyzing the results obtained by the fair-MAC extension, its superiority over the legacy-MAC becomes obvious. The simulations within the Munich and the Manhattan scenario reveal an average packet loss of less than 10%, independent from

the packet rate. Only within the synthetic Italy scenario, the packet loss is 12% and therefore greater than the 10% limit. This again illustrates the ability of the protocol to distribute the maximal available throughputs and therewith allow the limitation of exceeding source data rates.

After the presentation of the packet loss results, Figure 101 depicts the achieved throughputs and the fairness within the Munich and the free space scenario. Analyzing the results, it attracts attention that the Munich scenario rigorously limits the achievable throughput. The throughput only reaches 1300 kBit/s. In contrast, the free space scenario permits combined throughputs of up to 4800 kBit/s. As already mentioned, the legacy-MAC achieves up to 20% higher throughputs. These throughputs are greater than those in section 5.6.2. This becomes apparent when considering the significantly higher node densities used for these evaluations. As a result, the average route length is much shorter, which essentially allows higher network capacities and throughputs. Even so, it is evident that source data rates greater than 20 packets per second already exceed the network capacity. While the network scales well with the offered load, the legacy-MAC is able to keep the individually achieved throughputs almost equivalent. Consequently, the fairness index is beyond 90%. However, with higher loads and more challenging network conditions, the fairness index drops significantly. Especially if the simulated network must cope with the Munich environment, only a single connection accounts for the entire throughput. Due to the definition of the fairness index, *F* drops to almost zero.

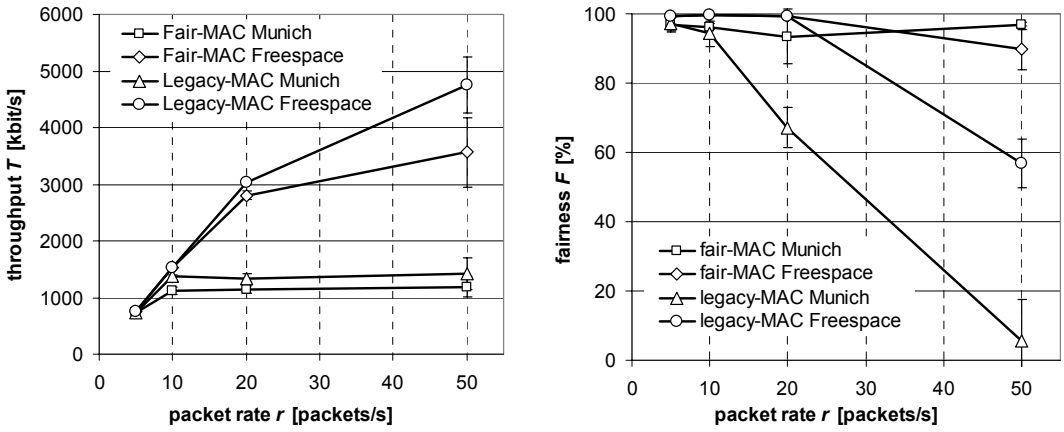

Figure 101: Throughput and fairness with respect to varying packet rates and environments.

In contrast to that, the fair-MAC extension keeps the fairness among different flows above 90%. As described in the previous section, the medium access is always more difficult towards the AP, and consequently these flows achieve less throughput. Therefore, fairness indices above 90% indicate that flows in the same direction achieve perfectly equivalent throughputs.

As a summery, the fair-MAC extension is able to uniformly share the available data rate among all available sources, although urban environments are significantly more challenging than line-of-sight scenarios.

5.7 Summary

The utilization of ad hoc networks without any connection to the fixed infrastructure is possible. However, it is not likely that gateways connecting to the Internet are unreachable, or all nodes solely communicate with other ad hoc participants. Therefore, the usage of the network to extend the service area of APs is probably the most common usage scenario. With varying scenarios the network load distribution alters as well. Within pure ad hoc networks, each node is source or sink with equal probability and therefore the load is also perfectly uniformly distributed. However, if an AP is present in the center of the network, interior nodes must forward data to and from exterior nodes, and therefore observe higher loads. Consequently, the load is radial symmetric distributed.

Simulations illustrate that the existing legacy-MAC has very balanced per-node fairness. While this is optimal for pure ad hoc networks, it shows significant performance degradations with non-uniformly distributed load. Exterior nodes do not recognize that interior nodes are overloaded. They continue to generate packets while interior nodes already drop packets. End-to-end flow control mechanisms like TCP are unable to overcome this shortcoming. If a TCP control realizes a packet loss, it reduces its throughput, allowing others to increase their throughput. This amplifies the misbehavior. The already reduced flows must reduce their throughput even further. Without valuable maximal throughput information from underlying layers, TCP creates oscillating individual flow throughputs in dynamic networks.

An analysis revealed two independent shortcomings, namely the queuing of the 802.11 MAC protocol and the distributed access scheme. Both shortcomings occur even in small networks and independent from a dynamic network topology. With only a single queue, the analysis illustrates that nodes favor own packets, while packets received from other nodes experience higher dropping probabilities. Two queues within each node overcome the unfairness. The first queue manages packets from the node's own higher layer while the second queue is solely responsible for forwarding traffic. Both queues independently compete for medium access. This improves the fairness, but does not achieve full equality among flows. Therefore, the newly developed fair-MAC extension piggybacks throughput information within packets. Especially the number of flows and the number of hops to the AP are relevant. The AP continuously evaluates the received information and calculates the globally optimal access rates. With the help of the already exchanged MAC ACK packets, nodes receive the information, and calculate their own individually optimal access rates. A cross-layer approach allows the MAC layer to notify higher layers about the maximal achievable throughput. Therewith, nodes prevent packet drops in their own queues.

Simulations illustrate that the extension is able to distribute the optimal parameters within 200 ms. It achieves this quick adaptation without the necessity to generate additional packets and therefore keeps the overhead minimal. As additional advantage, the extension is fully transparent for upper layers, and therefore modifications of higher layer protocols are unnecessary. Results depict that the fair-MAC extension outperforms the legacy-MAC under high load conditions. This is independent from the velocity, the packet rate, and the network size. While the overall throughput of the fair-MAC is reduced by about 10% in comparison to the legacy-MAC, the fairness among different network flows is almost optimal. Networks utilizing the legacy-MAC protocol favor sources with direct links to the AP. While those nodes achieve maximal throughput,

other nodes are almost cut from data exchange with the AP. In contrast to that, networks using the fair-MAC extension evenly spread the available bit rate and allow all nodes to communicate with the AP, independent from their distance to the APs. A positive side effect is that the extension significantly reduces the packet loss of the network. It prevents overload situations in which the data traffic load exceeds the network capacity.

Finally, simulations in urban scenarios emphasize this conclusion. The superior performance of the fair-MAC extension is independent from the underlying urban scenario. Its achieved throughputs are again reduced by 5-10%, whereas the packet loss remains constant and below 10%. This allows even the utilization of QoS critical applications like streaming and VoIP.

6 Conclusion

Ad hoc networks form a completely novel approach and provide support for novel scenarios. It opens new possibilities and equality among network participants and the established telecommunication companies must rethink their existing business models. However, it also comprises novel challenges and yet unsolved questions. Therefore, much research is necessary to understand the theoretical background as well as to develop well-performing algorithms and protocols.

This thesis covers three sources of possible performance improvements within mobile ad hoc networks. The lifetimes of arbitrary ad hoc paths are considered important for the network performance, but up to now, they are not examined in-depth. Whereas the behavior of ad hoc networks within urban environments and the limiting factors of networks around access points are new research areas. Besides simulative evaluations, the thesis makes use of theoretical methods and tools as well. Therewith, it is of practical and theoretical relevance for researchers as well as for developers working in the area of ad hoc networks.

Initially, the thesis theoretically analyzes **lifetimes of ad hoc network paths**. The approach models arbitrary path lengths utilizing independent mobile nodes. The mathematical evaluation reveals that the probability of path lifetimes decreases negative exponentially. Routes over more than six hops show very short average path lifetimes, and therewith prevent reasonable networking. A possibility to overcome this limitation or to generally improve the overall performance is the utilization of multipath routing strategies. The analysis illustrates that disjoint backup routes have the potential to lengthen the average path lifetime by about 50%. However, the positive effect lessens with every additional path.

The subsequent simulations extend the model and evaluate four different routing strategies. Besides the theoretically examined shortest path and disjoint multipath strategies, it supports non-disjoint paths and multipath obtained with the help of the flooding mechanism. Evaluations reveal that the non-disjoint multipaths routing strategy outperforms all other algorithms. This is independent from the network size and the number of discovered multipath. However, the node density alters the results and other multipath strategies show equivalent results.

The theoretical routing overhead analysis illustrates that the shortest path strategy obviously induces least overhead, while multipath routing strategies cause some additional overhead. Only source routing strategies like DSR induce significantly more overhead. When considering maximal path lifetimes with reasonable additional overhead, the non-disjoint routing strategy depicts the best available routing strategy for most scenarios. Only for highly dynamic scenarios, like cars on highways, the shortest path algorithm presents the optimal choice.

The second contribution focuses on the performance of **ad hoc network protocols in urban environments**, because previous evaluations only considered flat environments. These artificial scenarios do not allow statements about network performance in real-

world scenarios. Therefore, the accomplished evaluations include urban environments as well. The chosen Walfisch-Ikegami propagation model has a balanced relation between computational complexity and prediction accuracy. Due to the building deployment, existing mobility models are inappropriate. The novel city-motion mobility model interacts with the buildings and consequently node movements show realistic characteristics.

The evaluation is based on three different urban environments to minimize statistical effects. Urban scenarios imply much higher network dynamics, and connections commonly require longer paths. As known from previous considerations, longer paths also lead to considerable performance degradations. The evaluation emphasizes that reasonable networking is only possible for routes with few hops and with significantly lowered overall network load. Results illustrate that the conditions of urban environments turn ad hoc networks unmanageable even for small network sizes. To overcome these shortcomings, different solutions are discussed. One possibility is the introduction of time-to-live fields within route request packets to limit the maximal hop length, which emphasize the evaluation from the previous chapter. Another option to reduce the network dynamic is the detection of low mobility or even static nodes.

Chapter 5 focuses on the **performance of ad hoc networks around access points**. A protocol evaluation of the existing WLAN 802.11 MAC reveals its inability to achieve reasonable performance. An in-depth analysis exposes two independent shortcomings. Both the queuing algorithm and the balanced per-node fairness are appropriate for uniformly distributed networks, but fail if radial symmetric loads are present. The legacy-MAC favors sources close to the AP, while at the same time cuts off the communication attempts of all other sources.

The newly developed fair-MAC extension is able to overcome both limitations. It uses two queues and nodes continuously exchange information about individually optimal access rates. The extension is transparent and consequently it does not necessitate modifications at higher layer protocols. Access rate information are piggybacked on data packets which keeps the overhead as low as possible. Simulations illustrate its outstanding performance in comparison with the legacy-MAC protocol. It achieves completely fair throughput distributions among individual flows, while lowering the overall throughput only by 10%. Additionally, the exactly determined access rates lead to minimal packet losses.

The final test is the utilization of the extension within urban environments. Results emphasize the previous findings. The throughput is reduced, but the fairness among individual flows is greatly increased. The low packet loss even allows the utilization of QoS critical applications like streaming or VoIP.

The consideration of all these performance-limiting factors will allow the development or redesign of ad hoc network protocols, which are able to cope with arbitrary scenarios. Such redesigned protocols will outperform existing protocols with respect to packet loss and fairness. As a matter of fact, some nodes will experience limited services, but the remaining participants will significantly benefit and on the whole, user satisfaction rises. I am confident that despite the remaining challenges and shortcomings, ad hoc networks present unique possibilities and scenarios. On the medium term, ad hoc networks will replace currently existing wireless solutions due to their greater adaptability and simplicity.

A Abbreviations

AAA	Authentication, Authorization, and Accounting
ACK	Acknowledgement
AODV	Ad hoc On-Demand Distance Vector Routing
AODV-DMP	Non-Disjoint Multipath AODV
AODVM	AODV Multipath
AP	Access Points
CAC	Call Admission Control
CBR	Constant Bit Rate
cdf	Cumulative Density Function
CM	City Motion Mobility Model
CTS	Clear-To-Send
DBF	Distributed Bellman Ford
DCF	Distributed Coordination Function
DLC	Data Link Control Layer
DMP	Disjoint Multipath Routing
DSN	Destination Sequence Number
DSR	Dynamic Source Routing
EDCF	Enhanced Distributed Coordination Function
FL	Flooding Based Routing
FS	Free Space Model
GPS	Global Positioning System
GSR	Global State Routing
HCF	Hybrid Coordination Function
ISM	Industrial-Scientific-Medical Band
ITU	International Telecommunication Union
LL	Link Layer
LLT	Link Lifetime
LOS	Line-Of-Sight
MAC	Medium Access Control

MANET	Mobile Ad Hoc Networks
NDM	Non-Disjoint Multipath Routing
NLOS	Non-Line-Of-Sight
OM	Obstacle Mobility Model
PCF	Point Coordination Function
pdf	Probability Density Function
PHY	Physical Layer
PLT	Path Lifetimes
RD	Random Direction Mobility Model
RDER	Route Discovery Error
RERR	Route Error Packets
RRCM	Route Confirmation Message
RREP	Route Reply
RREQ	Route Request
RTS	Ready-To-Send
RWP	Random Waypoint Mobility Model
SM	Shadowing Model
SP	Shortest Path Routing
TCP	Transmission Control Protocol
TMT	Theoretical Maximal Throughput
ToS	Type of Service
TRG	Two Ray Ground Model
TTL	Time-To-Live
UDP	User Datagram Protocol
UWB	Ultra Wide Band
WEP	Wired Equivalent Privacy
WIM	Walfisch Ikegami Model
WLAN	Wireless Local Area Networks
WPA	Wi-Fi Protected Access
ZRP	Zone Routing Protocol

B Terminology

In order to achieve a common understanding of significant phrases and terms, the most important ones are described in the following:

- *Node*: A participating device within an ad hoc network.
- *Link*: Wireless interconnection between two adjacent neighbors from a medium access layer perspective. Links break in case nodes move out of each other radio range, or the antenna of the transmitter or receiver is turned off.
- *Connectivity*: The number of direct neighbors of a node determines its connectivity. High average node connectivity ensures full network connectivity and prevents isolated nodes.
- *Topology*: The topology determines the structure of the network on a link basis. It continuously changes because of the mobility of nodes.
- *Hop and Multi-Hop:* A hop describes the interconnection between adjacent nodes from a network layer perspective. A multi-hop system consists of several independent but contiguous hops. Besides the hops towards the terminating nodes, every hop has a predecessor and a successor hop. Over a multi-hop system, data packets are wirelessly transmitted multiple times.
- *Path and Route*: Consists of one or several hops, and connects distant participating nodes from a network layer perspective. Paths break whenever one or several links contained within the paths break. The network layer protocol is responsible to reestablish the path or to create a new one between the nodes.
- *Path set:* Consists of multiple paths between the same source and destination. Individual paths within a set can be node or link disjoint, or utilize equal nodes and links. However, a set never contains multiple equivalent paths.
- *Flow*: A flow is a transport layer data stream created by the source node and bound for the destination node. A flow may contain data packets (e.g. UDP or TCP) with arbitrary packet sizes and generating rates.
- *Connection*: Communication setup between two distant nodes from a higher layer perspective. The connection is uninterrupted, and remains valid as long as one or both nodes need to exchange data.
- *Infrastructure*: Network entities which are necessary to maintain the network, while they do not act as sources or sinks for communication. It is usually managed by network providers and network participants are commonly charged for utilizing the services provisioned by the infrastructure.

C Necessary Simulation Duration and Repetitions

Previous evaluations of ad hoc network algorithms consider different scenarios with varying traffic and mobility patterns. They also take into account the physical, the link-, and sometimes the transport-layer and classify existing routing algorithms based on certain metrics. However, they commonly use predetermined simulation durations of 300, 600 or 900 seconds and evaluate the obtained results after the simulation finished. They do not consider the simulation time as crucial parameter for the accuracy of the evaluation. Due to mobility and local node connectivity effects, the monitored metrics significantly vary over time. Therefore, the defined simulation time could be too short for reasonable results. Though, it remains unanswered, how long simulations must run to permit reasonable conclusions on the obtained results. To overcome this uncertainty a minimal simulation time is determined to achieve a sufficient accuracy. As most publications use the ns-2 as simulation tool, all following considerations focus on the ns-2 as well. It turns out that well-arranged ns-2 simulation setups are able to minimize the computational effort while preserving or even improving accuracy.

The other inadequately addressed shortcoming is the number of necessary simulation repetitions. Evaluations in earlier publications generally calculate the average out of ten or twenty comparable simulations. However, ongoing evaluations depict that results significantly vary with the initial network topology, although the relevant startup parameters are equivalent. It remains uncertain how many consecutive repetitions are necessary to reach a certain confidence in the achieved results. Therefore, the definition of a necessary number of simulation repetitions is essential. In the following, the observed results are briefly described, however more details can be found in [5].

C.1 Simulation environment

Monitored events must occur statically independent to realistically determine necessary evaluation times of ad hoc network simulations. Consequently, all input parameters for ns-2 simulations must have appropriate characteristics. The mobility model must guarantee a uniformly distributed node density for all points in time and the traffic model must allow a constant network load.

As described in chapter 2.5.1, the random direction (RD) mobility model perfectly fulfills the required characteristics. The commonly used ad hoc traffic generator for ns-2 simulations creates a constantly increasing network load over time. Obviously, this increasing load also modifies the monitored metrics and violates the requirement of time invariant event probabilities. Therefore, a different traffic generator is necessary, which keeps the network load as constant as possible. The new model initiates 20 simultaneous constant bit rate (CBR) flows. Each flow transmits 20 packets/s and each packet has a payload of 512 Bytes. Therefore, the average network load is 400 packets every second.

Each flow is again separated in several independent connections with alternating source-destination combinations. The first source-destination pair chooses a startup time between 0 and t_c and starts emitting packets. Consequently, the warm up phase lasts t_c seconds and thereafter the network load is approximately constant. All connection durations are randomly chosen from the time interval [0, t_c]. After a 5 second cool down phase, the next source starts transmitting its packets and keeps the connection for another period. With t_c = 100 seconds, the average holding time of a connection is about 50 seconds.

As always, the network simulator ns-2 version 2.1b9a is used as simulation tool. Variable simulation parameters for different scenario files are the size of the simulation area, the number of nodes, and their maximal speed. The node density is constant in all scenarios and every node has about 10 neighbors on average. This prevents nodes without links to the rest of the network or even separated network parts. Otherwise monitored events heavily depend on the current network topology. The evaluated scenarios contain either 50 or 100 nodes within simulation areas of 1000×1000 m^2 and 1400×1400 m^2, while the maximal node velocities is either 1 m/s or 10 m/s. Because AODV is the mainly considered routing algorithms within this thesis, it is as well used for the following evaluations.

C.2 Sufficient simulation times

The most important parameters in ad hoc network simulations are the packet loss and the achieved throughput. However, the throughput again depends mainly on the packet loss. Therefore, the packet loss p is chosen as monitored metric. It is defined as quotient between received and send data packets.

$$p = 1 - \frac{\rho_{recv}}{\rho_{send}} = 1 - \frac{recv.\,packets}{send\ packets} \tag{97}$$

Figure 102 and Figure 103 illustrate simulations with 900 seconds duration to screen the packet loss distribution over time. They depict the average packet loss and the short-term dependence. Initially accomplished simulations indicate the possibility of a premature termination of simulations. Figure 102 clearly indicates that after the 100 second warm up phase, the packet loss does not dramatically alter anymore. Only two out of 20 simulations (including the scenario in Figure 103) present behaviors, in which the packet loss after 900 seconds is clearly altered compared to the packet loss after 400 seconds.

To allow an early termination in case the mean packet loss remains constant, the network simulator ns-2 were extended with statistical analysis functions from the Communication Networks Class Library (CNCL) [212]. CNCL monitors the behavior of a variable metric, and in case its mean reaches sufficient accuracy, it suggests the premature termination of the simulation runs. As already mentioned, all scenarios use a 100 seconds warm-up phase, consequently CNCL also starts with its packet loss measurements after 100 seconds.

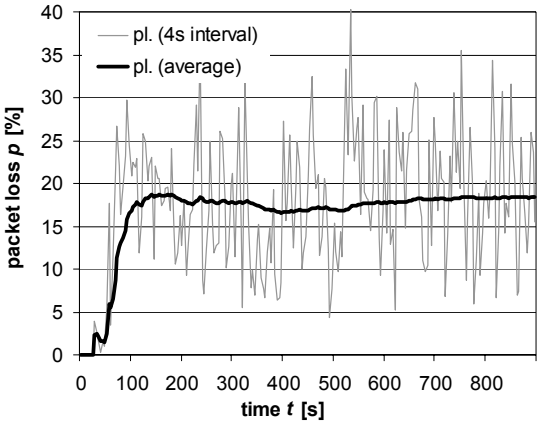

Figure 102: Packet loss distribution for a simulation with 100 nodes and 10 m/s max. node velocity.

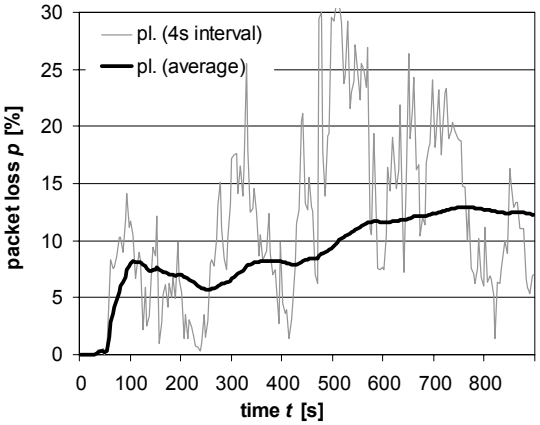

Figure 103: Packet loss distribution for a simulation with 100 nodes and 1 m/s max. node velocity.

Figure 104 presents the average packet losses for all four scenarios. In order to be able to indicate possible differences, only three simulations per scenario were carried out. Obviously, the packet losses of full length simulations and CNCL controlled simulations do not vary significantly. The mean deviation from the full length results are mostly below 10%. Two out of four scenario setups show deviations below 2%. The mean packet loss between full length and CNCL simulations with 50 nodes and 10 m/s node velocity differs by 9%, whereas 100 nodes and 1 m/s maximal velocity lead to a deviation of 25%. However, two out of three simulations for this parameter set generated comparable results. CNCL based results are almost equal to those after full length simulations. But the third result shows great deviations.

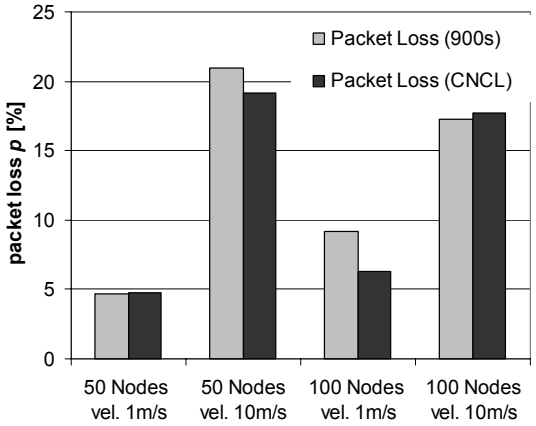

Figure 104: Comparison of mean packet losses of full length and CNCL controlled simulations.

The CNCL controlled simulation finished after 190 seconds with a mean packet loss of 5.7%, whereas the full length simulations presents a packet loss which differs by a factor of 2.5 (12.5%). Figure 103 emphasizes this behavior. It indicates that the packet loss after 300 seconds is 6.6%, while the average packet loss of the last 600 seconds is 14.7%. Obviously, CNCL is unable to forecast such significant packet loss variations.

Results based on CNCL cannot guarantee that they are equivalent to those results after 900 seconds. However it also remains questionable which result is more significant. Nevertheless, all other packet loss distributions show a more predictable behavior and CNCL is able to accurately forecast correct packet losses. Consequently, all simulations for this thesis use as basis the above stated simulation parameters and settings and rely on the CNCL extension.

C.3 Necessary numbers of simulations

After the specification of reasonable simulation times, the following section determines the necessary number of consecutive simulations to achieve sufficient confidence in the mean packet loss. The size of the confidence interval around the mean $\overline{x}$ of a random process is calculated as function of the variance s^2 and the mean. With n the number of measurements and x_i $(i \leq n)$ an independent result from the series, the variance s^2 computes as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{98}$$

The error probability α and the degree of freedom n-1 determine the confidence interval δ as value of the Student-t distribution $t_\alpha(n\text{-}1)$. The confidence interval indicates that the true average $\hat{x}$ of a random process is with an probability α within the interval $[\overline{x}-\delta, \overline{x}+\delta]$. The confidence interval δ is given as

$$\delta = \frac{s}{2\cdot\sqrt{n}} t_\alpha(n-1) \tag{99}$$

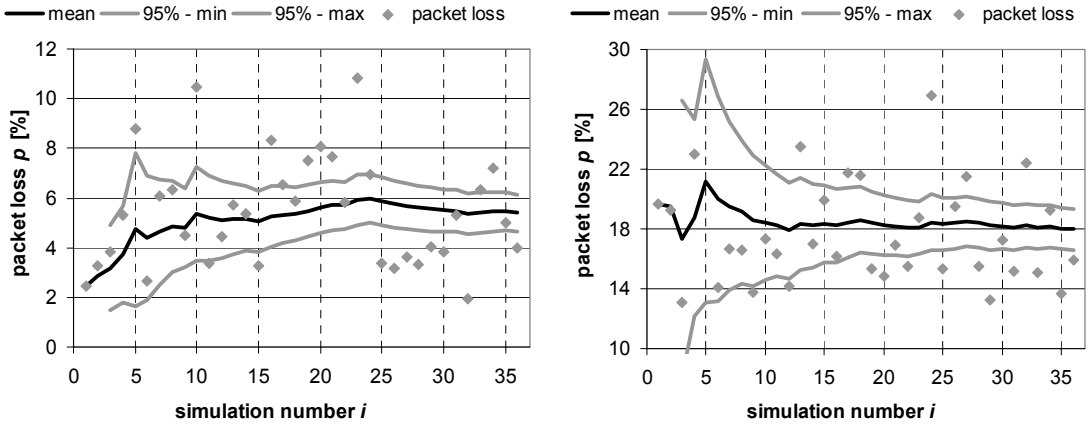

Figure 105: Progress of mean and confidence intervals for AODV simulations left: 50 nodes and 1 m/s; right: 100 nodes and 10 m/s.

Evaluations consist of 36 simulations with varying traffic and mobility traces. The confidence interval calculations always utilize an error probability α of 95%. Figure 105 shows two examples for behaviors of the mean packet loss and its confidence interval for increasing number of simulations. The dark line indicates the progress of the mean, the grey lines are the bounds for the upper and lower 95% confidence interval, and the dots are the respective simulation results. The example with 50 nodes and 1 m/s maximal node velocity shows that about 15 independent simulations are necessary to

calculate the mean with sufficient accuracy. After 15 simulations, the mean packet loss is 5.0% with a confidence interval ranging from 3.8% to 6.3%. For increasing numbers of consecutive simulations, the mean remains almost constant but the confidence interval continues to shrink.

The second simulation scenario in Figure 105 contains 100 nodes with a maximal node velocity of 10 m/s. Certainly the packet loss is an order of magnitude larger than in the previous example. However, the mean packet loss remains almost constant for more than 15 simulations. The mean packet loss is 18.3% after 15 simulations and has a 95%-confidence interval δ of 2.6%. Additional scenarios and evaluations can be found in [5]. As a conclusion, scenarios with mean packet losses of less than 20% require at least 10 simulations to reach confidence values of less than 3%. Scenarios with higher average packet losses require more simulations to reach the same confidence. However the possible relative deviation in comparison to the average packet loss is not as significant for reasonable performance statements. Therefore, all performance evaluations within this thesis are based on at least ten independent simulations in order to achieve a sufficient confidence in the made statements.

Bibliography

Own publications are depicted in bold.

[1] G. E. Moore, "Cramming More Components Onto Integrated Circuits," in *Electronics*, vol. 38, April 1965.

[2] L. Kleinrock, "Nomadic Computing - An Opportunity," *ACM SIGCOMM Computer Communication Review*, vol. 25, January 1995.

[3] M. Frodigh, P. Johansson, and P. Larsson, "Wireless Ad Hoc Networking: The Art of Networking without a Network," in *Ericsson Review*, 2000, pp. 248-263.

[4] J. Jubin and J. D. Tornow, "The DARPA Packet Radio Network Protocols," *IEEE Special Issue on Packet Radio Networks*, vol. 75, pp. 21-32, January 1987.

[5] **I. Gruber** and H. Li, "Sufficient Evaluation Times and Necessary Repetitions for Reasonable Ad Hoc Network Simulations," presented at International Workshop on Mobile Ad Hoc Networks and Interoperability Issues (MANETII'04), Las Vegas, USA, June 2004.

[6] **I. Gruber**, "Link Availability Times in Mobile Ad Hoc Networks," presented at 8th EUNICE Open European Summer School (EUNICE'02) and the IFIP Workshop on Adaptable Networks and Teleservices, Trondheim, Norway, September 2002.

[7] **I. Gruber**, A. Baessler, and H. Li, "Fair WLAN Scheduling for Ad Hoc Networks with Access Points (Poster)," presented at ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'04), Tokyo, Japan, May 2004.

[8] **I. Gruber** and S. Högg, "Experimental Results with a GPS and Signal Strength extended Ad Hoc Routing Protocol," presented at IEEE Workshop on Wireless Local Networks (WLN'03), Bonn, Germany, October 2003.

[9] **I. Gruber**, O. Knauf, and H. Li, "Performance of Ad Hoc Routing Protocols in Urban Environments," presented at European Wireless 2004, Barcelona, Spain, February 2004.

[10] **I. Gruber** and H. Li, "Behavior of Ad Hoc Routing Protocols in Metropolitan Environments," presented at IEEE Vehicular Technology Conference Fall 2004 (Fall VTC'04), Los Angeles, USA, September 2004.

[11] **I. Gruber** and H. Li, "Link Expiration Times in Mobile Ad hoc Networks," presented at IEEE Local Computer Networks (LCN) 2002, Tampa, USA, November 2002.

[12] **I. Gruber** and H. Li, "Path Expiration Times in Mobile Ad Hoc Networks," presented at IEE European Personal Mobile Communications Conference (EPMCC'03), Glasgow, Great Britain, April 2003.

[13] **I. Gruber** and C. Mathiesen, "The Broadcast based Ad Hoc Routing Protocol (BCBR) - A Novel Approach for Ad Hoc Routing," presented at 18th International Teletraffic Congress (ITC 18), Berlin, Germany, August 2003.

[14] **I. Gruber**, R. Schollmeier, and W. Kellerer, "Peer-to-Peer Communication in Mobile Ad Hoc Networks," *Ad Hoc & Sensor Wireless Networks (OCP Science Journals)*, vol. 1, May 2005.

[15] R. Schollmeier, **I. Gruber**, and M. Finkenzeller, "Routing in Mobile Ad Hoc and Peer-to-Peer Networks. A Comparison," presented at Networking 2002, International Workshop on Peer-to-Peer Computing, Pisa, Italy, May 2002.

[16] D. Yu, H. Li, and **I. Gruber**, "Path Availability in Ad Hoc Networks," presented at 10th International Conference on Telecommunications ICT'2003, Tahiti, French Polynesia, February 2003.

[17] H.-M. Zimmermann, **I. Gruber**, and C. Roman, "A Voronoi-based Mobility Model for Urban Environments," presented at European Wireless (EW'05), Nicosia, Cyprus, April 2005.

[18] M. Zitterbart, K. Weniger, O. Stanze, and **I. Gruber** et. al., "IPonAir - Drahtloses Internet der naechsten Generation," *PIK Themenheft Mobile Ad-hoc-Netzwerke*, September 2003.

[19] **I. Gruber**, "Evaluation of Ad Hoc Routing Strategies to Maximize Path Lifetimes," presented at IST Mobile & Wireless Communication Summit, Dresden, Germany, June 2005.

[20] **I. Gruber** and H. Li, "Patent: Routing Method for an Ad Hoc Network," Siemens AG, Ed. Munich, Germany, September 2004.

[21] **I. Gruber** and H. Li, "Patent: Method and Base station for the Transmission of Information in a Cellular Radio Communication System extending by means of Ad Hoc Connections," Siemens AG, Ed. Munich, Germany, September 2004.

[22] **I. Gruber**, A. Bäßler, and H. Li, "Patent Pending: Flow Fairness in Distributed Ad Hoc Networks with Access Point," Siemens AG, Ed. Munich, Germany, 2004.

[23] **I. Gruber**, A. Bäßler, and H. Li, "Patent Pending: Verfahren zur Übertragung von Datenströmen," Siemens AG, Ed. Munich, Germany, 2004.

[24] W. Kellerer, R. Schollmeier, **I. Gruber**, and F. Niethammer, "Patent Pending: Mobile Peer-to-Peer Networking," NTT DoCoMo Euro-Labs, Ed. Munich, Germany, 2004.

[25] W. Roush, "10 Emerging Technologies That Will Change the World," in *MIT Technology Review*, February 2003.

[26] C. Santivanez, B. McDonald, I. Stavrakakis, and R. Ramanathan, "On the Scalability of Ad Hoc Routing Protocols," presented at IEEE INFOCOM, New York, USA, June 2002.

[27] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388-404, March 2000.

[28] S. Shenker, C. Partridge, and R. Guerin, "RFC 2212 - Specification of Guaranteed Quality of Service," Internet Engeneering Task Force (IETF) September 1997.

[29] M. H. Ahmed, "Call Admission Control in Wireless networks: A Comprehensive Survey," *IEEE Communications Surveys and Tutorials*, vol. 7, pp. 2-21, April 2005.

[30] M. Möske, H. Füßler, H. Hartenstein, and W. Franz, "Performance measurements of a vehicular ad hoc network," presented at IEEE Vehicular Technology Conference (Spring VTC'04), Milan, Italy, May 2004.

[31] K. Weniger and M. Zitterbart, "Address Autoconfiguration in Mobile Ad Hoc Networks: Current Approaches and Future Directions," *IEEE Network Magazine Special issue on Ad hoc networking: Data Communications & Topology Control*, July 2004.

[32] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," presented at IEEE INFOCOM, Hong Kong, China, March 2004.

[33] J. Eberspächer, H.-J. Vögel, and C. Bettstetter, *GSM - Switching, Services and Protocols*, 2nd ed. Chichester, England: John Wiley & Sons, Ltd., March 2001.

[34] L. Zhou and Z. J. Haas, "Securing Ad Hoc Networks," *IEEE Network*, vol. 13, pp. 24-30, November 1999.

[35] J.-P. Hubaux, L. Buttyan, and S. Capkun, "The Quest for Security in Mobile Ad Hoc Networks," presented at ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'01), Long Beach, USA, October 2001.

[36] H. Yang, H. Luo, F. Ye, S. Lu, and L. Zhang, "Security in Mobile Ad Hoc Networks: Challenges and Solutions," *IEEE Wireless Communications*, vol. 11, pp. 38-47, February 2004.

[37] DAIDALOS, "Designing Advanced Network Interfaces for the Delivery and Administration of Location Independent Optimized Personal Service, from www.ist-daidalos.org," April 2005.

[38] IEEE Standards Department, *ANSI/IEEE Std. 802.16: Air Interface for Fixed Broadband Wireless Access Systems*. New York, USA, October 2004.

[39] European Telecommunications Standards Institute, *ETSI Standard TS 102 177 V1.2.1: Broadband Radio Access Networks (BRAN): HIPERMAN: Physical (PHY) layer*. Sophia-Antipolis Cedex, France, January 2005.

[40] D. J. Johnston and M. LaBrecque, "IEEE 802.16 WirelessMAN Specification Accelerates Wireless Broadband Access," in *Technology@Intel Magazine*, August 2003.

[41] The WiMAX Forum, "from http://www.wimaxforum.org." Vista, USA, April 2004.

[42] IEEE Standards Department, *ANSI/IEEE Std. 802.11: Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Network - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1999 Edition (ISO/IEC 8802-11: 1999)*. New York, USA, June 2003.

[43] European Telecommunications Standards Institute, *ETSI Standard TS 101 475 V1.3.1: Broadband Radio Access Networks (BRAN): Hiperlan 2: Physical Layer*. Sophia-Antipolis Cedex, France, December 2001.

[44] S. Xu and T. Saadawi, "Does the IEEE 802.11 MAC protocol work well in multihop ad hoc networks?" *IEEE Communication Magazine*, vol. 39, pp. 130-137, June 2001.

[45] IEEE Standards Department, "IEEE 802.11 WG & Activities, from http://grouper.ieee.org/groups/802/11," in *IEEE Standards for Information Technology*. New York, USA, April 2005.

[46] A. Gershman and N. Sidiropoulos, *Space-Time Processing for MIMO Communications*. Chichester, England: John Wiley & Sons, Ltd., April 2005.

[47] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311-335, 1998.

[48] M. Bludszuweit, "Mobile Communication Speed Record: One Gigabit per Second over the Air, from http://www.siemens.com." Munich, Germany: Siemens AG, December 2004.

[49] C. Hartmann, "Wireless Access with Smart Antennas and SDMA," presented at EUNICE'99, Fifth EUNICE Open European Summer School, Barcelona, Spain, September 1999.

[50] B. A. Miller and C. Bisdikian, *Bluetooth Revealed 2nd Edition*. Englewood Cliffs, USA: Prentice Hall Inc., 2001.

[51] The ZigBee Alliance, "http://www.zigbee.org," April 2004.

[52] IEEE 802.15.3 Task Group, "802.15.3, from http://www.ieee802.org/15/pub/TG3.html," April 2005.

[53] MultiBand OFDM Alliance SIG, "MultiBand OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a," September 2004.

[54] R. Kolic, "Wireless USB Brings Greater Convenience and Mobility to Devices," in *Technology@Intel Magazine*, February 2004.

[55] E. Royer and C.-K. Toh, "A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks," *IEEE Personal Communications*, April 1999.

[56] A. Boukerche, "Performance Evaluation of Routing Protocols for Ad Hoc Wireless Networks," *Mobile Networks and Applications: The Journal of Special Issues on Mobility of Systems, Users, Data and Computing*, vol. 9, pp. 333-342, August 2004.

[57] C. Perkins, "Highly Dynamic Destination Sequence Distance Vector Routing (DSDV) for Mobile Computers," presented at ACM SIGCOMM'94, 1994.

[58] R. Bellman, "On a Routing Problem," *Quarterly of Applied Mathematics*, vol. 16, pp. 87-90, 1958.

[59] L. R. Ford, "Technical Report P-932 - Network Flow Theory," The RAND Corperation, Santa Moncia, California, August 1956.

[60] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, USA: Prentice Hall Inc., 1992.

[61] B. Bellur and R. G. Ogier, "A Reliable, Efficient Topology Broadcast Protocol for Dynamic Networks," presented at IEEE InfoCom, New York, USA, March 1999.

[62] R. Ogier, F. Templin, and M. Lewis, "IETF RFC 3684 - Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)," February 2004.

[63] T. Clausen and P. Jacquet, "IETF RFC 3626 - Optimized Link State Routing Protocol (OLSR)," October 2003.

[64] T. Clausen, P. Jacquet, A. Laouiti, P. Muhlethaler, a. Qayyum, and L. Viennot, "Optimized Link State Routing Protocol," presented at 5th IEEE International Multitopic Conference (INMIC), Lahore, Pakistan, December 2001.

[65] C.-C. Chiang, H.-K. Wu, W. Liu, and M. Gerla, "Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel," presented at IEEE Singapore International Conference on Networks (SICON'97), Singapore, April 1997.

[66] M. Joa-Ng and I.-T. Lu, "A Peer-to-Peer zone-based two-level link state routing for mobile Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks*, pp. 1415-25, August 1999.

[67] A. Iwata, C. Chiang, G. Pei, M. Gerla, and T. Chen, "Scalable Routing Strategies for Ad-Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1369-1379, August 1999.

[68] M. R. Pearlman and Z. J. Haas, "Determining the Optimal Configuration for the Zone Routing Protocol," *IEEE Journal on Selected Areas in Communications*, vol. 17, August 1999.

[69] G. Pei, M. Gerla, and T.-W. Chen, "Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks," presented at IEEE ICC 2000, New Orleans, USA, June 2000.

[70] T.-W. Chen and M. Gerla, "Global State Routing: A New Routing Scheme for Ad-hoc Wireless Networks," presented at IEEE ICC'98, Atlanta, USA, June 1998.

[71] M. Gerla, X. Hong, and G. Pei, "Landmark Routing for Large Ad Hoc Wireless Networks," presented at IEEE GLOBECOM 2000, San Francisco, USA, November 2000.

[72] C. Perkins and E. Royer, "Ad Hoc On demand Distance Vector Routing," presented at 2nd IEEE Workshop on Mobile Computing Systems and Applications, February 1999.

[73] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva, "A Performace Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols," presented at ACM MobiCom 1998, Dallas, USA, October 1998.

[74] C. Perkins, E. Belding-Royer, and S. Das, "IETF RFC 3561 - Ad Hoc On-Demand Distance Vector (AODV) Routing," July 2003.

[75] D. B. Johnson, D. A. Maltz, and Y.-C. Hu, "IETF Draft: The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)," July 2004.

[76] V. D. Park and M. S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," presented at IEEE INFOCOM'97, Kobe, Japan, April 1997.

[77] P. Johansson, T. Larsson, N. Nedman, B. Mielczarek, and M. Degermark, "Scenario-based Performance Analysis of Routing Protocols for Mobile Ad-hoc Networks," presented at ACM MobiCom 1999, Seattle, USA, August 1999.

[78] S. R. Das, C. E. Perkins, and E. M. Royer, "Performance Comparison of Two On-Demand Routing Protocals for Ad Hoc Networks," presented at IEEE InfoCom, Tel-Aviv, Israel, March 2000.

[79] University of California in Los Angeles (UCLA), "GloMoSim, from http://pcl.cs.ucla.edu/projects/glomosim." Los Angeles, USA, April 2005.

[80] X. Zeng, R. Bagrodia, and M. Gerla, "GloMoSim: a library for parallel simulation of large-scale wireless networks," presented at ACM Workshop on Parallel and Distributed Simulation (PADS'98), Banff, Canada, May 1998.

[81] Scalable Network Technologies, "QualNet 3.8, from http://www.scalable-networks.com." Los Angeles, USA, April 2005.

[82] OPNET Technologies, "OpNet, from http://www.opnet.com." Bethesda, USA, April 2005.

[83] The VINT Project, "The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/," April 2005.

[84] M. Greis and VINT Project, "Tutorial for the Network Simulator ns-2, from http://www.isi.edu/nsnam/ns/tutorial/index.html," April 2005.

[85] J. Chung and M. Claypool, "NS by Example, from http://nile.wpi.edu/ns/," April 2005.

[86] S. Aust, M. Sessinghaus, C. Pampu, and C. Görg, "Hierarchical Mobile IP ns-2 Extensions for Mobile Ad hoc Networks," presented at 4th IASTED International Multi-Conference on Wireless Networks and Emerging Technologies (WNET 2004), Banff, Canada, July 2004.

[87] The CMU Monarch Project, "The CMU Monarch Project: Wireless and Mobility Extension to ns," Work in Progress, available from http://www.monarch.cs.cmu.edu, December 2002.

[88] K. Fall and K. Varadhan, "The ns Manual, http://www.isi.edu/nsnam/ns/doc/index.html," The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC, February 2005.

[89] D. Johnson and D. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks*. The Netherlands: Kluwer Academic Publishers, 1996.

[90] C. Bettstetter, G. Resta, and P. Santi, "The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 2, pp. 257-269, July 2003.

[91] J. Yoon, M. Liu, and B. Noble, "Random Waypoint Considered Harmful," presented at IEEE InfoCom 2003, San Francisco, USA, April 2003.

[92] C. Bettstetter, "Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects," *ACM Mobile Computing and Communications Review*, vol. 5, pp. 55-67, July 2001.

[93] A. B. McDonald and T. Znati, "A Mobility based Framework for Adaptive Clustering in Wireless Ad Hoc Networks," *IEEE Journal on Selected Areas in Communication*, vol. 17, August 1999.

[94] M. Dillinger, K. Madani, and N. Alonistioti, *Software Defined Radio: Architectures, Systems and Functions*. Chichester, England: John Wiley & Sons, Ltd., June 2003.

[95] M. Lott, R. Halfmann, E. Schulz, and M. Radimirsch, "Medium access and radio resource management for ad hoc networks based on UTRA TDD," presented at ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'01), Long Beach, USA, October 2001.

[96] M. Berg, "A Concept for Hybrid Random/Dynamic Radio Resource Management," presented at IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'98), Boston, USA, September 1998.

[97] Internet Engineering Task Force, "IETF Mobile Ad-hoc Networks (MANET) Working Group," April 2005.

[98] M. Joa-Ng and I.-T. Lu, "A Peer-to-Peer Zone-Based Two-Level Link State Routing for Mobile Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1415-1425, 1999.

[99] Y.-B. Ko and N. H. Vaidya, "Location-Aided Routing (LAR) in Mobile Ad Hoc Networks," presented at ACM/IEEE International Conference on Mobile Computing and Networking, Dallas, USA, October 1998.

[100] C. E. Perkins, "Mobile IP," *IEEE Communications Magazine*, vol. 35, pp. 84-99, May 1997.

[101] S. Aust, D. Proetel, N. A. Fikouras, C. Pampu, and C. Görg, "Policy based Mobile IP Handoff Decision (POLIMAND) using Generic Link Layer Information," presented at IEEE International Conference on Mobile and Wireless Communication Networks (MWCN 2003), Singapore, October 2003.

[102] K. Kuladinithi, N. A. Fikouras, A. Könsgen, A. Timm-Giel, and C. Görg, "Enhanced Terminal Mobility through the use of Filters for Mobile IP," presented at IST Summit on Mobile and Wireless Communications (IST Summit), Aveiro, Portugal, June 2003.

[103] **I. Gruber** and H. Li, "A novel Ad Hoc Routing Algorithm for Cellular Coverage Extension," presented at Networks 2004, Vienna, Austria, June 2004.

[104] Y. Lin and Y. Hsu, "Multihop Cellular: A New Architecture for Wireless Communications," presented at IEEE InfoCom 2000, Tel Aviv, Israel, March 2000.

[105] R. Stevens, *TCP/IP Illustrated: The Protocols*, vol. 1. Reading, USA: Addison-Wesley, 1994.

[106] C. Casetti, M. Gerla, S. Mascolo, M. Y. Sansadidi, and R. Wang, "TCP Westwood: End-to-End Congestion Control for Wired/Wireless Networks," *Wireless Networks*, vol. 8, pp. 467-479, 2002.

[107] J. Liu and S. Singh, "ATCP: TCP for Mobile Ad Hoc Networks," *IEEE Journal of Selected Areas in Communication*, vol. 19, pp. 1300--1315, July 2001.

[108] K. Chandran, S. Raghunathan, S. Venkatesan, and R. Prakash, "A Feedback based Scheme for Improving of TCP Performance in Ad hoc Wireless Networks," *IEEE Personal Communications Magazine*, vol. 8, pp. 34-49, February 2001.

[109] A. Klemm, C. Lindemann, and O. Waldhorst, "A Special-Purpose Peer-to-Peer File Sharing System for Mobile Ad Hoc Networks," presented at IEEE Vehicular Technology Conference (VTC2003-Fall), Orlando, USA, October 2003.

[110] J. Eberspächer, R. Schollmeier, S. Zöls, and G. Kunzmann, "Structured P2P Networks in Mobile and Fixed Environments," presented at International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs '04), Ilkley, England, July 2004.

[111] M. Conti, G. Maselli, G. Turi, and S. Giordano, "Cross-layering in mobile ad hoc network design," *IEEE Computer*, vol. 37, pp. 48-51, February 2004.

[112] A. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, pp. 8-27, August 2002.

[113] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirc, "A Survey on Sensor Networks," *IEEE Communications Magazine*, vol. 40, pp. 102-114, August 2002.

[114] J.-P. Hubaux, J.-Y. L. Boudec, S. Giordano, M. Hamdi, L. Blazevic, L. Buttyan, and M. Vojnovic, "Towards mobile ad-hoc WANs: Terminodes," presented at IEEE Wireless Communications and Networking Conference (WCNC 2000), Chicago, USA, September 2000.

[115] N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy, "PATHS: Analysis of PATH Duration Statistics and their Impact on Reactive MANET Routing Protocols," presented at ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'03), Annapolis, USA, June 2003.

[116] Z. Ye, S. V. Krishnamurthy, and S. K. Tripathi, "A Framework for Reliable Routing in Mobile Ad Hoc Networks," presented at IEEE InfoCom, San Francisco, USA, April 2003.

[117] S. R. Das, R. Castañeda, and J. Yan, "Simulation-based Performance Evaluation of Routing Protocols for Mobile Ad Hoc Networks," *ACM/Baltzer Mobile Networks and Applications (MONET)*, vol. 5, pp. 179-189, September 2000.

[118] B. McDonald and T. Znati, "A Path Availability Model for Wireless Ad-Hoc Networks," presented at IEEE Wireless Communications and Networking Conference (WCNC), New Orleans, USA, September 1999.

[119] M. Marina and S. Das, "On-demand Multipath Distance Vector Routing in Ad Hoc Networks," presented at IEEE International Conference on Network Protocols (ICNP), 2001.

[120] A. Nasipuri, R. Castañeda, and S. R. Das, "Performance of multipath routing for on-demand protocols in mobile ad hoc networks," *ACM/Kluwer Mobile Networks and Applications (MONET)*, vol. 6, pp. 339-349, August 2001.

[121] A. Valera, K. G. Seah, and S. V. Rao, "Champ: A highly resilient and energy-efficient routing protocol for mobile ad hoc networks," presented at The Fourth IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002), Stockholm, Sweden, September 2002.

[122] A. Valera, W. K. G. Seah, and S. Rao, "Cooperative Packet Caching and Shortest Multipath Routing in Mobile Ad hoc Networks," presented at IEEE INFOCOM 2003, San Francisco, USA, April 2003.

[123] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly-Resilient, Energy-Efficient Multipath Routing in Wireless Sensor Networks," presented at ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'01), Long Beach, USA, October 2001.

[124] S.-J. Lee and M. Gerla, "Split multipath routing with maximally disjoint paths in ad hoc networks," presented at IEEE International Conference on Communications (ICC), Helsinki, Finland, June 2001.

[125] J. C. Navas and T. Imielinski, "Geocast: Geographic Addressing and Routing," presented at ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom'97), Budapest, Hungary, September 1997.

[126] Y.-B. Ko and N. H. Vaidya, "GeoTORA: A Protocol for Geocasting in Mobile Ad Hoc Networks," presented at 2000 International Conference on Network Protocols, Osaka, Japan, November 2000.

[127] W.-H. Liao, Y.-C.-. Tseng, K.-L. Lo, and J.-P. Sheu, "Geogrid: A Geocasting Protocol for Mobile Ad Hoc Networks Based on Grid," *Journal of Internet Technology*, vol. 1, pp. 23-32, 2000.

[128] S. Jiang, Y. Liu, Y. Jiang, and Q. Yin, "Provisioning of Adaptability to Variable Topologies for Routing Schemes in MANETs," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1347-1356, September 2004.

[129] A. Tsirigos and Z. J. Haas, "Multipath routing in mobile ad hoc networks or how to route in the presence of frequent topology changes," *IEEE Communications Magazine*, vol. 39, pp. 132-138, November 2001.

[130] A. Tsirigos and Z. J. Haas, "Analysis of Multipath Routing Part I: The Effect on the Packet Delivery Ratio," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 138- 146, January 2004.

[131] A. Tsirigos and Z. J. Haas, "Analysis of Multipath Routing, Part 2: Mitigation of the Effects of Frequently Changing Network Topologies," *IEEE Transactions on Wireless Communications*, vol. 3, March 2004.

[132] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An Analysis of the Optimum Node Density for Ad hoc Mobile Networks," presented at IEEE International Conference on Communications 2001, June 2001.

[133] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," presented at 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), Seattle, USA, August 1999.

[134] T. Camp, J. Boleng, and V. Davies, "Mobility Models for Ad Hoc Network Simulations," *Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, pp. 483-502, 2002.

[135] C. Bettstetter, "Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks," presented at MSWiM'01, July 2001.

[136] E. W. Weisstein, "Probability Function," in *MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/ProbabilityFunction.html*, December 2004.

[137] E. W. Weisstein, "Disk Point Picking," in *MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/DiskPointPicking.html*, January 2005.

[138] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2 ed. Cambridge, USA: Cambridge University Press, 1997.

[139] W. Su, S. Lee, and M. Gerla, "Mobility Prediction in Wireless Networks," presented at IEEE Military Communications Conference (MILCOM), Los Angeles, USA, October 2000.

[140] R. Sedgewick, *Algorithms*, 2nd ed. Reading, USA: Addison-Wesley, 1988.

[141] Algorithmic Solutions Software GmbH, "The LEDA User Manual, Version 4.5," July 2004.

[142] E. W. Weisstein, "Circle Line Picking," in *MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/CircleLinePicking.html*, December 2004.

[143] C. Bettstetter, "On the Connectivity of Ad Hoc Networks," *The Computer Journal, Special Issue on Mobile and Pervasive Computing*, vol. 47, pp. 432-447, July 2004.

[144] E. W. Weisstein, "Square Line Picking," in *MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/SquareLinePicking.html*, January 2005.

[145] D. M. Lazoff and A. T. Sherman, "An Exact Formula for the Expected Wire Length Between Two Randomly Chosen Terminals," Computer Science Department, University of Maryland Baltimore County, Baltimore, USA July 1994.

[146] H. T. Friis, "A note on a simple transmission formula," *IRE Waves and Electrons*, vol. 34, pp. 254-256, May 1946.

[147] T. S. Rappaport, *Wireless Communications, Principles and Practice*. Englewood Cliffs, USA: Prentice Hall Inc., 1996.

[148] B.-C. Seet, G. Liu, B.-S. Lee, C.-H. Foh, K.-J. Wong, and K.-K. Lee, "A-STAR: A Mobile Ad-hoc Routing Strategy for Metropolis Vehicular Communications," presented at Networking 2004, Athens, Greece, May 2004.

[149] N. Schult, M. Mirhakkak, and D. LaRocca, "Routing in mobile ad hoc networks," presented at IEEE Military Communications Conference Proceedings (MilCom'99), Atlantic City, USA, October 1999.

[150] A. Kamat and R. Prakash, "Effects of link stability and directionality of motion on routing algorithms in MANETs," presented at IEEE International Conference on Computer Communications and Networks (ICCCN'2000), Las Vegas, USA, October 2000.

[151] K.-J. Wong, B.-S. Lee, B.-C. Seet, G. Liu, and L.J.Zhu, "BUSNet: Model and Usage of Regular Traffic Patterns in Mobile Ad Hoc Networks for Inter-vehicular Communications," presented at International Conference on Information and Communication Technologies (ICTE'03), Bangkok, Thailand, April 2003.

[152] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," presented at ACM International Conference on Mobile Computing and Networking (MobiCom'03), San Diego, USA, September 2003.

[153] J. B. Andersen, T. S.Rappaport, and S. Yoshida, "Propagation Measurement and Models for Wireless Communication Channels," *IEEE Communication Magazine*, pp. 42-49, January 1995.

[154] J. Beyer and R. Jakoby, "Two Semi-Empirical and Fast Prediction Models for Urban Microcells Compared with Measurements at 919 and 1873 MHz," presented at European Personal Mobile Communication Conference (EPMCC'97), Bonn, Germany, September 1997.

[155] T. Kürner and A. Meier, "Prediction of Outdoor and Outdoor-to-Indoor Coverage in Urban Areas at 1.8 GHz," *IEEE Journal on Selected Areas in Communications*, vol. 20, April 2002.

[156] I. Forkel and M. Salzmann, "Radio Propagation Modeling and its Application for 3G Mobile Network Simulation," presented at 10th Aachen Symposium on Signal Theory, Aachen, Germany, September 2001.

[157] J. Walfisch and H. L. Bertoni, "A Theoretical Model of UHF Propagation in Urban Environments," *IEEE Transaction on Antennas and Propagation*, vol. 36, pp. 1788-1796, December 1988.

[158] F. Ikegami, S. Yoshida, T. Takeuchi, and M. Umehira, "Propagation Factors controlling Mean Field Strength on Urban Streets," *IEEE Transactions on Antennas and Propagations*, vol. 32, pp. 822-829, August 1994.

[159] B. H. Fleury and P. E. Leuthold, "Radiowave Propagation in Mobile Communications: An Overview of European Research," *IEEE Communications Magazine*, vol. 34, pp. 70-81, February 1996.

[160] European Cooperation in the Field of Scientific and Technical Research (COST), *COST231 TD(90) 119 Rev. 2: Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz bands*. The Hague, September 1991.

[161] L. R. Maciel, H. L. Bertoni, and H. H. Xia, "Unified Approach to Prediction of Propagation over Buildings for all Ranges of Base Station Antenna Height," *IEEE Transaction of Vehicular Technology*, vol. 43, pp. 41-45, February 1993.

[162] J.-E. Berg, "A Recursive Method for Street Micro-Cell Path Loss Calculations," presented at IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'95), Toronto, Canada, September 1995.

[163] European Telecommunications Standards Institute, *ETSI Standard TR 101 112 v3.2.0: Universal Mobile Telecommunications System (UMTS);Selection procedures for the choice of radio transmission technologies of the UMTS*. Sophia-Antipolis Cedex, France, April 1998.

[164] J. E. Berg, *Building penetration, Digital Mobile Radio toward Future Generation Systems (COST 231 Final Report), sec. 4.6, pp. 167-174*. Brussels, Belgium, 1999.

[165] M. Lott and I. Forkel, "A Multi-Wall-and-Floor Model for Indoor Radio Propagation," presented at IEEE Vehicular Technology Conference (Spring VTC'2001), Rhode Island, Greece, May 2001.

[166] P. Bernardi, R. Cicchetti, and O. Testa, "An Accurate UTD Model for the Analysis of Complex Indoor Radio Environments in Microwave WLAN Systems," *IEEE Transactions on Antennas and Propagations*, vol. 52, pp. 1509-1520, June 2004.

[167] M. L. Rubio, A. García-Armada, R. P. Torres, and J. L. García, "Channel modeling and characterization at 17 GHz for indoor broadband WLAN," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 593-601, April 2002.

[168] M. Lott, "On the Performance of an Advanced 3D Ray Tracing Method," presented at European Wireless (EW'99), Munich, Germany, October 1999.

[169] O. Knauf, "Mobilitäts- und Kanalmodell für Ad-hoc-Netz Simulationen in Innenstädten," in *diploma thesis*. Munich, Germany: Technische Universität München, Lehrstuhl für Kommunikationsnetze, May 2003.

[170] M. Hutter, "Signalausbreitungs- und Bewegungsmodell für Ad Hoc Netz Simulationen in Urbanen Umgebungen," in *diploma thesis*. Munich, Germany: Technische Universität München, Lehrstuhl für Kommunikationsnetze, March 2004.

[171] 3rd Generation Partnership Project (3GPP), *Technical Specification Group Radio Access Networks; Radio Frequency (RF) system scenarios (Release 5): Report 3GPP TR 25.942 V5.3.0*, June 2004.

[172] Erik Nordström, B. Wiberg, and H. Lundgren, "AODV Routing Protocol Implementation." Uppsala, Finland: created at Uppsala University, http://user.it.uu.se/~henrikl/aodv, December 2004.

[173] B. Crow, I. Widjaja, J. Kim, and P. Sakai, "IEEE 802.11: Wireless Local Area Networks," *IEEE Communications Magazine*, pp. 116-126, September 1997.

[174] R. D. J. v. Nee, G. A. Awater, M. Morikura, H. Takanashi, M. A. Webster, and K. W. Halford, "New High-rate Wireless LAN Standards," *IEEE Communications Magazine*, vol. 37, pp. 82 - 88, December 1999.

[175] Wi-Fi Alliance, "802.11n: Q&A," in *http://www.wi-fi.org/OpenSection/pdf/802.11n_Q_A.pdf*, February 2005.

[176] J. Winters, "802.11n throttles up WLAN throughput," in *Network World, http://www.nwfusion.com/*, November 2004.

[177] European Telecommunications Standards Institute, *ETSI Standard TR 101 683: HIPERLAN Type 2: System Overview*. Sophia-Antipolis Cedex, France, February 2000.

[178] M. Radmirisch and V. Vollmer, "HIPERLAN Type 2 Standardization - an Overview," presented at European Wireless, Munich, Germany, October 1999.

[179] J. Khun-Jush, G. Malmgren, P. Schramm, and J. Torsner, "Overview and Performance of HIPERLAN Type 2 - A Standard for Broadband Wireless Communications," presented at IEEE Vehicular Technology Conference (VTC), Tokyo, Japan, May 2000.

[180] N. Esseling, E. Weiss, A. Krämling, and W. Zirwas, "A Multi Hop Concept for HiperLAN/2: Capacity and Interference," presented at European Wireless 2002, Florence, Italy, February 2002.

[181] N. Esseling, "Extending the Range of HiperLAN/2 Cells in Infrastructure Mode using Forwarding Mobile Terminals," presented at European Personal Mobile Communication Conference (EPMCC), Vienna, Austria, February 2001.

[182] T.-S. Ho and K.-C. Chen, "Performance Analysis of IEEE 802.11 CSMA/CA Medium Access Control Protocol," presented at IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'96), Taipei, Taiwan, October 1996.

[183] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal of Selected Areas in Communications*, vol. 18, pp. 535-547, March 2000.

[184] C. H. Foh and M. Zukerman, "Performance Analysis of the IEEE 802.11 MAC Protocol," presented at European Wireless, Florence, Italy, February 2002.

[185] L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-Access Modes and their Throughput-Delay Characteristics," *IEEE Transactions on Communications*, vol. 23, pp. 1400-1416, December 1975.

[186] IEEE Standards Department, *ANSI/IEEE Std. 802.3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. New York, USA, March 2002.

[187] C. Ware, T. Wysocki, and J. Chicharo, "On the Hidden Terminal Jamming Problem in IEEE 802.11 Mobile Ad Hoc Networks," presented at IEEE International Conference on Communications (ICC), Helsinki, Finland, June 2001.

[188] K. Xu, M. Gerla, and S. Bae, "How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks?" presented at IEEE Global Telecommunications Conference (GlobeCom'02), Taipei, Taiwan, November 2002.

[189] P. Karn, "MACA - A New Channel Access Method for Packet Radio," presented at ARRL/CRRL Amateur Radio Computer Networking Conference, New York, USA, April 1990.

[190] F. A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II - the Hidden Terminal Problem in Carrier Sense Multiple-Access Modes and the Busy-Tone Solution," *IEEE Transactions on Communications*, vol. 23, pp. 1417-1433, 1975.

[191] J. R. Walker, "IEEE P802.11 Wireless LANS, Unsafe at any Key Size: An Analysis of the WEP Encapsulation," Technical Report, 03628E, IEEE 802.11 committee, March 2000.

[192] Skype Technologies S.A., "Skype, from http://www.skype.com/," February 2005.

[193] P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek, and M. Cheng, "Using IEEE 802.11e MAC for QoS over Wireless," presented at IEEE International Performance, Computing, and Communications Conference (IPCCC), Phoenix, USA, April 2003.

[194] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e Wireless LAN for Quality of Service," presented at European Wireless, Florence, Italy, February 2002.

[195] A. Grilo and M. Nunes, "Performance Evaluation of IEEE 802.11e," presented at IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'02), Lisbon, Portugal, September 2002.

[196] S. Mangold, S. Choi, G. R. Hiertz, O. Klein, and B. Walke, "Analysis of IEEE 802.11E for QoS support in wireless LANs," *IEEE Wireless Communications*, vol. 10, pp. 40-50, December 2003.

[197] C. Brandauer, G. Iannaccone, T. Ziegler, S. Fdida, and M. May, "Comparison of tail drop and active queue management performance for bulk-data and web-like internet traffic," presented at Sixth IEEE Symposium on Computers and Communications (ISCC'01), July 2001.

[198] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397-413, August 1993.

[199] J. Jun, P. Peddabachagari, and M. Sichitiu, "Theoretical Maximal Throughput of IEEE 802.11 and its Applications," presented at IEEE International Symposium on Network Computing and Applications, Cambridge, USA, April 2003.

[200] H. Luo, S. Lu, and V. Bharghavan, "A New Model for Packet Scheduling in Multihop Wireless Networks," presented at ACM International Conference on Mobile Computing and Networking (MobiCom), Boston, USA, August 2000.

[201] H.-Y. Hsieh and R. Sivakumar, "Improving Fairness and Throughput in Multi-hop Wireless Networks," presented at International Conference on Networking (ICN), Colmar, France, July 2001.

[202] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed Fair Scheduling in a Wireless LAN," presented at ACM International Conference on Mobile Computing and Networking (MobiCom), Boston, USA, August 2000.

[203] K.-C. Wang and P. Ramanathan, "End-to-End Throughput and Delay Assurance in Multihop Wireless Hotspots," presented at ACM international Workshop on Wireless mobile applications and services on WLAN hotspots, San Diego, USA, September 2003.

[204] J. Jun and M. L. Sichitiu, "Fairness and QoS in Multihop Wireless Networks," presented at IEEE Vehicular Technology Conference (VTC), Orlando, USA, October 2003.

[205] J. Jun and M. L. Sichitiu, "The Nominal Capacity of Wireless Mesh Networks," *IEEE Wireless Communications*, vol. 10, pp. 8-14, October 2003.

[206] F. H. P. Fitzek, D. Angelini, G. Mazzini, and M. Zorzi, "Design and Performance of an enhanced IEEE 802.11 MAC Protocol for multihop coverage extension," *IEEE Wireless Communications, Special issue on the evolution of wireless LANs and PANs*, vol. 10, pp. 30-39, December 2003.

[207] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Distributed Multi-Hop Scheduling and Medium Access with Delay and Throughput Constraints," presented at ACM International Conference on Mobile Computing and Networking (MobiCom), Rome, Italy, July 2001.

[208] A. Bäßler, "Fairer Medienzugriff in mobilen Ad hoc Netzen mit Internet Gateway," in *diploma thesis*. Munich, Germany: Technische Universität München, Lehrstuhl für Kommunikationsnetze, May 2004.

[209] J. Walrand, *An Introduction to Queueing Networks*. Englewood Cliffs, USA: Prentice Hall Inc., 1988.

[210] E. Gelenbe and G. Pujolle, *Introduction to Queueing Networks, Second Edition*. Chichester, England: John Wiley & Sons, Ltd., April 1998.

[211] D. C. Lee, "Effects of leaky bucket parameters on the average queueing delay: Worst case analysis.," presented at IEEE Infocom'94, Toronto, Canada, June 1994.

[212] M. Junius, A. Speetzen, M. Steppler, M. Büter, D. Pesch, and others, "ComNets Class Library and Tools (CNCL) Documentation version 2.5," Aachen University of Technology, Ed. Aachen, Germany, 2002.