Dependable Routing for Cellular and Ad hoc Networks

Vom Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertation

von

Dipl.-Ing. Matthias Hollick

geboren am 2. März 1973 in Frankfurt

Erstreferent: Prof. Dr.-Ing. Ralf Steinmetz
Korreferent: Prof. Klara Nahrstedt (Ph.D.)

Tag der Einreichung: 1. November 2004
Tag der mündlichen Prüfung: 21. Dezember 2004

D17
Darmstädter Dissertation

Kurzfassung (Deutsch)

Diese Dissertation behandelt den Themenkomplex der verlässlichen Kommunikation in Mobilfunknetzen sowie in Ad hoc Netzen. Verlässlichkeit charakterisiert hierbei das konsistente Verhalten hinsichtlich Systemleistung und Systemstabilität. Der Schwerpunkt der Arbeiten liegt auf dem Routing *(Leitwegbestimmung/Paketvermittlung)* für die genannten Netzklassen.

Die Erforschung von Technologien zur Unterstützung der heutigen, mobilen Gesellschaft hat insbesondere auf dem Kommunikationssektor für Impulse gesorgt. Der erzielte Fortschritt manifestiert sich im Bereich der mobilen und drahtlosen Kommunikation mit dem Paradebeispiel des Mobilfunks, der eine dominierende Rolle einnimmt. Im Gegensatz hierzu kann das Internet die im Festnetzbereich errungene Vormachtstellung nicht auf den mobilen Bereich übertragen. Wir sind der Auffassung, dass die Dichotomie von Internet und Mobilfunk einer kritischer Überprüfung bedarf. In der vorliegenden Arbeit liegt der diesbezügliche Fokus auf Kommunikationsnetzen der nächsten Generation und im Speziellen auf deren Routingsystem. Als Forschungsgegenstand für die weitere Untersuchung wurde die Klasse der zellularen Netze (geschlossene Nutzergruppe und basierend auf Infrastrukturkomponenten) sowie die Klasse der Ad hoc Netze (offene Nutzergruppe und infrastrukturlos) gewählt.

Im ersten Teil der Arbeit wird das Konzept der Verlässlichkeit, bezogen auf das Routing in Kommunikationsnetzen (*Routing Dependability*), definiert. Die herausragenden, sich auf die Verlässlichkeit auswirkenden Charakteristika der betrachteten Netzklassen werden identifiziert. Im Weiteren wird ein Mobilitäts- und Lastmodell formuliert und instanziiert, welches die makroskopischen Effekte von Mobilität in Ballungsgebieten beschreibt. Unter der Arbeitshypothese einer Reduzierung der Zellgrößen sowie einer Integration von drahtlosen Netzzugangspunkten hoher Bandbreite (*Hotspots*) zeigen die Modellvorhersagen, dass der Einfluss von Mobilität auf das Routingsystem nicht vernachlässigt werden kann.

Im zweiten Teil der Arbeit werden Architekturen und Algorithmen vorgestellt, um die durch die Mobilität hervorgerufenen Lastschwankungen zu kompensieren. Unser

Ansatz ist, das Routingsystem über das Kernnetz hinausgehend auf das Funkzugangsnetz zu erweitern. Hierzu werden die existierenden hierarchischen und baumförmigen Netzstrukturen aufgebrochen und durch eine stärker vermaschte Netztopologie ersetzt. Diese neuartige Topologie erhöht die Flexibilität des Netzes und ermöglicht den Einsatz verteilter Routing- und Kontrollmechanismen. Mit Hilfe einer vergleichenden Leistungsbewertung moderner Routingverfahren wird die Anwendbarkeit ausgewählter Strategien zur Ressourcenkontrolle in den untersuchten Netzen gezeigt. Weiterhin erfolgt der Entwurf einer neuartigen Systemarchitektur, die Konzepte aus dem Gebiet variabler Topologien auf zellulare Netze überträgt. Speziell auf die Flexibilität dieser Architektur abgestimmt wird ein Routingverfahren entwickelt und hinsichtlich seiner Funktionsweise überprüft. Die damit in einer Simulationsstudie erzielten Ergebnisse zeigen im Vergleich zu anderen aktuellen Verfahren einen signifikanten Leistungsgewinn.

Die Herausforderungen auf dem Gebiet der Mobilität bleiben für die Klasse der Ad hoc Netze bestehen. Gleichzeitig setzt diese Netzklasse eine Kooperation von Netzknoten zwingend voraus. In Teil drei der Dissertation wird diese Kooperation in Frage gestellt. In einem ersten Schritt werden die durch Knotenfehlverhalten induzierten Effekte analytisch betrachtet. Ein theoretisches Modell des Routenfindungsprozesses wird unter Berücksichtigung verschiedener Klassen von Knotenfehlverhalten entwickelt und validiert. In einem zweiten Schritt schließt sich eine umfassende Simulationsstudie an. Untersuchungsgegenstand ist die Leistungsbewertung der Faktoren Mobilität, optimierte Routingverfahren sowie Knotenfehlverhalten bezüglich der Verlässlichkeit des Routingsystems. Die erzielten Ergebnisse zeigen mit zunehmendem Grad von Fehlverhalten, respektive zunehmender Mobilität, einen signifikanten Leistungsabfall bis hin zum Versagen des Netzes.

Abstract

This dissertation addresses an expedient set of research challenges in the area of cellular and ad hoc networks. In particular, we discuss the dependability—that is, the consistency of performance and behavior—of the routing system within such networks.

Our restless, mobile society has inspired diverse fields of research in the area of computer communication networks. The results manifest in recent advances in the area of wireless and mobile communications, with cellular networks being an exemplar of success. Despite the fact of its supremacy in fixed networks, the Internet has not yet caught up in the mobile domain. We argue that next generation networks should overcome the dichotomy between the aforementioned networks and decide to focus our attention to the role of the routing system. We select cellular networks (closed to subscribers and based on infrastructure) and ad hoc networks (open/public and infrastructureless) to represent our objects of research.

In Part I of the dissertation, we define the concept of routing dependability and identify the most important characteristics that influence this concept for the aforementioned networking paradigms. A mobility/workload model that captures the macroscopic effects of mobility for metropolitan areas is formulated and instantiated. Our results show that the effects of mobility acting upon the routing system cannot be neglected if we move towards smaller cell sizes and integration of hot spots in cellular systems.

To cope with the observed mobility-induced traffic dynamics, we develop and evaluate architectures and algorithms in Part II of our work. In particular, we argue to broaden the scope of the routing domain towards the radio access tier of cellular networks ("smart edge"). In a first step, we break up today's mainly hierarchically and tree-structured cellular networks and introduce a meshed network topology, which adds flexibility to deploy distributed network control mechanisms. A comparative performance analysis of state-of-the-art routing protocols shows the feasibility of such control mechanisms for basic resource management in the surveyed networks. In a second step, a novel system architecture is designed that augments concepts from variable topology networks to cellular architectures. We also develop a routing algorithm to take advantage of the achieved flexibility. As a proof of concept, the algorithm is implemented and a simulation study is performed to obtain deeper insights in the algorithm's operation. Our results show significant performance gains compared with current state-of-the-art algorithms.

Ad hoc networks are effected by mobility as well. Moreover, they are built upon the premise of node cooperation. In Part III of this dissertation, we challenge such cooperation. Therefore, we analytically investigate the effects impaired by node misbehavior and contribute a model of the route acquisition process that includes different classes of node misbehavior. The model is formulated analytically and validated by means of an experimental analysis. Subsequently, an extensive simulation study is conducted to investigate the dependability trade-off between node mobility, routing protocol performance optimizations, and node misbehavior. Our results show clearly the performance degradation of the routing system, leading to network frailty for an increasing number of misbehaving nodes or increasing mobility, respectively.

Acknowledgements

The compilation of a list of all people who deserve acknowledgement is an endeavor on its own. Most probably this list cannot be exhaustive—and even more likely arranging the list in proper order is impossible. This thesis is concerned with routing dependability (and performance), hence, we apply this concept to ensure the correct routing of our acknowledgements to all recipients:

Acknowledgement to whom acknowledgement is due.

Table of Contents

Part I
Introduction, Background, and Motivation

Part II
Dependable Routing for Cellular Networks

Part III
Dependable Routing for Ad hoc Networks

Part IV
Summary, Conclusion, Outlook, and Appendices

Preface

Over centuries, urban areas and cities have emerged and flourished by creating environmental conditions that encouraged the inhabitants to share and consume a variety of services, while attracting strangers to join their community. Transportation networks such as roads, railways, or other means of public transportation played an essential role in this process, because they facilitated trading of goods, exchanging human resources, and transferring knowledge. Considering communication networks and especially the Internet as a "virtual" counterpart of the real world transportation system, we observe similar development patterns. The Internet attracts an increasing number of people to participate in urban life in the virtual cities it empowers. However, the roots of the virtual communities are not as deep as the ones of their physical counterparts. The network layer (and hence the routing system) is the core of the Internet's transportation system. We consider the service delivered by this routing system as the key utility to influence whether or not the Internet is able to mirror the magnetism of cities on a scale of global proportions. However, we find that today's routing architectures are insufficiently prepared to handle the influence of real world phenomena such as mobility. At the same time, we witness an increasing demand for a seamless and pervasive nature of communications. This thesis highlights some important aspects of the basic utilities of wireless communication networks. In particular, we investigate challenges related to the dependability of the routing system in cellular and ad hoc networks.

"Human beings, who are almost unique in having the ability to learn from the experience of others, are also remarkable for their apparent disinclination to do so."
—Douglas Adams

Part I
Introduction, Background, and Motivation

1 Introduction

This chapter introduces and motivates our work in the area of routing dependability. Starting from a non-technical perspective we observe that our modern society increasingly depends on communication and information technology. In combination with the desire for mobile and pervasive communications this leads to a set of unsolved problems in the area of dependability in cellular and ad hoc networks. Since we are especially interested in network layer issues, we briefly describe the routing foundations of the Internet and discuss the shortcomings of its architecture with respect to mobile and pervasive communications. This discussion leads to the identification of our working area. We then detail the goals and contributions of our work for the area of cellular and ad hoc networks. At the end of the introduction we briefly survey work that is closely related to ours, and outline the structure of the dissertation. The discussion of our working area is further refined in Chapter 2 and Chapter 3. Moreover, the individual goals are addressed in Chapter 4 and Chapter 5 of this dissertation for the cellular network domain and in Chapter 6 and Chapter 7 for the ad hoc network domain.

1.1 Motivation

While the preface envisioned virtual cities on top of the Internet, such a development is only in its early stages. At the same time, however, computer communication networks are getting increasingly important for enabling and maintaining complex information systems, which empower our society. Governments as well as all kinds of organizations and individuals increasingly rely on these systems in every day life. Over the last decade the Internet evolved to be the platform of choice for communication at large—despite the fact that the Internet's best-effort routing system has never been designed for this usage. However, if the trustworthy and dependable (that is, consistent in performance and behavior) operation of the routing system cannot be guaranteed, then this may lead to failures in systems that depend on communication networks. This, in fact, may cause catastrophic disruptions of normal life or the halting of the virtual cities in our example.

The desire for pervasive communications and the wish to communicate while being mobile makes matters even more challenging, because it goes hand in hand with increasing dynamics at different levels of the communication system. We observe that applications such as electronic banking, electronic commerce, or electronic government currently evolve towards *mobility*, that is, *mobile* e-banking, etc. [LC03]. Hence, it is necessary to investigate wireless networks, such as cellular networks or ad hoc networks, with respect to their ability to fulfill the demanded quality of service (QoS) to meet the strict requirements of such applications and services. From an application or end user perspective the internals of the network are not visible. Nonetheless, these internals—and here especially the routing system—allow or hinder the predictability and determinism of the delivered service and define the achievable performance region of the communication system. The goal of our research is to investigate various aspects coupled with the dependability of routing networks. We specially focus on the packet-switched domains of cellular networks and ad hoc networks.

To detail the motivation of our work, in the following subsection, we introduce the design principles of the Internet. We focus on the Internet because it can be considered the state-of-the-art design for packet-switched (routing) networks. In particular, we introduce the foundations of the Internet's routing system and point out its shortcomings. Subsequently, we discuss the challenge of mobility and its impact on Internet routing. Both discussions are further refined in Chapter 2 respectively Chapter 3 of this dissertation. We then detail the goals and contributions of our work for the area of cellular and ad hoc networks. This chapter is concluded with a survey of related work similar in scope to ours and an outline of the structure of the dissertation.

1.1.1 The Routing Foundations of the Internet

The history of and the reasoning for the design of the Internet cannot be explained in a few words. Here, we give a brief introduction to the routing foundations of the Internet required as background for our work.

The basic concept of the Internet, the descendant of the military ARPANET established around 1973, is based on the transparent transmission of datagrams across an arbitrary network of networks. Before the Internet networks had mostly been connected by application layer gateways. This resulted in a loss of functionality as well as difficult application deployment. The need to overcome these gateways motivated the creation of an *internet layer*. This layer allows to form a bigger network and was intended for global addressing. Moreover, the network is virtualized and end-to-end protocols are isolated from network details and changes. To maximize interoperability while minimizing the number of service interfaces, a *single* internet protocol has been defined, the Internet Protocol (IP), which can be considered the substrate of the Internet's routing system. IP follows the packet-switched routing paradigm of Kleinrock [Kle78]. To assure that the number of usable networks is maximized, IP version 4 (IPv4) has been specified as a *narrow* protocol with the lowest common denominator of network functionality. At this point in time logical addresses were unique, datagrams were not changed in transit, and end systems would handle error detection, retransmission, security, naming, and binding. These concepts have determined the basic design of most Internet applications.[1]

Due to the simplicity of its basic principles and the integration into the UNIX platform, the Internet became a useful tool for collaboration among researchers. Just one decade ago awareness of the potential of the Internet was stimulated at large with the opening to public and commercial applications and the introduction of novel services including the World Wide Web. Today, the Internet is regarded as the most successful communication network ever. However, the incredible success of Internet technology also led to a set of emerging problems, most of which have been addressed by "workarounds" or protocol extensions. To name a few, network layer security has been addressed with the IP Security framework [KA98], network layer mobility has been addressed with Mobile IP [Per02], and network address translation [SE01] allows for private networks and to deal with limitations of the Internet's addressing scheme. However, IP Security as well as Mobile IP have never gained wide acceptance. While the add-ons to IP have mostly been specified to leave the core of the network untouched, they also induced problems and the need to address interdependencies between the various extensions. As a matter of fact, the actual Internet comprises firewall systems, virtual private networks, proxies, caches, dynamic and unstable addresses, etc., which severely affect the Internet and result in a loss of transparency [Car00]. As a result applications either fail completely, need modification, or must be specially handled by novel components[2]. Despite these facts there is a strong wish to keep the key principles of the Internet. Because of the imperfections of the Internet, some hard problems, such as the provision of QoS, have never been solved successfully, though[3]. Also, to bridge the gap between

1) A more detailed discussion on the design principles and the routing architecture of the Internet is provided by Cerf and Kahn [CK74], Clark [Cla88], and Carpenter [Car00].

2) See, for example, Roedig [Roe02] in this context for the study and development of novel firewall architectures to allow for seamless multimedia communications.

the heterogeneous systems residing on top of the internet layer, gateways have again gained momentum[4]. The introduction of a next generation internetworking protocol called IP version 6 (IPv6) has yet to succeed. However, it is doubtful if IPv6 allows to reestablish the transparency of the network. From our routing centric perspective IPv6 does not solve the majority of the existing problems. It is neither more secure than IPv4, nor does it provide solutions for the QoS problem or routing related issues. It does, however, regard the existence of these problems and introduces features to address the problems more easily.

As a consequence, there is a strong motivation to layer new applications over old ones and to design application layer solutions for problems that should naturally be approached in the network layer. As a central question remains, which functionality should be addressed on the internet layer, and, which requirements are of highest importance for the routing system. We believe that the basic characteristics allowing for high level concepts (such as QoS or security) should be explicitly addressed in the routing system. Moreover, the performance characteristics of new classes of applications (which are thought to be the nucleus for the virtual communities we imagined in the preface) need to be supported adequately. This includes demanding applications such as videoconferencing or teleimmersive applications, peer-to-peer systems or grid computing, multimedia messaging or augmented reality. From a networking perspective, challenges such as delay or disruption tolerant networks, an interplanetary Internet, or large scale content distribution networks have elaborate requirements as well.

The success of the Internet has largely been due to the simplicity of its routing architecture, which has fueled the development of a variety of applications on top of it. We believe that this success will only continue if the pure functionality of the routing service evolves from a simple best-effort scheme to deliver the datagrams towards more powerful and dependable routing paradigms to support demanding applications and services.

1.1.2 The Mobility Challenge—The Rise or Fall of Internet Technology

The last century was strongly influenced by the trend of personal *mobility* of the masses. The invention and success of the automobile, the bridging of large distances by means of airplanes, and the further development of public transportation contributed to this trend. A parallel trend can be observed in communication networking research. The invention of wireless transmission enabled networks to support mobile communications. A multitude of wireless networks (such as wireless extensions of telecommunication networks and wireless data networks) has been developed.

3) See, for example, Schmitt [Sch01] and Karsten [Kar00] for open issues and solutions coupled with resource reservation and the provisioning of QoS in heterogeneous networks. Heckmann [Hec04] discusses QoS from the perspective of a service provider.

4) See, for example, Ackermann [Ack02] for a discussion of gateway issues in the application area of voice over IP.

We motivate our work on basis of two scenarios that are of particular interest for mobile and wireless communications. First, Kleinrock [Kle96] coined the term *nomadic computin*g, which he later refined [Kle01] to cover a life and working situation. His definition includes the necessary prerequisites for the underlying network:

> *„But, in fact, most of us are nomads, moving between office, home, airplane, hotel, automobile, branch office, conference room, bedroom, etc. In so doing, we often find ourselves decoupled from our «home base» computing and communications environment. […] Nomadicity may be defined as the system support needed to provide a rich set of computing and communication capabilities and services to nomads as they move from place to place in a transparent, integrated and convenient form.“*

Solutions for Kleinrock's ideas are currently under development. However, the impact of mobility incurs various problems such as adequate mobility support or the discovery of viable services for the nomadic user [Hol01] [HS01].

Second, Weiser's vision of *seamless* and *ubiquitous (pervasive)* communications was the starting point for one of the most challenging and demanding quests in networking research [Wei91] [Wei93]. While pervasive computing is not to be confused with mobile communications, it necessitates a communication infrastructure to allow for seamless mobile communications that satisfies the higher level paradigms envisioned. Weiser's ideas summarize some of the most intriguing aspects to mediate the effects of physical environments and introduce visions of virtual mobility not realized yet. While focusing on application concepts, the projection of Weiser's ideas has impacts on the network layer, though, and mandates a mobility-aware routing system.

If we investigate the mechanisms of today's Internet supporting nomadic or pervasive communications, we find various problems linked to the mobility support in the routing system. The severity of these problems increases with the expected growth in mobile communications usage. The number of mobile phones has already crossed one billion [GSM04] and devices such as mobile music players or gaming devices are currently widely deployed, thus, pioneering a wave of always-on and always-connected user equipment. On top of individual and device mobility, we see emerging scenarios such as networks attached to people (personal area networks) or mobile ad hoc networks, which move with respect to their point of attachment. In summary, we expect mobility to introduce dynamics on various timescales and dimensions effecting the routing system.

The principles of the Internet mandate to address mobility on the network layer, because this layer provides for identity resolution, path computation, and forwarding of datagrams. However, the proposed network layer solution to mobility, Mobile IP [Per02], has not succeeded in fixing the problems related to mobility, one reason being the inadequate support for security [Sol98] [Hol00]. Because of this, researchers as well as application developers have been studying and implementing partial and proprietary solutions for the Internet's mobility problem. As an underlay solution, one can run Internet technology on top of mobility supporting networks such as cellular telecommunication networks [3GPP04]. As long as these networks do not support native IP, such solutions inherit, however, some unwanted limitations of the underlying technology. There also exist overlay solutions addressing the mobility problem. For example, Stoica et al. propose an application level approach capable of dealing with mobile users [SAZ+04]. This comes at the expense of re-introducing functionalities at the application layer that typically reside in the network layer: application level data routing, addressing, naming, etc. Moreover, overlays are often not aware of the underlying routing topology and may introduce even higher dynamics into the system [QYZS03].

1.2 Goals

The motivation of our work has highlighted the importance of dependable operation of communication networks. Moreover, we have described the basics of the Internet's routing system and introduced various challenges in the area of mobility-aware routing. We found a chasm, which separates the above envisioned scenarios of seamless nomadic and ubiquitous computing from being realized in the near future: today's prevalent but ill-suited Internet technology, which has never been designed to support the dynamics coupled with seamless mobile and pervasive communications. We believe that the challenge of mobility support marks an important decision point for the Internet. Depending on whether or not the existing architecture and technologies are able to adapt to the requirements, one may foresee this challenge to lead to the fall of Internet technology if unsolved—bearing also the possibility to further accelerate the Internet's none-to-second rise if solved.

The main goal of our work is to cross this chasm and to provide building blocks allowing for dependable, mobility-aware routing. The scope of our work is determined as follows. We build on the Internet paradigm of packet-switched routing. We further assume that user and end system mobility is supported by means of wireless networks. For the infrastructure-based domain we consider cellular radio access networks and for the infrastructureless domain we consider ad hoc networks. These two classes of networks have dissimilar characteristics and applications today. However, we believe that both network classes can benefit from each other and we thus cover aspects from both classes in our work. Our particular focus within these networks is the routing domain. Here, various challenges coupled to the dependable operation of the routing system persist. In this dissertation we address the following goals:

- The *quality of service* delivered by packet-switched networks that support mobility heavily depends on the performance of the routing system. There exists a close relationship between the network performance and the *dependability* of the underlying routing system. However, the effects of mobility on the dependability of routing systems are not well understood. Goal of our work is to establish a common understanding of the concept of routing dependability and to derive the most important characteristics influencing this concept for cellular and ad hoc networks.

- In *cellular networks* we witness a trend towards data communication. In consequence, the transition from connection-oriented towards packet-switched networks is widely discussed. Architectures to support packet-switched communication without compromising the requirements of the existing applications are necessary to allow for this transition. The performance of the overall system is at stake if the underlying routing system does not fit properly with the requirements of these architectures. Goal of our work is to develop a next generation cellular network architecture in combination with an efficient routing strategy.
- *Ad hoc networks* allow for the spontaneous formation of communication networks without a dedicated infrastructure. These networks are built upon the premise of cooperative nodes. The influence of node misbehavior on the dependable operation of the routing system and the trade-off between network performance and network frailty cannot easily be predicted today. This hinders the large scale deployment of ad hoc networks. Goal of our work is to develop models to predict the ad hoc network performance of realistic protocols including node misbehavior.

Figure 1 summarizes the roadmap of our work. The increasing dynamics introduced by nomadic and pervasive computing paradigms take their toll on breaking up established mechanisms of the quasi-static Internet. We borrow concepts from the Internet, cellular networks, and ad hoc networks to leverage solutions mitigating the effects of mobility on the routing system. Here, we focus on the basic concept of dependable routing, that is, the survivable and robust operation of the network to provide a consistent performance and behavior of the delivered routing service.

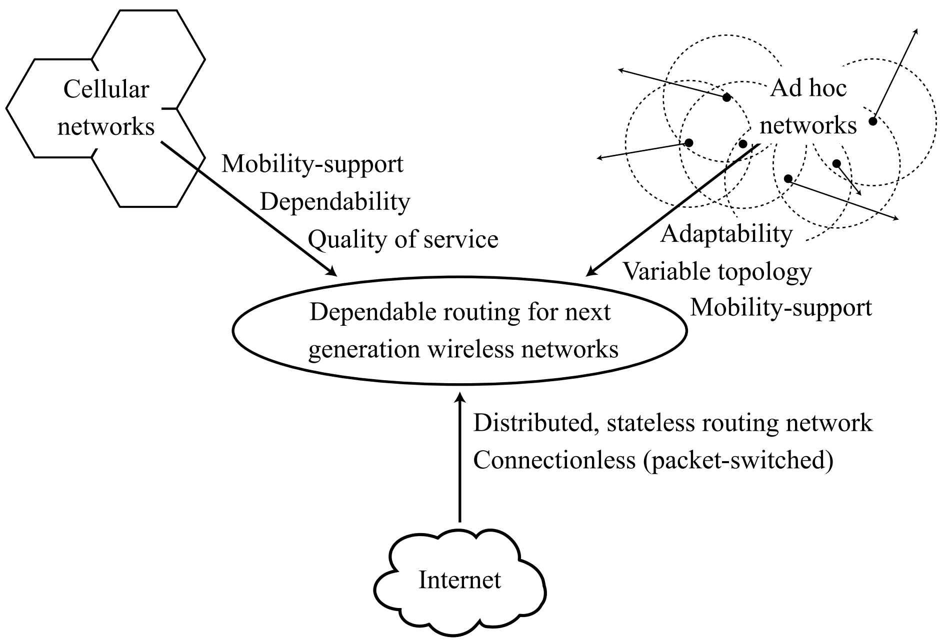


Figure 1: Roadmap of the thesis.

1.3 Contributions

As described in the previous sections, this dissertation addresses the challenge of dependable routing in cellular and ad hoc networks. Our contributions are as follows:

- *Conceptualization of routing dependability.*
  Based on a thorough literature survey of dependability in communication systems, we precisely define the subject of our dissertation, namely routing dependability for ad hoc and cellular networks. For both classes of networks we derive the most important characteristics influencing dependability. Based on these characteristics we deduce the dimensions of dependability for further study.
- *Macroscopic mobility/workload model for metropolitan areas.*
  To study the effects of mobility on the network performance in cellular networks, we formulate a mobility and workload model to capture the macroscopic effects of mobility in metropolitan areas. Our model describes the effects of mobility more precisely compared with currently available models. It is a hybrid of an empirical mobility model and a synthetic traffic model and allows for flexible parameterization. We instantiate the model for the city of Darmstadt and discuss the resulting workload dynamics induced by mobility.
- *Network architecture for radio access networks with static infrastructure.*
  We develop a static cellular network architecture to improve the performance of radio access networks for metropolitan scenarios. Our architecture extends the scope of the routing domain towards the edge of the network. Moreover, we depart from the strict tree hierarchy of today's cellular networks and introduce interconnections to form a meshed access network. We instantiate our architecture for the case of the Darmstadt area and study the performance of state-of-the-art routing algorithms in our architecture by means of a comparative analysis. The workload for the study is generated using our macroscopic mobility/workload model.
- *Network architecture for radio access networks with variable infrastructure.*
  We augment concepts from variable topology networks to our static cellular network architecture to form the novel variability-augmented cellular architecture. The architecture allows for flexible reconfiguration of the network topology to adapt to the traffic demand. We discuss the requirements of our architecture for routing algorithms.
- *Routing algorithm for radio access networks.*
  We design a routing algorithm to address the special requirements of radio access networks including our variability-augmented cellular architecture. The algorithm supports the differentiation of traffic classes by means of disjoint routing graphs and provides for near-optimal minimum-delay routing. As a proof of concept the algorithm is implemented in the ns-2 simulation environment. Moreover, we perform a comparative performance analysis of our algorithm with selected state-of-the-art algorithms for multiple topologies and various traffic conditions.

- *Modeling of ad hoc routing.*
  As a basis for our study of dependability in ad hoc networks, we formulate a model of an ideal route acquisition process. The model predicts the route-length distribution, which we put into novel use to describe the overall performance of ad hoc routing systems. We extend this model to cover transmission errors of the wireless channel and the features of a realistic ad hoc protocol. We experimentally validate our model.
- *Investigation of node misbehavior.*
  We formulate an extension to our analytical model covering various classes of node misbehavior. To do so we derive a classification of node misbehavior to suit analytical models. We include inactive, selfish, and malicious node behavior in our model. An extensive experimental analysis studies the effects of node misbehavior for multiple ad hoc routing protocol variants. In particular, we evaluate the trade-off between node misbehavior vs. performance optimization.

1.4 Related Work

Since routing is one of the basic functionalities in communication networks, the network layer is one of the most prominent topics in networking research. Despite the fact that there is a huge body of related work in this area, our special focus on routing dependability issues in cellular and ad hoc networks leaves only little work, which we can compare with directly. At the time of writing we are not aware of any other work investigating routing dependability for next generation wireless networks. For the areas of cellular and ad hoc networks, there exists related work, though. Here, we present the state-of-the-art frameworks that are related in scope to our work areas. We separate the discussion into architectural, algorithmic, and modeling related issues. A more detailed discussion of related work is performed in the corresponding chapters of this dissertation.

Architectural

A couple of routing frameworks for the next generation Internet have been proposed. This includes Nimrod [CCS96], BANANAS [KKW+03], NIRA [Yan03], SNF [JFA03], FARA [CBFP03], and Plutarch [CHM+03]. These frameworks have been designed to replace the current routing architecture of the Internet. As a result, the focus of this work lies on solutions of global proportions. The goal is to provide a powerful routing substrate for a global communication system. Our work focuses on the special requirements of mobile networking in networks of much smaller size, though. Moreover, we discuss the effects of mobility not on global scale but with emphasis on metropolitan areas.

In the area of mobile communications, there are various proposals for beyond 3rd generation network architectures. For example, network operators such as NTT-DoCoMo [YIY01] or hardware manufacturers such as Nokia [Usk03] propose their own architectures. However, these architectures do not extend the scope of the routing domain to the radio access part of the network (even if the end systems are directly addressable using IPv6). Only the core network is considered to be a routing network that follows a packet-switched paradigm. Differing from our work, the radio access network tunnels all IP datagrams to so-called edge routers, which mark the beginning of the core network. In contrast, the MIND project [MBH+02] [MLM+02] assumes a pure IP radio access network as we do. The work discusses mobility and quality of service issues. However, the proposed solution is limited to transfer traditional QoS mechanisms such as DiffServ [BBC+98] and IntServ [BCS94] to the radio access network.

Various research works propose to combine multihop strategies to extend current cellular architectures, ACENET [Yeh02], Multihop Cellular [LH00], iCAR [WQDT01], and the work of Li, Lott et al. [LLW+02] being examples. The proposed solutions are based on IP routing. However, a discussion of QoS and dependability issues misses but only best-effort ad hoc routing algorithms such as [JMH04] or [PBRD03] are proposed. In [HS02] Hsieh and Sivakumar investigate the performance of cellular vs. ad hoc networks. However, their study focuses on the link layer and does not consider the routing system.

Algorithmic

The algorithmic part for infrastructure-based routing networks is well covered in related work. A variety of routing and QoS routing algorithms exists for the wired domain, OSPF [Moy98], OSPF-OMP [Vil02], or MDVA [VGLA01], being examples. However, these algorithms do not consider the topology of the access network to be variable as we do. While algorithms from the ad hoc routing domain [PBRD03] [JMH04] are designed to support variability, most of them only insufficiently address QoS, though. However, there exists previous work to address QoS routing in ad hoc networks as well. For example Chen and Nahrstedt [CN99] propose a distributed QoS routing scheme that is based on ticket-based probing. The protocol proposed in [CN99] is able to deal with variability in topology but operates on flow level, which would induce a very high overhead in our scenario.

Dependability in the area of ad hoc routing cannot directly be compared with dependability in infrastructure-based networks. We are interested in analyzing the performance of the network if non-cooperative and malicious nodes participate in the network. There exists various work to study the influence of node misbehavior from Michiardi et al. [MM02], Marti et al. [MGLB00], Kargl [Kar03], and Wang, Bhargava et al. [WLB03]. However, this work sticks to simulation studies that are restricted to very small networks. In contrast, we present an analytical model to describe the node misbehavior and perform an extensive simulation study for medium to large scale networks. Additionally, we go beyond the basic protocol mechanisms and also include various protocol optimizations in our work. While the mitigation of misbehavior in ad hoc networks is beyond the scope of our work, there exists work such as [XN03] that fits nicely with our results and can be used to increase the dependability of the network.

Modeling

From a perspective of basic research, a body of work has been devoted to map the dependability problem to graph theory to allow for analytical evaluation. Kelleher in [Kel91] surveys the related work. However, this work cannot be easily transferred to real world routing systems because of the limited set of assumptions. Nevertheless it yields interesting results for basic graph/network properties such as connectivity or network diameter.

Orthogonal to the areas of cellular and ad hoc networks there exists some loosely related work in the area of overlay networks to support mobile communications, Stoica's work about an Internet Indirection Infrastructure [SAZ+04] being one example. However, due to the special capabilities and functionalities that distinguish overlay networks from physical networks, this work does not apply to our scenario. Hence, we do not further explore this work.

The review of related work shows that the challenges outlined earlier persist and need to be solved to enable next generation wireless networks. In particular, we perceive the solution of the problems coupled to routing dependability to be of high importance.

1.5 Structure of the Dissertation

This dissertation is structured in four main parts. In Part I we introduce our work and presents the necessary foundations in the areas of dependability and mobility. We present the main contributions of our work in Part II and Part III. These parts cover the areas of cellular networks and ad hoc networks, respectively. We conclude the dissertation in Part IV and give the appendices to our work. See Figure 2 for the coarse structure of the dissertation.

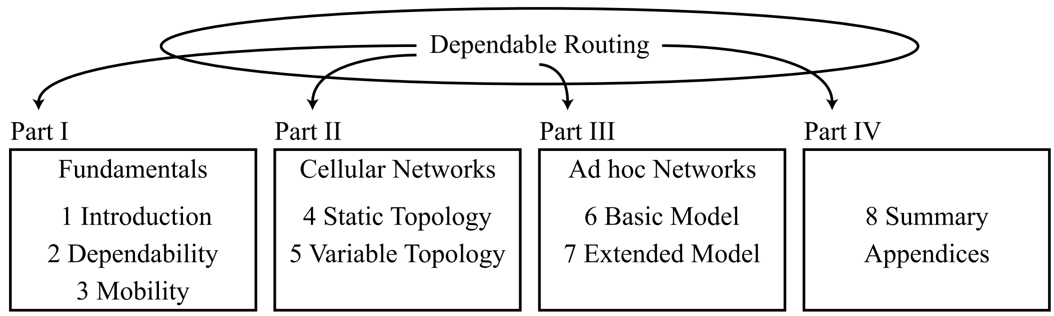


Figure 2: Structure of the dissertation.

The more detailed structure of the dissertation is as follows. Chapter 2 and Chapter 3 form the first part of our work. In Chapter 2 we define the concept of routing dependability, which serves as key concept of our work. Two major domains of mobile communication networks are covered by our investigation: cellular networks and ad hoc networks, respectively. A synthesis of existing results from dependability research is performed and the results are transferred to our application domain. In Chapter 3 we present the challenge of mobility as perceived in routing networks. We investigate the effect of user and end system mobility, which we find is coupled with the concept of dependability. A realistic mobility/workload model is formulated and the mobility induced dynamics are analyzed. We instantiate the model to provide the workload for an experimental analysis of routing architectures and -algorithms in subsequent chapters.

We investigate routing dependability in the area of cellular networks in Part II, that is, in Chapter 4 and Chapter 5. Departing from existing radio access networks an evolutionary architecture to enhance the routing dependability for cellular networks with static topology is designed in Chapter 4. We discuss and study the impact of various routing strategies to leverage the capabilities of this architecture by means of an extensive simulation study. In Chapter 5 architectural concepts from the domain of variable topology networks are introduced to deal with the effects of mobility described earlier in Chapter 3 and to address the shortcomings identified in Chapter 4. Moreover, we formulate a novel near-optimal multiclass minimum-delay routing algorithm to operate on top of the variable infrastructure proposed. The algorithm is thoroughly analyzed by means of a simulation study.

Part III covers the investigation of routing dependability in ad hoc networks in Chapter 6 and Chapter 7. In Chapter 6 we propose an analytical model of the route acquisition process in ad hoc networks to serve as a powerful and generic tool for analysis of routing protocols. The model is validated by means of simulation and instantiated for further usage. The model is extended to cover the effects induced by node misbehavior for usage in the context of routing dependability in Chapter 7. We give a classification of node misbehavior and study the effects induced by non-cooperation of nodes by means of an extensive simulation study.

Part IV concludes our work. In Chapter 8 we summarize the results and outline our contribution to the areas studied. Moreover, we give an outlook to future work on top of our results. Subsequently, the appendices to our work are presented.

2 Routing Dependability

As described in Chapter 1, the social and economic welfare heavily depend on communication and information technology. This dependency becomes more important with the ever increasing integration of digital communications in all days life. Chapter 1 also introduced the technical scope of our work, namely cellular and ad hoc networks to support mobile communications. We motivated our interest in network layer issues and highlighted the special focus on routing. In this chapter we establish the concept of routing dependability, which is not well understood in the domain of cellular and ad hoc networks, yet. Based on related work in the area of dependable computer systems, telecommunication networks, and the Internet, we derive the most important factors influencing the dependability of the routing system. Moreover, we give precise definitions of the subject matter, which serve as foundation for the remainder of our work. We discuss the dimensions coupled with the dependability of routing systems with respect to the investigated classes of networks. On top of this foundation, we approach actual problems of routing dependability for the cellular network domain in Chapter 4 and Chapter 5 and for the ad hoc network domain in Chapter 6 and Chapter 7, respectively.

2.1 Motivation

The transportation of digital information can be seen as a commodity in the developed countries. There exist multiple types of networks, which provide service to the user. The two most common classes of networks are telecommunication networks and the Internet, respectively. The former group of networks bases upon a connection-oriented paradigm that allows for a highly predictable and reliable service for both, wired networks as well as wireless networks. In contrast, the substrate of the Internet, namely the network layer, is connectionless and unreliable while the transport layer augments additional service paradigms such as reliable and ordered delivery of messages. From a user or application perspective, the internals of the network are not visible. As a result the user tends to chose the service providing the highest convenience—irrespective of the underlying network.

Until only recently the telecommunication world focused on a small set of high quality services, while the Internet provided a diversity of services with differing scope and varying quality. The latter is due to the open and flexible architecture of the Internet that allows for rapid development and deployment of application services compared to the closed and strictly regulated telecommunication networks. For novel services such as multimedia services [SN04b] [SN04a], peer-to-peer applications [SW04], or spontaneous and mobile communications [Wei91] neither networking paradigm is perfectly suited, though. While telecommunication networks lack the flexibility to introduce novel services easily, the best-effort nature of the Internet's routing layer poses severe constraints to the achievable performance region of the applications under harsh conditions. Providing adequate quality of service in challenging environments can be considered a hard problem, which may not be solvable at all, if the necessary foundations are not available. The aspect of dependability is one of these foundations but not studied sufficiently yet for the environments we imagine. Especially the class of wireless networks to support user and end system mobility features various dimensions of characteristics/constraints that strongly influence the routing system.

In this chapter we introduce the concept of *routing dependability* to describe the trustworthiness of a routing system such that reliance can justifiably be placed on the consistency of behavior and performance of the routing service delivered. The contributions of this chapter are as follows:

- We give *basic definitions of routing systems* and study the *roots of dependable communication systems*. We include the *related concepts of survivability and trustworthiness* in our discussion.
- We derive and define a *conceptual model of routing dependability.* In particular, we base on related work in the areas of *telecommunication networks* and the *Internet*.
- We derive the *predominant characteristics* that *contribute to routing dependability* in *infrastructure-based cellular networks* and *ad hoc networks.* We extend our survey to cover the characteristics of *future wireless networks* as well.

- Based on our findings we derive a set of *important dimensions that influence routing dependability* in *cellular* and *ad hoc networks*. We also define the areas where we *address the deprivation of dependability* coupled with these dimensions.

2.2 Outline

This chapter is organized as follows. We give precise definitions to define the boundaries of our work in Section 2.3. This includes particularly the conceptualization and modeling of dependability in the area of routing. In Section 2.4 we describe the most important characteristics of cellular and ad hoc networks that effect routing dependability in these networks. We also include next generation wireless networks in our study. Subsequently, in Section 2.5, we define the dimensions of routing dependability that are further studied in this dissertation. We summarize our findings in Section 2.6 and discuss the next steps in addressing dependability in highly dynamic systems. The coarse structure of this chapter is shown in Figure 3.

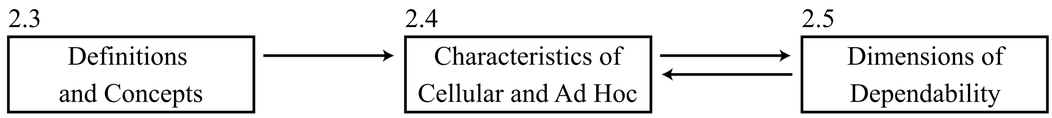


Figure 3: Structure of Chapter 2.

2.3 Dependability in Routing Systems

This section defines the meaning of dependability in the context of this dissertation. We give working definitions of routing systems that mark the boundaries of our research. After a brief discussion of the roots and the meaning of dependability and trust in common, we derive a conceptual model of *routing dependability,* which is the core of this chapter and serves as the principle definition for the remainder of the dissertation. We approach the concept theoretically and from the perspective of related work. For the domain of routing networks we discuss the most prominent definitions, which are based on work from the International Telecommunication Union (ITU), the Internet Engineering Task Force (IETF), as well as various research in this area.

2.3.1 Routing Systems

We define the term routing system using an end-to-end perspective. There exist a source and a destination node, each hosting application processes. These nodes are attached to the routing system or network. To allow for communication, the delivery of messages from source to destination is a logically required architectural functionality for the network. We denote this with the term *routing* and the functional components of the system as *routers*. In particular, the process of routing consists of (a) a service for *identity resolution (resolution of the address where a uniquely identifiable node can be* *reached)*, (b) the *computation* of an appropriate *path*, and (c) the capability to *forward (transport) messages* through the system. Let us define a (unicast) routing system to be:

Definition (1). *"A **routing system** delivers messages from a source node to a destination node by means of networked intermediate nodes (**routers**), which implement the functional process (**routing**) of identity resolution, path computation, and message forwarding."*

We can mainly distinguish two types of service provided from the network and routing system [Per99] [Tan03]:

- The service can be *reliable* or *best effort (unreliable, datagram)*. A reliable service model guarantees the delivery of packets without duplicates and in temporal order. A best effort network delivers the packets as they arrive at the destination.
- The service can be *connection-oriented* or *connectionless*. Connectionless communication uses individual packets that are transmitted independently. Connection-oriented communication establishes a connection prior to message delivery.

Figure 4 shows two possible instantiations of a routing system employing a single-path routing strategy and a multipath routing strategy, respectively. The former strategy models today's prevalent Internet routing paradigm and is shown with the black packets. The latter one is pictured with gray packets. For the remainder of this work, we connote routing system synonymously with the unreliable and connectionless Internet model, which is the subject of our investigation. In particular, we assume the routing decisions to be decentralized and distributed. The datagram routing in the Internet is transparent to the end systems and as a consequence the routing system/network is often treated as a black box as shown in Figure 4. However, the black box model is not sufficient for the investigation of routing dependability. Hence, we aim to have at least a translucent view into the routing system. Our perspective also separates between the data plane (end-to-end traffic between source and destination) and the control plane (control traffic between

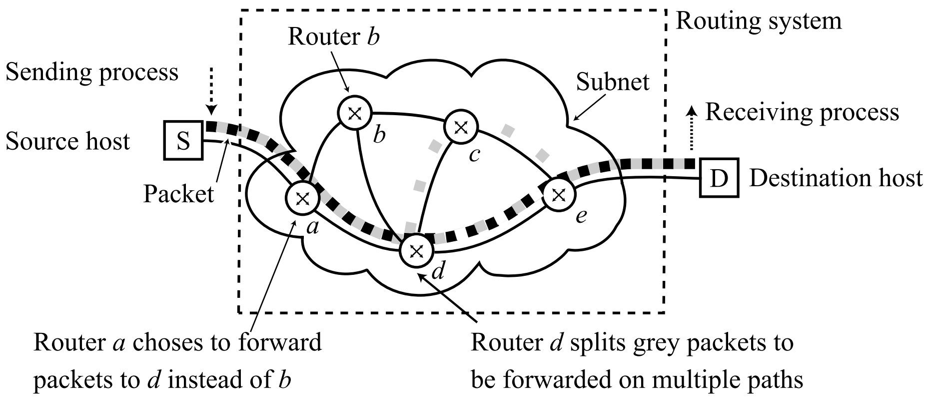


Figure 4: Sample routing system. The black packets show the standard Internet routing and follow the shortest path. The grey packets visualize a multipath routing behavior for a cost metric different than hop count.

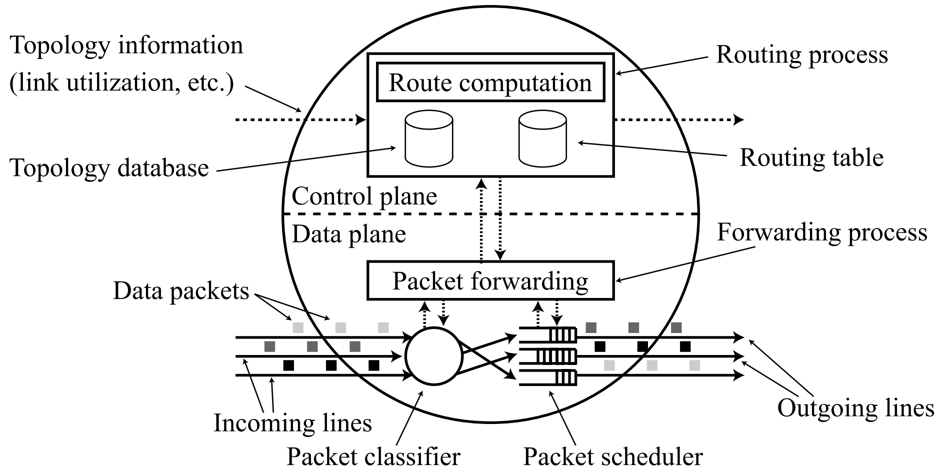


Figure 5: Conceptual model of a router. We distinguish between control plane and data plane. The control plane includes the functionality to exchange topology information and to compute the routing table. The data plane includes the functionality to forward data packets according to the routing table.

individual routers) of the routing system. Despite the presence of various cross-layer interactions, especially with regard to lower layers, we consider the routing service to be a network layer discipline only.

Figure 5 shows the conceptual model of a router. The individual strategies and procedures to implement routing can vary. Usually one distinguishes into a routing algorithm or strategy to compute the next hops or paths and a routing protocol to distribute the captured dynamics of the routing system. Both, algorithm and protocol reside on the control plane of a router/routing system. The data plane of a router implements the forwarding directive according to the calculated next hops/paths. As shown in Figure 5, a router might classify the incoming datagrams and enforce routing policies by means of different schedulers depending on the type of service.

2.3.2 Conceptual Model of Routing Dependability

The root of the adjective *"dependable"* is dated back at least to 1735 [MWE02], its meaning being *reliable*. In modern written and spoken english the meaning of dependable has slightly evolved and is commonly referred to as:

Definition (2). *"**Dependable**—worthy of reliance or trust." and "**Dependable**—consistent in performance or behavior."* [WN04]

These definitions form the basis for defining the notion of dependability in a technical sense. They emphasize the value and importance of dependability, that is, a failure in dependable operation may lead to the failure of the overall system. Moreover, the consistency of performance and behavior conveys the intention of dependability to be measured and guaranteed as a technical metric. To be able to qualify and quantify the dependability in the context of technical systems, it is necessary to focus on the individual characteristics defining a particular system. As already stated in Def. (2), dependability is closely connected to the concept of trust. We borrow the definition of trust from Josang [JP04]:

Definition (3). *"**Trust**—the extent to which one party is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible."* [JP04]

For routing systems, the concept of trust conveys aspects from an end user as well as network perspective. Within our investigation we focus on the latter aspects. In static environments trust relationships may be preconfigured and controlled from the network operator. In contrast, the notion of trust is implicitly subjective and dynamic in nature in networks with heterogeneous and/or autonomous nodes operating in distributed and decentralized fashion. Here, each node can adapt its trust level based on different factors like prior knowledge and context information, which are dynamically increased or decreased by collaboration and observation of each node. A more general discussion of trust and dependability related issues in the digital domain can be found in [Sch00].

The origins of the technical meaning of dependability can be traced back to the early days of computing and communications [ALR00]. In the context of the early and pioneering work of Babbages, Larnder in 1834 proposed to eliminate errors in computation by using separate and independent computers and even more decisive by using different computation methods. The early electronic computers and communication systems used highly unreliable components. As a result the research focused on enhancing the reliability and dependability of operation. Outstanding basic theories of redundancy to enhance the reliability of logical structures, to mask faults, and to enhance the quality of communication have been developed from von Neumann, Moore, Shannon and their successors. Today, the fundamental concepts of dependability in computer systems are discussed from a technical perspective in various research groups and committees including the International Federation for Information Processing's working group on dependable computing and fault tolerance (IFIP WG 10.4) [IFIP04] as well as in the IEEE's technical committee on fault-tolerant computing (IEEE TC-FTC) of the IEEE computer society [IEEE04]. Recent advances of dependability in distributed systems are covered by Suri et al. [SWH95].

The concept of dependability discussed is closely related to the concepts of survivability [EFL+99] and trustworthiness [Sch99], which are defined as follows:

Definition (4). *"**Survivability**—the capability of a system to fulfil its mission in a timely manner."* [EFL+*99*]

Definition (5). *"**Trustworthiness**—the assurance that a system will perform as expected."* [Sch99]

Both concepts also introduce impairments or threats, which thwart the dependable operation of the system. These may be attacks, failures, environmental conditions, or accidents to name a few. In general, these are determined by the characteristics of the system and the external influences surrounding the system. This includes the system's design as well. Dependability, survivability, and trustworthiness are three different names describing the same essential property of a system and cannot easily be distinguished. For our discussion we have chosen to use the term dependability, which is defined below for our special application:

Definition (6). *"**Dependability** of a computer system is the ability to deliver service that can justifiably be trusted."* [ALR00]

The work of Avizienis, Laprie and Randell [ALR00] as well as earlier work of Laprie [Lap85] discusses a systematic approach to describe dependability. See Figure 6 for an adoption of the dependability trees introduced in [Lap85] and [ALR00]. The three main parts of dependability are the impairments (threats) to the system's dependability, the dependability goals (attributes) of the system, and the means by which dependability is reached. The impairments are categorized into faults, errors and failures. The faults and their sources may vary. They lead to errors of parts or components of the system, while the failure is the transition from correct to incorrect or optimal to suboptimal function of the overall system. The impairments can also be investigated with respect to their causes, consequences, and characteristics. The means to achieve dependability are to deal with the faults. In detail, this includes fault forecasting, fault prevention, fault tolerance, and fault removal. We add the categorization in proactive and reactive measures to refine the original model of [ALR00]. The attributes of dependability are manifold and partially overlap. In [ALR00] they are defined as availability, reliability, safety, confidentiality, integrity, and maintainability. These basic attributes are closely related to other work areas, such as security [Eck04] and quality of service [Sch01]. It is important to notice that we need to put different emphasis on these attributes depending on the intended sys-

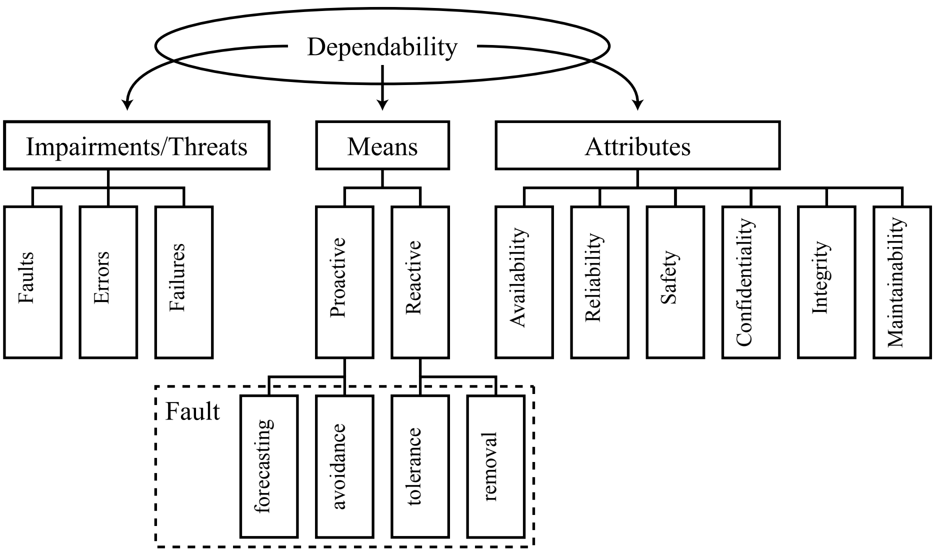


Figure 6: Dependability tree. Impairments describe the threats to dependability, means describe the mechanisms to achieve/restore dependability, and attributes specify the dependability goals in a system.

tem and its application. The description of some abstract system characteristics such as robustness cannot directly be performed from the basic attributes, though. Related work from Sanders et al. [DDD+00] [SM91] implements tools to make these basic models applicable to real world systems.

Dependability in Telecommunication Networks

For traditional public switched telephone networks (PSTN) and integrated service networks (ISDN), the ITU defines concepts related to quality of service and network performance including dependability [ITU94] to aid in planning, provisioning and operation of telecommunication networks. See Figure 7 for a visualization of the ITU's conceptual QoS model [ITU94]. The basic model for performance concepts of the ITU has four major building blocks that are related to our work: quality of service, serveability, trafficability performance, and dependability. The individual building blocks can be described as follows. *Quality of service* is the most abstract concept in the model and describes the satisfaction of a *user* of the *service*.

Definition (7). *"**Quality of service**—the collective effect of service performance, which determines the degree of satisfaction of a user of the service."* [ITU94]

The service related primitives of QoS are described with the concept of *serveability*, which includes the components *service accessibility performance*, *service retainability performance*, and *service integrity performance* [ITU94]. To be able to maintain a certain quality of service level, the perspective of items (infrastructure components) is described in the network performance part of the diagram. The *trafficability performance* building block acts as technical description of the ability of an infrastructure component

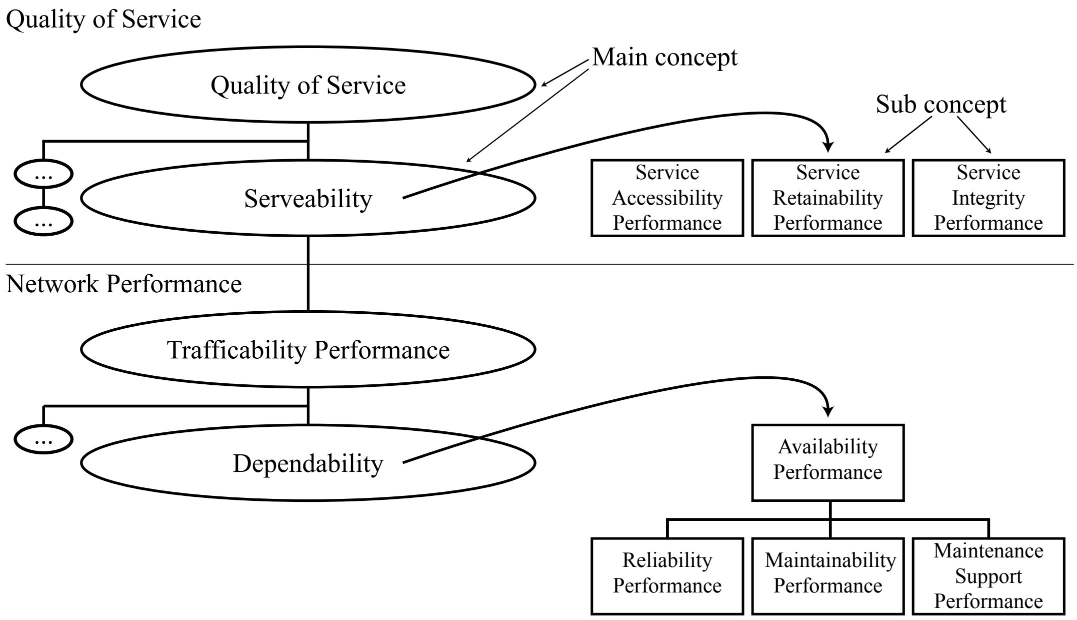


Figure 7: The ITU's classification of quality of service. Each concept may affect the ones above collectively or individually.

to deliver a certain performance level. Finally, the foundation of the aforementioned concepts is given by the concept of *dependability*, which is further refined into *availability performance*, *reliability performance* and two *maintainability* related blocks. We take the definitions from [ITU94] to describe the outlined concepts:

Definition (8). *"**Trafficability performance**—the ability of an item to meet a traffic demand of a given size and other characteristics, under given internal conditions."* [ITU94]

According to [ITU94], dependability is the key performance measure for the trafficability performance concept and can be defined as follows:

Definition (9). *"**Dependability**—the collective term used to describe the availability performance and its influencing factors: reliability performance, maintainability performance and maintenance support performance."* [ITU94]

Definition (10). *"**Availability performance**—the ability of an item to be in a state to perform a required function at a given instant of time or at any instant of time within a given time-interval, assuming that the external resources, if required, are provided."* [ITU94]

Definition (11). *"**Reliability performance**—the ability of an item to perform a required function under given conditions for a given time-interval."* [ITU94]

We exclude the maintenance related performance in our work, that is, the restoration of the function by means of maintenance, since our focus is on the technical but not the operational aspects of dependability. In the context of dependability in telecommunication networks the terms *availability* and *reliability* are used as the respective performance measures. Unfortunately, the ITU definitions are tightly coupled with operational considerations as well as with resources and facilities of telecommunication networks only. As a consequence they are not generally applicable to scenarios outside the telecommunication sector.

Dependability in the Internet

Likewise in telecommunication networks, we perceive dependability in the Internet to be an enabler for higher level concepts such as, for example, quality of service. However, Def. (7) cannot be easily interpreted in the engineering domain. A technically more precise definition of QoS was introduced by Schmitt [Sch01]:

Definition (12). *"**Quality of service**—the well-defined and controllable behavior of a system with respect to quantitative parameters."* [Sch01]

The current Internet lives without QoS, which is in part due to the complexity of the QoS concepts proposed. Moreover, the basic routing of the Internet can hardly be considered dependable. For example, the predecessor of the Internet, the ARPANET, suffered from catastrophic failures because of its routing protocol, which could only be repaired with manual interaction (see, for example, [Per99] for details of this malfunction). Based on this experience, the Internet community decided to require routing protocols to fulfil some basic dependability criteria such as, for example, the ability of the protocol to stabilize after the failure condition is removed (self-stabilization). Influenced by the failure of the ARPANET, routing protocols for the Internet have been kept very simple, though. Existing work in this area mainly focuses on routing security and operational considerations (see [Shi00]). In [BMY04] the focus lies on security in routing protocols but excludes the overall routing system explicitly. In contrast, the work of Perlman [Per99] fits well in the scope of our work. According to her, a (routing) network design process should obey the following principles:

- *Scope, scalability, robustness, autoconfigurability, tweakability, determinism*, and *migration*.

These goals approach dependability from various angles. Autoconfigurability describes the network's ability to operate in plug and play fashion without manual intervention. Tweakability describes the possibility to manually optimize the system if the autoconfiguration does yield only suboptimal results. The goal of robustness is subdivided into four subgoals that are:

- *Safety barriers*, which hinder the spreading of faults.
- *Self-stabilization* after the defect or malfunctioning device is eliminated.
- *Fault detection* as an ability of the network.
- *Byzantine robustness* in case of improperly operating components or attacks.

See also Kenyon [Ken02] for a more operational perspective on performance and reliability characteristics of internetworks. There exists other work in the area of dependability for smart networks by Helvik [Hel99], which focuses on the dependability of the network transport underneath the routing system. His results are based on earlier general work in the area of dependability [Lap85]. Helvik closely relates dependability to the survivability of the core transport functionality even for failure conditions. He presents various mechanisms to enhance the survivability of the network. However, Helvik investigates only connection oriented paradigms in his work.

In summary, the important factors of dependability contribute to higher level concepts such as quality of service or security. The fundamental requirements of a reliable, predictable, and in terms of traffic well performing system form the intersection that allows to bootstrap these higher level concepts. Given a basic dependability we are able to move forward in both domains—quality of service as well as security—to cover special aspects in greater detail. For our work we do both: investigate the core of dependability, namely the operational requirements to ensure a baseline dependability for the routing system as well as address higher level optimization goals coupled with the concept of dependability. For within our work we define routing dependability to be:

Definition (13). *"**Routing dependability**—the trustworthiness of a routing system such that reliance can justifiably be placed on the consistency of behavior and performance of the routing service it delivers."*

Def. (13) is open enough to embrace all types of networks we investigate. This includes hybrids of different network classes such as beyond third generation (B3G) cellular systems that integrate ad hoc concepts. For general applicability, we distinguish between the data-plane and the control-plane of the network in our investigation. In each network class the dimensions of dependability may vary depending on the inherent characteristics of the network's routing system. However, the concept of routing dependability is often overlooked. In summary, viable solutions for higher level concepts such as network QoS and security are only possible if the groundwork, in particular, sufficient mechanisms to ensure the dependable operation, is laid.

2.4 Characteristics of Cellular and Ad hoc Networks

Having derived the general concept of routing dependability, we further refine this concept to be able to capture the important aspects for the case of cellular and ad hoc networks. We give concise definitions of these network classes and describe their most important characteristics. Subsequently, we deduce the most important dimensions related to routing dependability within these networks.

2.4.1 Characteristics of Cellular Networks

We refer to cellular networks[5] as the wireless extension to the traditional telecommunication networks. During the last decade these systems have been tremendously successful. The number of mobile subscribers has already surpassed the number of landlines according to the GSM Association [GSM04]. Cellular networks are built around infrastructure components. Their primary goal is to enable seamless voice-communication. They support user mobility and roaming between different networks/providers. Because of their roots in telecommunication networks, the system architectures are mainly hierarchical and tree-structured, network control is centralized, and the predominant service paradigm is connection-oriented [EVB01]. In current state-of-the-art networks we have a "smart" core and a "dumb" edge, the edge referring to the access network consisting of the base stations and the mobile equipment (see Section 4.6 of our work for a more detailed introduction to cellular network architectures). Only recently, the agenda of cellular network development and operation has been refined to include data communication as well. The most important characteristics surrounding cellular networks can be summarized as follows (see, for example, [3GPP04] for the corresponding specifications

5) We use the term cellular networks interchangeable with radio access networks in the remainder of this dissertation.

of second and third generation cellular networks or [Sch03] for a more general textbook covering this topic):

- The network deals with *user* and *end system mobility.*
- The *communication channel* is *wireless* but *controlled by infrastructure.*
- The network is *closed,* only subscribers participate.
- The network *architecture* relies on an *infrastructure-based* paradigm.
- The *topology design* is *hierarchical and tree-structured.*
- The *routing/forwarding system* is *centrally controlled*, which leads to the notion of an *"dumb" edge* and a *"smart" core* of the network.
- The *access* network is a *non-routed* network and only the *core network* provides for *routing functionality.*

Cellular networks are highly managed and maintained networks. The operators provide for fault tolerance by means of redundant components. As a result, the dependability of the routing system is very high. Please keep in mind, however, that failures in central control may lead to catastrophic failures of the system. Moreover, the achieved performance of data services is only suboptimal because optimizations are only applied in the routing core of the network.[6]

2.4.2 Characteristics of Ad hoc Networks

Likewise performed for cellular networks in Section 2.4.1, we now discuss the special characteristics of ad hoc networks. The term *"ad hoc network"* or *"mobile ad hoc network"* (MANET) can be seen as an umbrella to cover various species of networks. This includes, for example, inter-vehicular environments [HBE+01], disaster recovery, multimedia home entertainment, and zero-configuration personal area communication. Furthermore, there are also proposals for wide area ad hoc networks [HGBV01]. All these networks have certain demands in common: either there is impromptu need for communication, or the absence of infrastructure commands that the network has to be fashioned from whatever resources are immediately available. Moreover, the autonomous and cooperative operation is inherent to the network nodes, which are terminals (end systems) and routers (intermediate systems) at the same time. Likewise performed for cellular networks, we are interested in the predominant network characteristics. These can be summarized as follows:

- The network operates *infrastructure-less.*
- The network is composed by *nodes* that are both, *end systems* and *routers*.
- The network is *open*, everybody can participate.
- The nodes need to *cooperate* to allow for *multihop routing.*

6) Imagine you read this dissertation and want to hand it to a colleague sitting next door to you. Following today's cellular network paradigm the postal service would fetch the dissertation and bring it to a central post office where the routing would be performed. Finally, the postal service would deliver the dissertation to the recipient.

- The nodes may be *mobile* with unpredictable speed and direction of movement.
- The *communication channel* is *wireless.* Transmission errors, limited range, hidden and exposed terminals, etc. are *not controlled* by *infrastructure*.
- The nodes or the communication between nodes may be *impacted* by *external forces* such as environmental conditions.
- The network has to support *heterogeneity* in nodes, communication mechanisms, connectivity, etc. This induces severe *constraints* in computational power, energy supply, bandwidth, link asymmetry, etc.
- The *autonomy of nodes* can lead to *intermittent availability/unpredictable failure of nodes*.
- The routing system operates *self-organized*, that is, there is need for *adaptation* to *changing network conditions* on various *timescales*.
- The network has to deal with special *application characteristics*: peer-to-peer applications, geocasting, etc.

The characteristics are dominated by the absence of infrastructure, the highly dynamic nature of the topology, the heterogeneity of the devices, and the dependency on the wireless communication channel. Moreover, the system is qualified by the need for cooperative operation of the nodes to allow for multihop operation of the network. Most routing protocols do silently assume only well-behaving and cooperative nodes.

2.4.3 Characteristics of Future Cellular and Ad hoc Networks

We expect that cellular and ad hoc networks are integrated to form future network architectures [Yeh02] [WQDT01] [LH00] [LLW+02] [JMH04]. The main challenges addressed by these architectures are the heterogeneity of communication technologies, the variability in topology, and the mobility of users, end systems, and networks. Moreover, these architectures anticipate a convergence of cellular telecommunication networks with the Internet. While the core of such architectures is usually seen as relatively static, the edges are designed to be highly adaptive to allow for seamless integration of ad hoc networks, etc. As a consequence, the network experiences high dynamics in various dimensions. The characteristics of future network architectures are:

- The network has to deal with *heterogeneity* in *access*, in *protocols*, in *components*, in *services*, in *algorithms*.
- The architecture is *open* and *scalable* and provides a *framework for interworking.*
- The network experiences *extreme dynamics* in various areas including topology.
- Some network components have *huge resources* (computational power, memory, storage, bandwidth, energy supply) while others are *constraint in resources* as described above for ad hoc networks.
- The network has to deal with a multitude of *application characteristics*.

In summary, hybrids of ad hoc and cellular networks have to cover an incredibly huge spectrum of functional and non-functional requirements. Particularly, the design of routing systems has to keep up with the resulting dynamics. Although the vision of future cellular and ad hoc networks is discussed at large, the consequences for the dependability and performance aspects of the routing system are currently not in the focus of research.

2.5 Routing Dependability in Cellular and Ad hoc Networks

We identified the most important characteristics of cellular and ad hoc networks in the previous subsections. Based on these characteristics we are able to uncover some important dimensions of dependability for the surveyed networks. In particular, we focus on the dimensions of (1) *network control*, (2) *dynamics of topology*, and (3) *node autonomy* in our work and specially concentrate on the evaluation of impairments related to these dimensions. In routing systems the first dimension can be closely related to the centralized or distributed nature of the routing strategies and algorithms while the second dimension is coupled with the protocols that capture and distribute the state of the network. The third dimension is of special interest in distributed routing systems such as ad hoc networks. Our further emphasis on specific aspects of these dimensions is outlined in the corresponding chapters of this dissertation.

- The dimension *"network control"* includes the routing architecture and strategy. The surveyed systems cover the full range from strictly hierarchical, tree-structured, and centrally controlled cellular networks to the spontaneous formation of ad hoc nodes operating in autonomous manner. While the network architecture and design can be seen as a relevant factor on a longer timescale, there are as well medium-term constraints such as traffic engineering and short-term issues such as adaptive routing.
- The dimension *"dynamics of topology"* covers aspects such as, for example, user and end system mobility. It is obvious that the topology dynamics are directly related to the dynamics of the routing system itself, which in fact influences the network dependability. Moreover, the wireless nature of the communication channel contributes to the link stability and error rate in mobile ad hoc networks. The possibility to run out of energy and the vulnerability of ad hoc nodes to environmental conditions or adversaries is of importance as well. Moreover, in autonomous systems without centralized control, the misbehavior of individual network nodes may cause fairly complex problems.
- The dimension *"node autonomy"* describes the ability of the individual system entities to allow for dependable self-organization of the routing system. While the self-organized formation of routing networks has been studied in various flavors in the field of ad hoc communication networks, the concept of self-organized dependability or trust has not been sufficiently addressed yet. We perceive the degree of autonomy to contribute to the most complicated challenge in dependable routing, though. Autonomy is closely intertwined with the openness of the system as well as the cooperation among nodes.

There clearly exist orthogonal or partially correlated dimensions such as the heterogeneity of end systems or nodes, or the open/closed nature of the system, which also contribute to the problem as such. We do, however, limit our discussion to the core of routing systems, which we specify by network design, routing strategies, and node behavior. Figure 8 shows a qualitative graphical representation of the first two dimensions of the investigated classes of networks (see [HMKR04] for an extended version that also covers sensor networks). Because the classification of the surveyed networks is subjective, it has been discussed in a panel of experts.

While the knowledge of the factors influencing routing dependability is of utmost importance towards more dependable routing systems, it is only the first step. The second equally important step is the development of mechanisms and systems, which establish the envisioned trustworthiness in the consistency of behavior and performance of the routing system. There is no panacea to cure all surveyed types of networks immediately. However, substantial improvements can be achieved by addressing the particular characteristics, which have been identified to be the weak link of the system. Departing from topology dynamics and network control, we now describe the special area of interest of our work. We focus on a subset of dependability related issues for the case of cellular networks and ad hoc networks.

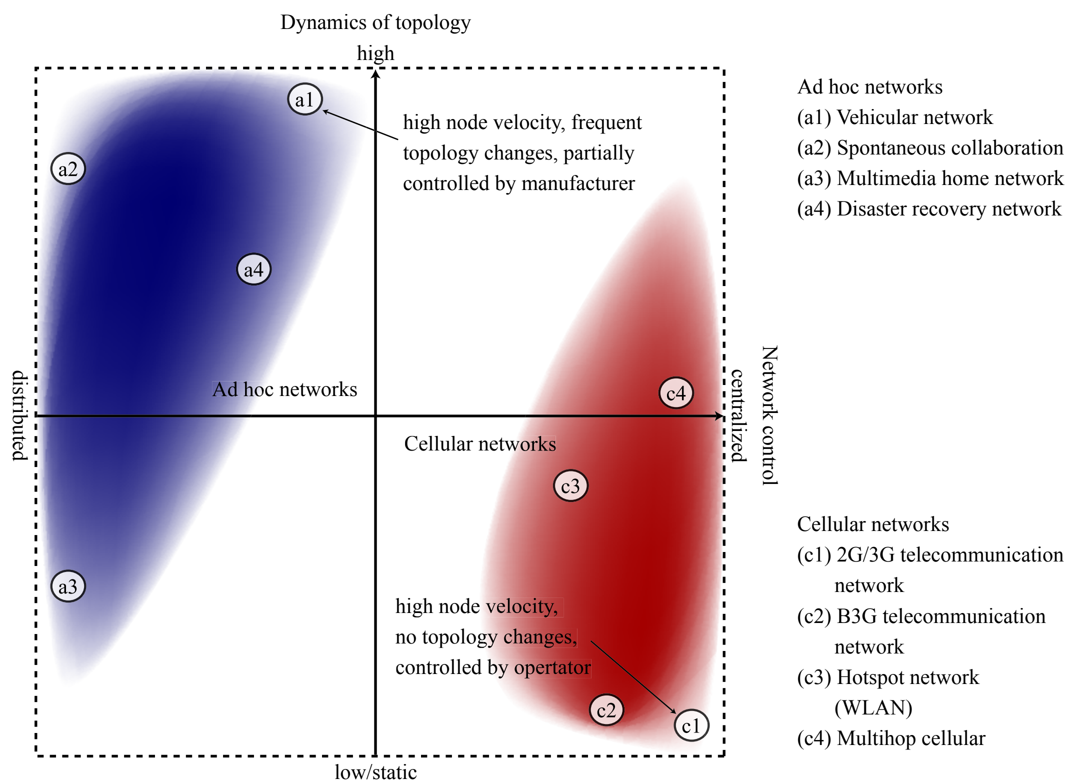


Figure 8: Classification of the surveyed network classes. We show the dimensions of network control and topology dynamics.

2.5.1 Dimensions of Dependability in Cellular Networks

Today's cellular networks are tightly controlled by network providers. Hence, the dimension of network control is of particular importance for our study of future cellular networks. As described earlier, the topology is mostly static in these networks. From a technical perspective, the centralization of functionality allows to build highly reliable, available, and dependable systems by means of redundancy and highly available system components. In our work we focus on the *routing* aspects of cellular networks. In existing cellular systems the message routing is part of the core network only. The existing dynamics in end systems and wireless communications are not directly affecting the routing system because the radio access network is a non-routed network. Nevertheless, secondary effects of mobility can be observed. These effects are due to potential hot spots that lead to a degradation of the routing performance of cellular networks, because the routing cannot easily adapt to the traffic fluctuations. The performance aspects of dependability are of major importance in the area of cellular networks. However, if we regard future cellular networks, topology dynamics need also to be carefully investigated. Drafts for next generation cellular networks propose to introduce variable and mobile base stations. At the same time the routing domain of cellular networks is expected to move towards the edges of the network. In Chapter 5 and Chapter 6 of our work we emphasize on the problems and possible solutions for future cellular networks.

The performance degradation of today's cellular networks because of hot spots and the expected routing nature of future cellular networks make a case for a closer investigation of routing dependability issues. The dependability of existing cellular systems sets a very high benchmark because of the closed network paradigm, the tight network control and the ownership of infrastructure. Moreover, the related network performance aspects of dependability for this class of networks are of interest. In particular, we identify two interesting research directions coupled with routing dependability in cellular networks. First, one main research direction in cellular networks is to optimize the utilization of the deployed infrastructure components. Our investigation aims at smarter systems than available today. Novel network designs propose that the routing process starts at the edge of the network. This leaves room for optimization of the network performance; a smart cellular network may gain in network performance while maintaining the current level of dependability. This increase in performance can lead to an increase in perceived QoS from a user's perspective and, thus, to an increase in perceived dependability. Second, possible dependability gains of cellular networks can be achieved by introducing a simplified and self-adaptive network control. While this does not necessarily increase the dependability of the network under normal conditions, it may allow for self-healing and self-organizing network operation, which includes the possibility of increased dependability in case of failure of the central network control components.

In summary, for cellular networks, our goal is to deliver better performance of the routing system. That is, to optimize the throughput of the network while ensuring the dependability and stability of the system even in case of failures. For our investigation, we concentrate on the wireless and wired infrastructure part in the radio access network of the cellular networks. There the impact of mobility can directly be observed. Objectives are the investigation of the self-healing and self-adaptive operation of infrastructure components and corresponding routing strategies. In summary, for the case of cellular networks we investigate the following "impairments" to routing dependability.

- *Mobility.* We focus on the load fluctuations induced by user and device mobility. Our target scenario is a metropolitan area radio access network with hot spots.
- *Decentralized network control.* In particular, we investigate the move from centralized to distributed and decentralized control. We consider a *routed radio access network*, that is, the routing domain of current cellular networks is extended from the core network to cover the radio access network as well.
- *Variable network topology.* We assume that the radio access network can be reconfigured on demand. We consider a meshed network structure to be formed using wireless point-to-point communication links.

2.5.2 Dimensions of Dependability in Ad hoc Networks

Similar to cellular networks, the dimensions of dependability in ad hoc networks are determined by the key characteristics of the underlying networking paradigm. Here, our key focus is on topology dynamics and node autonomy. In terms of technological maturity, ad hoc networks are in an early phase of development and have not yet crossed the chasm to be ready for the mass market. For our quest on dependable systems, this mandates for a different approach compared to the cellular network case. Because the concept of dependability is not well established in ad hoc networking research yet, we aim to develop it to this point. While the routing system in cellular system relies on dedicated infrastructure components, the same functionality is realized in ad hoc networks by systems that are both, end systems and routers. Coupled with the harsher constraints of ad hoc environments this strongly influences the routing system's dependability.

We identify the misbehavior and non-cooperation of nodes to be one factor hindering the routing system of ad hoc networks to reach a satisfactory level of dependability. In combination with the above outlined constraints such as decentralized network control and high variability in topology, the routing system cannot be trusted per se. Goal of our work is to allow for a qualification and quantification of the effects of node (mis)behavior on the dependability of the overall routing system. Dependability for this class of networks, thus, includes survivability and trustworthiness in the traditional sense to ensure the minimum requirement of dependable transport of datagrams. We aim in investigating building blocks of routing primitives to ensure the self-adaptation of the routing system even in case of failures or attacks. This includes probabilistic mechanisms to reduce the network load as well as to enhance the resilience of the routing system. Again we evaluate the network performance.

While we expect the nodes of future networks to become smarter, the impromptu and self-organized nature of ad hoc networks also exhibits communication paradigms that are not fully understood yet. This includes the establishing of knowledge about the network by means of context information and the visionary concept of establishing dependability in open and distributed systems by means of trust-building on top of only minimal information. The realization of the latter ideas is, however, postponed to future work. In summary, for the ad hoc network domain we investigate the following "impairments" to routing dependability.

- *Node autonomy/misbehavior.* Ad hoc nodes are autonomous and cannot easily be controlled. We study the effects node misbehavior induced by inactive, selfish, and malicious nodes.
- *Adaptation to topology dynamics.* We investigate the influence of protocol mechanisms to optimize the network with respect to node mobility, scale of the network, and robustness of routes.

2.6 Summary

This chapter established the concept of routing dependability, which is the main theme of our work. Based on related work in the area of the Internet and telecommunication networks, we derived a conceptual model of dependability in communication networks. We refined existing dependability definitions to exactly match the case of routing networks for two different types of networks; cellular networks and ad hoc networks. Our definitions as well as the conceptual model may easily be extended to cover other types of routing networks. See, for example, [HMKR04] for an extension of our work to also cover sensor networks, which will not be investigated during the course of this work. Unfortunately, designers and developers of routing architectures and protocols often neglect fundamental concepts and principles such as dependability, availability, and reliability. Driven by the market competition they solely focus on a more advanced feature set. However, the decoupling of fundamental concepts from high-level goals may result in only inferior routing systems. For example, the important yet complex high-level concept of quality of service is likely to fail if the dependability of the underlying network cannot be guaranteed adequately.[7]

Our analysis of dependability revealed open research issues for both classes of networks discussed and we have been able to narrow down the interesting dimensions to allow for proper treatment in our investigation. For infrastructure-based networks such as, for example, cellular networks, we are especially interested in the dimension of network control, which is closely coupled with the network performance. Novel networking architectures are investigated in Part II of this dissertation. The achievable performance

7) The Wireless Application Protocol (WAP) can be considered to be a prominent example for such an unsuccessful service. While the WAP specification covers a huge set of possible and full-featured applications, the user experience was very poor. One important part of the problem was the underlying (circuit-switched) routing system, which was not able to deliver the characteristics for the applications transferred from the packet-switched Internet domain.

region for these architectures is optimized in combination with decentralized and distributed approaches to resource management. For ad hoc networks, our focus shifts to the effects induced by node misbehavior. We carefully analyze these effects and provide for means to adapt the network to mitigate the loss of dependability in Part III of our work.

Having described the most important characteristics of routing dependability, which make up the optimization goal in our work, the next chapter provides for the application scenario that motivates and stimulates our research: the mobility of users and end systems. This stimulus induces various effects coupled with the concept of routing dependability, which is to be studied in the subsequent chapters.

3 Mobility in Cellular and Ad hoc Networks

In Chapter 2 we have introduced the concept of routing dependability. In particular, we derived the most important factors influencing the dependability of the routing system for the special case of cellular and ad hoc networks. In this chapter we study the effect of user and end system mobility, which is tightly coupled with the availability and performance aspects of dependability. We formulate a realistic mobility/workload model to describe metropolitan areas and instantiate it for the case of Darmstadt, a city in Germany. The mobility-induced dynamics on various timescales are described and analyzed within the context of wireless networks. The obtained results are subsequently used to form the workload for an experimental analysis of various routing strategies and to justify the development of novel architectures and algorithms for cellular networks in Chapter 4 and Chapter 5, respectively. The model predictions also describe future usage scenarios for ad hoc networks and motivate our work to push the limits of technology in this area farther in Chapter 6 and following.

3.1 Motivation

The analysis of dependability in cellular and ad hoc networks needs to reflect the characteristics imposed by their special context of operation. The influence of the wireless channel as well as the user and terminal mobility are of utmost importance for both scenarios. Wireless communication differs in many respects from wireline communication, the time varying nature of the channel being the most important constraint. The influence of mobility is limited to moving end systems, that is, handsets in cellular networks, while in the case of ad hoc communication both, the end as well as the intermediate systems are expected to be mobile. The analysis of routing dependability in cellular and ad hoc networks is inherently defined by the characteristics described above. Large scale wireless networks offer broadband capacity while supporting user and terminal mobility. To allow for a detailed study of these networks, we need to carefully model these peculiarities. However, research in this field suffers from the lack of realistic mobility and workload models. Besides the usage for our investigation there is also a strong need for such models to be able to perform sound simulations supporting important yet difficult tasks like network planning and resource management in future networks.

The effects induced by mobility are manifold. These effects cannot be easily captured, however. The pure description of position, speed, and direction serves only as a snapshot of the current mobility situation. Prediction of future mobility patters is much harder [ZS04] [SN02]. This is especially true in our context, where we consider the macroscopic effects of mobility instead of the single-user perspective. Moreover, in mobile and wireless communications, mobility has to be investigated in combination with the wireless technology, the routing architecture, etc. This also includes the perspective of the observer: while a group of mobile users moving simultaneously lay a large burden on a cellular infrastructure because of the synchronized cell changes, they would allow for a stable topology in an ad hoc network if regarded from inside the group.

Proposals for next generation wireless local and metropolitan area networks are often characterized by smaller cells and hot spots. The reduction in size of the radio cells goes hand in hand with an increase in capacity. Also, we expect different traffic than in today's networks that are dominated by voice. As pointed out in Chapter 1 and Chapter 2, we are especially interested in future cellular networks that follow distributed paradigms for network control. Existing mobility models for use in cellular networks focus mainly on traditional teletraffic management, that is, the effects induced on the control plane of centrally controlled networks. Since we do not limit our work to control plane issues only, these models cannot be applied directly. The combination with network traffic to form a workload model makes matters even worse in existing models: either the traffic cannot be clearly separated from the observed mobility patterns in real-world traces or the traffic only accounts for connection-oriented voice communication. The mobility models used in ad hoc networking research are rather limited, too, and do not apply for the macroscopic scenario of cellular networks. These models do, however, strictly decouple mobility and traffic and allow for flexible parameterization of the workload. Summarized, we cannot use most existing models for our investigation because they are neither flexible enough to cover our scenarios nor as accurate as needed.

In this chapter we address the lack of adequate workload models to capture the novel requirements of next generation wireless networks. A novel approach towards realistic modeling of user mobility is proposed and studied. We formulate an analytical model, which is a hybrid of an empirical mobility model and a synthetic traffic model. The hybrid nature of the model is owed to the current state-of-the-art in mobility and traffic modeling as we describe below and results in great flexibility due to the clear separation of the influences of mobility and traffic. This comes at the expense of additional but reasonable complexity, however. The mobility part of the model is based on the combination of statistical zoning information with field data of movement patterns. This allows us to predict the density of users (classified into different groups) for a given area at a given time. We are able to integrate different traffic characteristics on top of our mobility model elegantly. The combination of user density with the predicted (synthetic) traffic of the modeled user groups gives the traffic and fluctuations of traffic throughout the network, thus, describing the workload for the envisioned scenario. The model is described analytically. This formalization allows for easy implementation of our model to cover real metropolitan environments. We perform an instantiation of our model for Darmstadt, a German city of approximately 140,000 inhabitants. Analysis and simulations are provided, which show that the proposed scheme is quite promising. Our findings are that our model is able to cover the macroscopic effects of real-world behavior more precisely than currently available mobility/workload models.

Our contributions are as follows:

- We *review related work in the area of mobility modeling.* In particular, we *study the accuracy, flexibility, and applicability of existing macroscopic and microscopic models* for the intended application in next generation wireless networks.
- We *formulate an analytical workload model*, which is a hybrid of an empirical mobility model and a synthetic traffic model. The *empirical mobility part of the model accurately captures the dynamics induced in metropolitan areas* and the *synthetic traffic model facilitates the integration of flexible traffic models*.
- We perform an *instantiation of our model for Darmstadt*, a german city, and *analyze the model predictions*.

3.2 Outline

The chapter is organized as follows. An extensive survey of related work reveals that the existing models in literature do not provide the granularity of detail in information we need for our work. In contrast to the rest of this dissertation, we advance the related work to Section 3.3 at the beginning of the chapter to emphasize our decision process. Section 3.4 describes our model. In Section 3.4.1 the fundamentals of our model, including the modeling of locations and user behavior are explained. Section 3.4.2 comprises the analytical description of our model. We give definitions as well as equations to exactly represent the model assumptions. In Section 3.4.3 we present the instantiation of our model. The model predictions are described and visualized in Section 3.5. Moreover, an in depth analysis is performed. The chapter is concluded by summarizing the main results and discussing the important aspects for the rest of this dissertation in Section 3.6. The coarse structure of this chapter is shown in Figure 9.

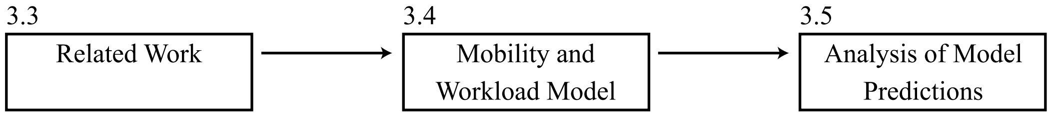


Figure 9: Structure of Chapter 3.

3.3 Related Work

There are different approaches towards realistic mobility models for mobile communications. A fairly comprehensive overview of mobility modeling in wireless networks can be found in the work of Bettstetter [Bet01b] while Camp et al. provide a more detailed survey of some random mobility models [CBD02]. There is, however, no common scheme to categorize these models because they are so manifold. Within our work, we use a categorization into two classes: *microscopic* vs. *macroscopic* mobility models, since these two concepts cover the essential distinction for our investigation.

Microscopic mobility models describe the mobility behavior of individuals. They are often based on analytical descriptions of user movement. Some microscopic models, however, build upon the exact tracking of trajectories of individuals. If microscopic models are aggregated to describe macroscopic behavior, the nature of the model hinders realistic predictions, they are accurate only on microscopic level. Most synthetic or random models fall into this class. *Macroscopic mobility models* describe the aggregated effects of mobility. They are mostly obtained using statistical data collection. To keep the complexity of macroscopic models reasonable, it is possible to aggregate the behavior of individuals instead of trying to track each user on macroscopic scale. While most microscopic models are formulated as general as possible, the value of macroscopic models gains by realistic instantiation. The price to pay for this realism is the loss of generality.

Within our context we are interested to realistically model mobility at the scale of a metropolitan area. Our target application domain are cellular networks, which cover large areas and support a high number of users. Within the context of ad hoc networks, due to scalability constraints of current technology, large networks are currently not tractable. Hence, we use microscopic models in this part of our dissertation. The insights gathered using our macroscopic model can, nevertheless, be used for ad hoc networks because they allow for more realistic parameterization of the microscopic models. Looking from this perspective, we reviewed the following related work.

3.3.1 Microscopic Mobility Models

The class of random models is the most prominent example for microscopic models. We start with random models to describe the movements of individuals: there are numerous ones including the *random walk model*, the *random gauss-markov model*, the *random mobility model*, the *markovian model*, and the *random waypoint model* [CBD02]. These models try to provide an exact analytical description of the behavior of individuals. Random models have been optimized for accuracy, [Bet01a] being one example. The characteristics of a generalized random model for cellular networks with respect to mobility patterns are studied by Zonoozi and Dassanayake [ZD97]. However, these models are not able to reflect macroscopic behavior and often exhibit undesirable properties if applied improperly as shown by Yoon et al. [YLN03].

Random models may operate on other levels to reflect group behavior, as described by Hong, Gerla et al. [HGPC99]. However, neither the results of individual mobility nor of group mobility do account for a city at large. This is mainly due to the mostly homogeneous nature of random models. If aggregated, these models lead to fairly equal user distributions for large areas. If we, however, regard a real city, we observe numerous attractions and hot spots. These are likely to be crowded during short periods of time or over day while being empty during the night. It is, from our perspective, extremely unlikely that pure random behavior fits such a scenario. There are some random models, which can be widely parameterized to match correct macroscopic behavior as well (see, for example Bettstetter et al. in [BHPC02]). The correct instantiation for microscopic models depends on macroscopic models, nonetheless. Since the microscopic models violate our assumption of realism at metropolitan scale, we do not further investigate this class of models in this section.

3.3.2 Macroscopic Mobility Models

There are various sub-classes of macroscopic mobility models. For our survey we start with models developed in the context of *transportation planning*, because these serve as foundation for various other models reviewed. Thereafter, we present *traditional teletraffic models* starting with general work and moving towards models more specifically targeted to describe metropolitan areas. Finally, we describe *measurement/observation based models* describing real network behavior.

Models from transportation planning are often used as a basis for large scale mobility models. TRANSIMS [LAN04] is an example of a complex synthetical model. It creates a virtual region with complete representation of individuals and their activities. The transportation is modeled and trips are planned to satisfy the individuals' activity patterns. The behavior of individuals and the geometrical distribution of households is derived from census data. The focus of TRANSIMS is transportation planning. It would be possible to extend the framework to generate data that fits the needs of teletraffic modeling for wireless networks. However, this comes at the expense of a very complex overall framework, which, for example, tracks the movements of all individuals and requires very costly calculations and extensive parameterization.

The work of Hong and Rappaport [HR86] describes a classical model in the area of teletraffic management. Here, the call blocking probability as well as the call termination probability are studied for varying cell sizes and different priority schemes in the handover process. A general work in the area of teletraffic models for urban environments was performed by Nanda [Nan93]. The author investigates various types of cell structures and especially the influence of variation in the size of cells. Findings of [Nan93] include the relation between handoff rate and cell size as well as between handoff rate and cell shapes. Despite the results in the areas investigated, the work cannot be applied to our scenario, since it does not provide the necessary detail and flexibility of the user model. Other work in this area includes [CST96] by Camarda et al. and [JB99] by Jugl and Boche, both of which suffer from the same shortcomings as [HR86] and [Nan93].

The work of Scourias and Kunz [SK99] limits its scope on the location management aspects of user mobility. However, the underlying mobility model presented makes uses of more detailed information of user behavior and activities, which are not further exploited. Moreover, the work borrows the concept of trips to describe the intention and whereabouts of users from work in the area of transportation modeling. The work of Rocha et al. [RMdS00] uses a mobility model similar to the one in [SK99] and implements a graphical tool to support analysis of the results. The instantiation of areas and user behaviors is, however, very coarse and does not reflect realistic environments.

Markoulidakis et al. [MLA98] describe a traffic model for “third generation cellular mobile telecommunication systems”. Part of this model is a user mobility model [MLTS97], which is based on the average distance between and the average velocity of users. These parameters are estimated for certain environments, including multiple classes of outdoor environments and different vehicles. The work is limited to the estimation of cell border crosses to predict handoff rates and call durations. While in [MLA98] the geographical instantiation assumes a manhattan grid model of a city center using relatively coarse characteristics for zoning information, [MLTS97] presents background information on mobility modeling and a comparison to classic area zone models for teletraffic modeling including work described in [HR86].

The work of Lam, Cox, and Wilson [LCW97] presents and evaluates teletraffic models for metropolitan, national, and worldwide scale. The work is based on results from transportation planning. The authors have been able to show that the model predictions can be validated using real world data. The focus lies—like in most work in the area of wireless teletraffic modeling—on the investigation of control plane issues such as location management traffic. Therefore, the model does not include a distinction into multiple user types, effectively rendering it unusable for most parts of our intended model usage, which couples data traffic with user classes.

After finishing our work on the model, Breyer et al. [BKOKR04] proposed an activity-based model to describe user mobility, user behavior, and the related network service usage. The model is very similar to our model, however, it focuses on much smaller areas, for example, a campus of a university.

Another group of models for large scale networks relies on the analysis of existing infrastructures and network traces. In particular, work investigating large scale wireless radio access topologies for data traffic with respect to user mobility has been proposed only recently. The work of Tang and Baker [TB02] is able to provide deep insights on user behavior for a metropolitan area wireless network. The work of Kotz and Essien [KE02] claims to be the largest and most comprehensive real world trace of a production wireless LAN. The results, however, do account for a special campus style network and mainly focus on traffic analysis—the mobility aspect is restricted by the campus setup and cannot be transferred to public networks. Balachandran et al. [BVBR02] concentrate on network performance of small-scale networks, which are not representative for the metropolitan scale.

There are two critical points about trace-based methodologies. First, they only account for services already deployed and in use, particularly only counting "early adopters". This is especially true if we regard wireless radio access networks for data communication at the time of writing. Second, the results often cannot be separated into mobility and traffic related parts, thus, prohibiting the parametrization of individual factors for simulation.

3.3.3 Conclusion

There exists numerous work in the area of mobility and workload modeling for communication networks supporting mobile users. Current approaches including [Nan93] and [LCW97] focus on traditional teletraffic management properties, that is, parameters coupled with handoff and handover rates of voice calls induced by user mobility. They do not differentiate between different classes of users or Internet like traffic demands, however. These restrictions are mainly due to the state-of-the-art of traditional cellular networks, which at that time were mainly used for connection-oriented voice communication. The activity-based mobility models described in [SK99], [RMdS00], and [MLTS97] provide for some basic mechanisms that may be used for our purpose. The formulation of these models and their instantiation is, however, not optimized for data traffic analysis but for the limiting case of classical teletraffic applications likewise [Nan93] and [LCW97].

In summary, random models do not properly account for mobility within large areas and the existing macroscopic models are either too restricted for our purpose or too complex for the intended application. The empirical studies of wireless local area networks are limited with respect to the strict decoupling of mobility and traffic. Thus, we decided to develop a novel model to fit the needs of the described scenario. We borrow concepts from transportation planning to realistically cover mobility in macroscopic areas, such as a center of a large city. The traffic part of the model describes the application related traffic characteristics and allows for flexible traffic generation on application, stream, flow, or packet level. See [HPSS03] for realistic traffic modeling on these levels.

3.4 Mobility and Workload Model for Metropolitan Areas

The mobility part of our model is developed in the context of transportation and land use modeling [Opp95]. We use the well-known *travel demand modeling* or *activity based approach* described in [Kit88], which is also used in related work including [SK99] [RMdS00] [MLTS97]. The basic elements of activity based transport modeling are the *trip*, which defines the movement of a user from an origin to a destination, and the *zone*, which defines areas with a certain attraction level. Trips are based on the intended *behavior* of users while zones represent homogeneous areas with respect to socio-economic characteristics. The size of zones can range from a few hundred square meters to several square kilometers in size. This basic model has been used since the 1960s to forecast travel demand and various techniques and sub-models have been developed for determining the variables for each of the elements of travel demand modeling. It is important to keep in mind that the developed model represents aggregated information of users and zones only. Thus, the model is well-suited for investigations covering macroscopic effects. In the following we describe the fundamentals of the activity based transport modeling.

3.4.1 Fundamentals

The fundamental concept of our model is the classification of different types of location and user behavior. Based on this classification, we are able to generate trips and calculate the density of users within the areas of interest. To form the workload, a traffic model is adjoined to our mobility model.

Classification of Location

The classification of locations has to reflect their different attraction levels over time. A reasonable granularity can be reached by assigning a base attraction level to a zone. The base attraction level is set depending on basic characteristics of the location. *Residential*

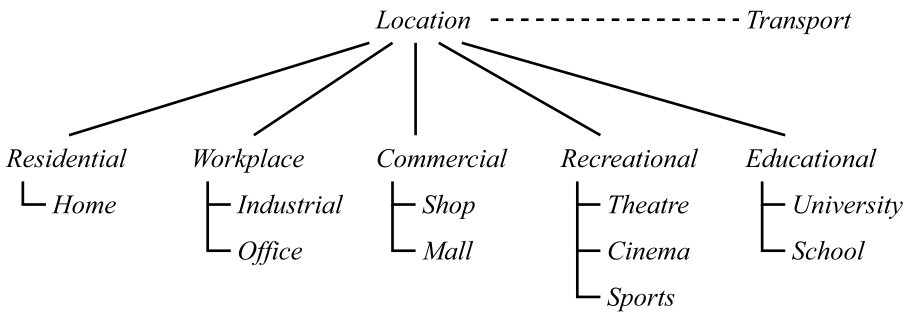


Figure 10: Classification of location.

areas and *workplaces* (we distinguish into industrial and office workplaces) account for the major fluctuation of users. Moreover, the attraction of *commercial*, *educational*, and *recreational* facilities is an important factor. These types of locations are further refined as follows. Commercial locations include shops and shopping malls, recreational locations include theatres and opera, museums, cinemas, sport events, bars, and pubs. Educational places can be further classified into schools and universities. A special (virtual) location, called *transport,* accounts for users on the move, for example, using cars, bikes, busses, or trams. We assume zones of reasonable size, like those given in most zoning plans for city development. Moreover, the zones should be homogeneous with respect to the attraction of the different locations within each zone, which is usually also true for public zoning information. See Figure 10 for a detailed classification of locations. Please note that the classification is non-exhaustive to keep the model's level of complexity reasonable.

The accuracy of the approach may be further increased if extra information about special places such as shopping malls, schools, universities, and sights possibly located in the zone is combined with the base attraction.

Classification of User Behavior

The most important criterion for the user classification is to adequately characterize the behavior of individuals. The intentions of the resulting groups need to match the attracting locations as well. The availability of information about the constitution and behavior of the derived groups is another crucial factor to allow for proper instantiation. This includes information to allow for case discrimination, in particular, to predict state transition or inactivity of users. Resulting from this information, we are able to predict the aggregated effect of user mobility.

We differentiate the following types of users (roles): *resident, worker, consumer, trainee,* and *traveler* (see Figure 11). Residents are the main group of interest during evening and at night. We distinguish between inactive residents (sleeping at home) and active ones (being active at home). We discriminate consumers into buyers and visitors in order to separate shopping activity from leisure. The latter ones have to be distin-

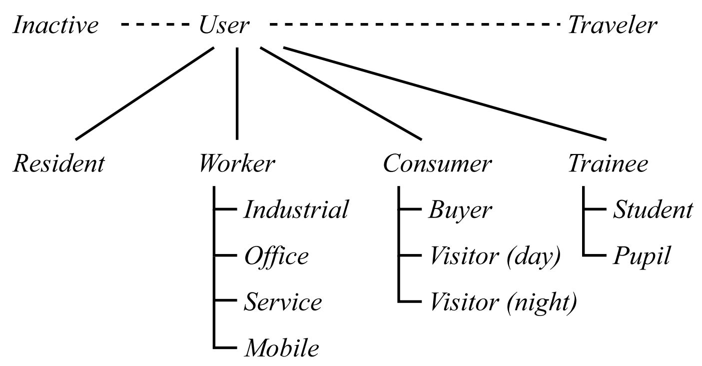


Figure 11: Classification of user behavior.

guished into daily and nightly visitors to model the different locations of recreation facilities as for example museums, swim-halls, and shops during day or cinemas, theatres, and pubs in the evening. The number of daytime visitors includes tourists as well.

The number of trainees, namely pupils and students, usually cannot be neglected, nor can the number of workers. The class of workers is divided into different sub-classes, which account for different zones in which they perform their job (commercial zone, industrial zone, etc.). Moreover, we differ with respect to the expected professional communication behavior such as, for example, mobile personnel to rely on the communication network to carry out their job, or office/industrial workers to communicate in breaks or on their way to/from work. The special role of a traveler accounts for the state transition between other states. For example, if a resident departs for a shopping district, he becomes a traveler before finally taking the role of a consumer. Besides their function as intermediate state, the travelers represent the number of people just crossing the investigated area by car, commuting to external work places or from external housings to internal workplaces/universities. See Figure 11 for the classification of user-behavior.

Trip Generation

The classification of location and user behavior are the basis for the generation of trips and the calculation of the effects induced by mobility. The trip generation estimates the total number of trips that depart and arrive in a specific zone. In particular, trips are classified by trip purpose, such as work or shopping, for example. The purpose is derived from the intended user behavior. The corresponding trip is scheduled to a location where this intention can be fulfilled, for the given examples to a workplace or shopping mall. The generated trips need to be distributed to the destination zones according to the attraction of these zones, which is determined by socio-economic characteristics as described using the classification of locations. Moreover, a modal split accounts for different transportation modes such as car, bike, bus, or tram. According to a trip assignment the individual trips are mapped to the transportation infrastructure and the best route is calculated. We are primarily interested in the user densities for given zones but not trips of each individual. This allows us to omit the determination of modal splits, trip assignment, and route calculation, which reduces the model complexity significantly. To allow for the remaining trip generation and distribution, the classification of location and user behavior is sufficient. Upon instantiation, real world data needs to be fitted to match this classification scheme. Finally, we are able to calculate the user densities for given zones.

3.4.2 Model

Our model distinguishes between user mobility and user traffic related issues to allow for flexible instantiation. We use the classification of locations, given in the granularity of zones, and the classification of users and behaviors (see Section 3.4.1). Moreover, we assume an overlay of cells onto the area of investigation. These may express individual radio cells or the aggregation of multiple radio cells, depending on the network model.

Modeling User Mobility

We formulate our mobility model building on the above introduced preconditions. Let

$L = \{residential,\ workplace,\ commercial,\ recreational,\ educational,\ transport\}$ denote the set of locations. The elements of L are denoted as l.

$B = \{resident,\ worker,\ consumer,\ trainee,\ traveler,\ inactive\}$ denotes the set of user roles (behavior). The elements of B are denoted as b. Exactly one behavior is assigned to a user at each point in time.

$Z = \{z_1, z_2, \ldots, z_n\}$ denotes the set of zones. Zones do not overlap in space. The union of all zones, z_i for all $i = 1, 2, \ldots, n$, covers the whole area of interest.

To allow for a mapping of locations to a cell-based network infrastructure let

$C = \{c_1, c_2, \ldots, c_m\}$ denote a set of cells.

The model should predict the number of users with behavior b, which can be expected in a zone/cell at a given time t. Let

$U_b(t)$ denote the total number of users with behavior b at time t and

$U_{z,b}(t)$ denote the time-dependent number of users with behavior b within zone z.

The total number of users is given by

$$U(t) = \sum_{b \in B} U_b(t) \text{ with } U_b(t) = \sum_{z \in Z} U_{z,b}(t). \qquad (1)$$

Hereby, $U(t)$ accounts for the total number of users being located within the investigated area at time t. The total number of active users might vary over time. During night most users are inactive, for example. They change their role to active residents after getting up and eventually to a worker, consumer, or trainee. In between, that is, while being mobile, they are modeled as travelers. To account for the *fraction of users* being active in a role b, we define a time-dependent split factor per behavior $f_b^{\ u}(t)$. In reality, moreover, the maximum number of users in the system is variable because commuters and travelers may enter and leave the system over day. Please note that $f_b^{\ u}(t)$ is independent from the location.

Using $f_b^{\ u}(t)$ with $\sum_{b \in B} f_b^{\ u}(t) = 1$ we can calculate $U_b(t) = U(t) \cdot f_b^{\ u}(t)$. (2)

The user population $U(t)$ needs to be assigned corresponding to the *attraction* of locations within zones. We use a zone- and behavior-dependent split factor $f_b^{\ a}(z)$ to account for this property. With $f_b^{\ a}(z)$ we are able to calculate the number of users $U_{z,b}(t)$ within each zone.

$$U_{z,b}(t) = U_b(t) \cdot f_b^{\ a}(z). \qquad (3)$$

To be able to account for communicating vs. non-communicating users, let

$\tilde{U}_b(t)$ denote the number of communicating (active) users with behavior b at time t (they emit traffic according to their type of behavior). Measured over a time-interval $[t_1, t_2]$ $\tilde{U}_b$ gives us the duration of communication activity in user hours. Let

$\tilde{U}_{z,b}(t)$ denote the number of active users with behavior b in zone z at time t.

To convert the number of users located in a zone into the corresponding number of active users, we introduce the notion of *intensity*. The intensity models the fraction of time a user dedicates to communication purposes. We introduce a factor $f_b^{\,i}$ to account for the communication intensity of users of class b. Thus, $\tilde{U}_b(t)$ denotes the sustained number of communicating users. For the sake of simplicity, $f_b^{\,i}$ may be combined with the traffic part of a workload model and neglected in the mobility part. For maximum flexibility of the mobility model, it would also be possible to introduce a zone and time dependent version of $f_b^{\,i}$, $f_b^{\,i}(z, t)$. Building on Equations (2) and (3) we obtain

$$\tilde{U}_{z,b}(t) = U_{z,b}(t) \cdot f_b^{\,i} = U_b(t) \cdot f_b^{\,a}(z) \cdot f_b^{\,i} \text{ with } \sum_{z \in Z} f_b^{\,a}(z) = 1 . \tag{4}$$

We obtain $$\tilde{U}_b(t) = \sum_{z \in Z} \tilde{U}_{z,b}(t) = \sum_{z \in Z} U_{z,b}(t) \cdot f_b^{\,i} . \tag{5}$$

Moreover, $$\tilde{U}(t) = \sum_{b \in B} \tilde{U}_b(t) . \tag{6}$$

The next step is the transformation of the results from zone to cell level. We use a network model similar to [24]. The cell shape is approximated as hexagon. We assume that p cells are arranged in $(2r-1)$ columns. Figure 12 shows that we obtain r columns of k cells and $(r-1)$ columns of $(k-1)$ cells. The total number of cells can be calculated as $p = rk + (r-1)(k-1)$. This cellular structure is overlaid on the zone-based area of investigation. To be able to transform the results from zone to cell granularity, let

A_z be the size of zone z in square meters and

A_c be the size of cell c in square meters. Let

$r_{z,c}$ denote the fraction of zone z, which is covered by cell c.

Assuming an equal distribution of attraction levels within zones, we are able to transform the results using

$$\tilde{U}_{c,b}(t) = \sum_{z \in Z} \frac{\tilde{U}_{z,b}(t)}{A_z} \cdot A_c \cdot r_{z,c} \text{, with } \sum_{z \in Z} r_{z,c} = 1 . \tag{7}$$

Using Equation (4) and Equation (7) we can calculate the user activity matrix $\tilde{U}_{c,b}(t)$.

$$\tilde{U}_{c,b}(t) = \begin{bmatrix} \tilde{U}_{1,resident}(t) & \tilde{U}_{1,worker}(t) & \dots & \tilde{U}_{1,inactive}(t) \\ \tilde{U}_{2,resident}(t) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \tilde{U}_{p,resident}(t) & \dots & \dots & \tilde{U}_{p,inactive}(t) \end{bmatrix} \quad (8)$$

Please note that the model described above can easily be extended to reflect even more details of user behavior or location attributes. It would be possible, for example, to model the time-varying aspects of user activity for different zones by introducing ${f_b}^a(z, t)$. Another extension we do not describe is the adaptation of ${f_b}^u(t)$ to account for additional influences as, for example, workers being on sick leave, etc.

The use of discrete time-intervals simplifies the instantiation of our model for practical purposes. The average value of active users per cell for a given time-interval can be combined with a synthetic traffic forecast for these users to form the workload for a cellular network. In particular, we obtain the user activity of all users of behavior b in a cell c for the interval $[t_1, t_2]$ to be

$$\bar{U}_{c,b} = \tilde{U}_{c,b}(t)\Big|_{t_1}^{t_2} \text{ given in user hours.} \quad (9)$$

Using Equation (9) we can also compute the user density per cell by dividing $\bar{U}_{c,b}$ by the cell size A_c. The overall user activity in user hours for behavior b for the interval $[t_1, t_2]$ is given by

$$\bar{U}_b = \sum_{c \in C} \tilde{U}_{c,b}(t)\Big|_{t_1}^{t_2}. \quad (10)$$

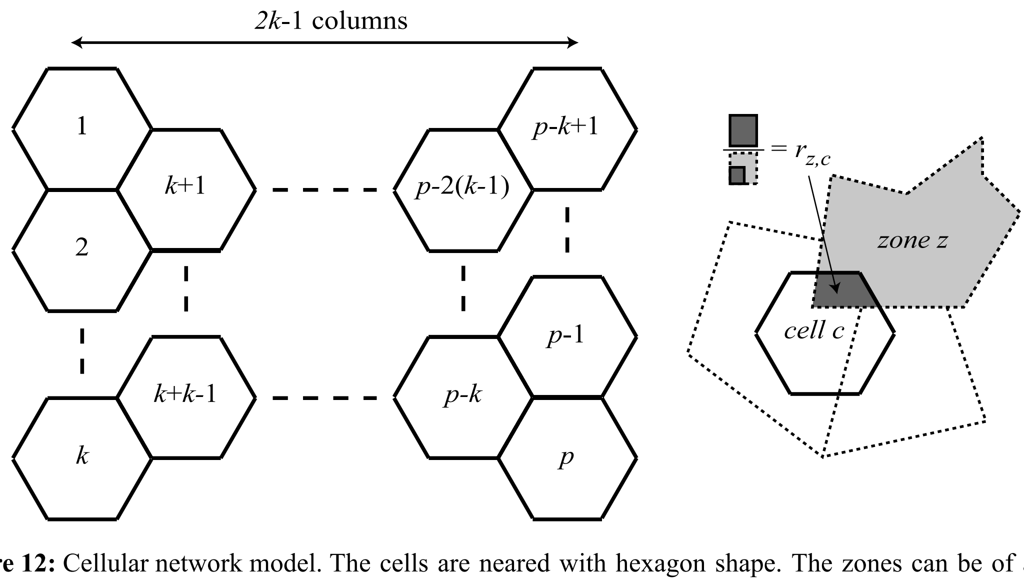


Figure 12: Cellular network model. The cells are neared with hexagon shape. The zones can be of arbitrary shape.

Modeling User Traffic

Likewise the mobility model, the traffic model is designed to allow for flexible parameterization. Various traffic classes are provided and may be instantiated independently for each user role. Let

$M = \{m_1, m_2, \ldots, m_g\}$ denote a set of g traffic classes. The elements of M are in general denoted as m. Let further

$Rate_{m,b}(t)$ be the time dependent traffic rate of class m for one user of behavior b in kbyte/s.

We are able to augment traffic estimates for different classes of users to form a workload matrix, $W_{m,c}(t)$, for each traffic class m. The workload matrix describes the traffic that is generated in a cell c to be

$$W_{c,m}(t) = \sum_{b \in B} \tilde{U}_{c,b}(t) \cdot Rate_{m,b}(t). \tag{11}$$

Using Equation (11) we are able to calculate the daily workload as well as the time varying loads for each traffic class m. The resulting traffic accounts for the traffic generated per cell. The distribution of the destinations of this traffic can be performed as appropriate.

3.4.3 Instantiation of the Model

In this section, we describe the instantiation of our model for the case of Darmstadt, a German city of around 140,000 inhabitants. The area covered is approximately 9.13 square kilometers and accounts for ~54,000 inhabitants. We present a typical 24 hour day within our model. The individual steps of the modeling process are:

1. Classification of users and user behavior B.
2. Classification of zones Z, cells C, and locations L.
3. Calculation of the time-dependent number of users $U_{z,b}(t)$ with behavior b within zone z and of the time-dependent activity $\tilde{U}_{z,b}(t)$ of users with behavior b in z.
4. Transformation of the results from zone to cell level.
5. Classification of traffic classes M per user behavior b.
6. Calculation of the workload matrix $W_{c,m}(t)$ per cell c and traffic class m.

We performed the instantiation of the model using statistical field data. The first two steps follow our description in Section 3.4.1 closely and build mainly on public and private census data, which has been obtained from the census department of the city of Darmstadt. Figure 13 shows a map of the modeled part of Darmstadt, including cells and zones. While the granularity of available data suited our model nicely in most parts, significant post-processing overhead as well as intimate knowledge of Darmstadt was necessary to increase the precision of the model. Using our model equations from Section 3.4, we calculated the time- and location-dependent activity of users. To obtain a valid workload model, we combined the activities with traffic estimates per user class. To allow for proper treatment of QoS aspects while keeping the complexity manageable we introduce four traffic classes similar to the ones proposed in [3GPP03]:

- *Conversational traffic* denotes traffic stemming from IP-telephony or video-conferencing. We assume constant bit rate traffic with a low delay bound.
- *Streaming traffic* accounts for services such as video on-demand or streaming audio. We assume variable bit rate traffic with low delay bound.
- *Interactive traffic* models transactional traffic such as web traffic. The delay bound is moderate to low and the traffic is self similar in nature.
- *Background traffic* accounts for traffic such as email delivery or peer-to-peer traffic without fixed delay bounds.

The traffic estimates used for instantiation reflect a prediction of user traffic in next generation mobile networks and have been provided by Siemens Corporate Technology as part of a collaboration project. Please note that we model the traffic aggregates on cell level. It is possible to combine the mobility model with other traffic vectors to account for different scenarios, easily. A detailed description of the instantiation process for both, the mobility part and the traffic part of our model as well as some implementation details can be found in [Kro02].

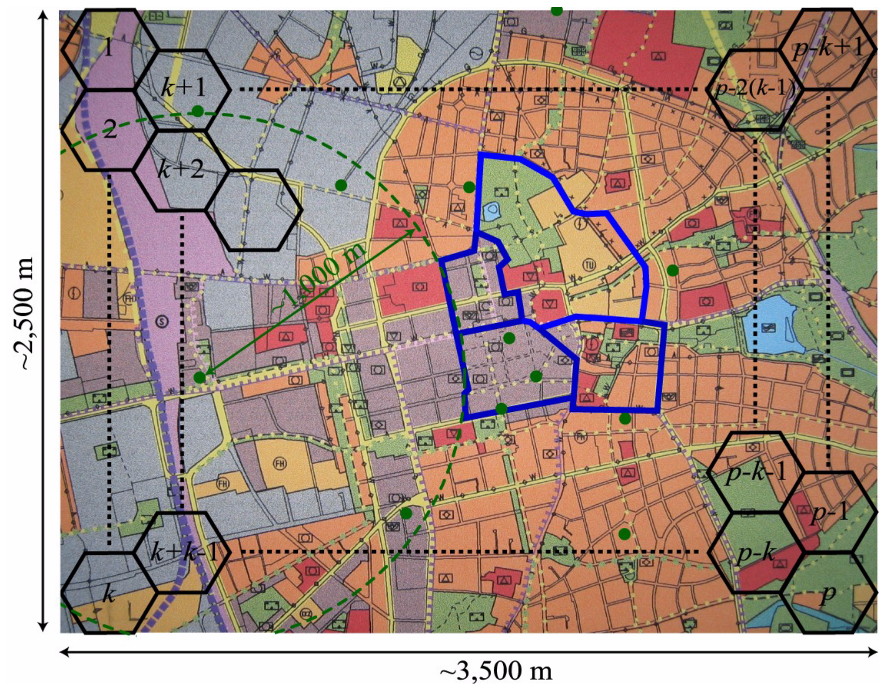


Figure 13: Zoning plan of the city center of Darmstadt. The zoning plan depicts the different urban areas. The arbitrary shape of some zones from census data is drawn in blue alongside with the hexagon-shaped black cells of our model. Base stations of actually deployed T-Mobile GSM network are show as green points, the dashed green line is a radius of 1,000 m to give an estimate for the GSM radio range.

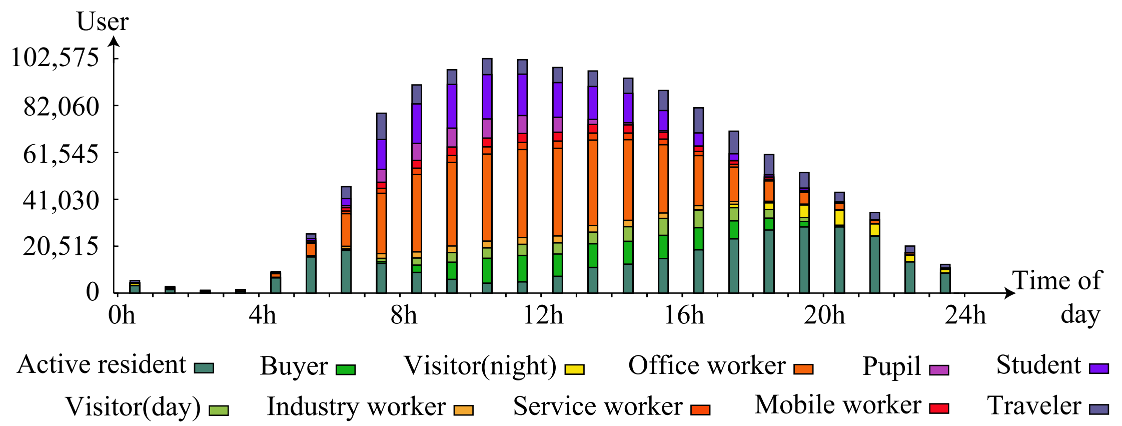


Figure 14: Number of active users over day in intervals of 1 h. See Section 3.4.1 for the classification of user behavior.

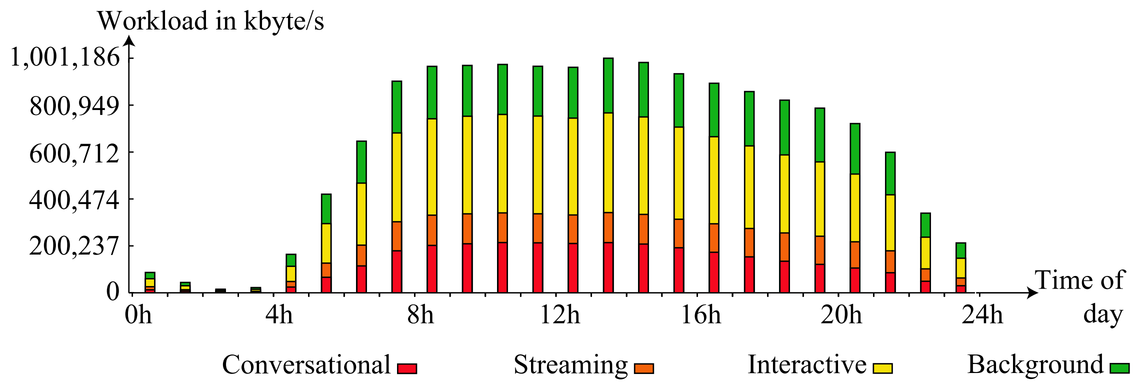


Figure 15: Predicted workload (in kbyte/s) over day in intervals of 1 h. See Section 3.4.3 for the modeled traffic classes.

Results of the model include the summarized number of active users over the time of day as shown in Figure 14 and the summarized workload over the time of day as presented in Figure 15. The numbers obtained need to be interpreted carefully. Despite the fact that we have a bounded simulation area, Darmstadt attracts a lot of commuters, consumers, and trainees from outside the modeled area. This is reflected if we regard the absolute numbers of inhabitants (~54,000) with the maximum number of active users being in the simulation area during the busy hour (~102,500). It is important to keep in mind that the geographical distribution of user density changes significantly over day. This is due to the different activities pursued by the different classes of users, which cannot be seen in Figure 14 or Figure 15. A detailed analysis of the model predictions is presented in Section 3.5 below.

3.5 Analysis of Model Predictions

The model predictions can be separated into predictions of the mobility related part of the model and those of the workload part of the model. Here, we concentrate on the mobility related results since the instantiation of the traffic related part is heavily dependent on the proposed application mix and may be performed according to the individual scenario in question. Since the instantiation of the model was performed for the case of Darmstadt, our findings may not be generalized easily. We believe, however, that the pathological cases observed manifest some typical effects of mobility, which can be expected to be of importance for other networks, too.

3.5.1 Model Predictions

Our model predictions include the estimated user density and user activity over time. We are especially interested in density/activity fluctuations since the emphasis of our study lies on the effects induced by mobility. Our observation is that due to the different activities pursued by the different users, the geographical distribution of user density changes significantly over day. The visualization of the user activity over day in Figure 14 shows, starting on the early morning, a steady increase of active users as they get up and go to work, school, or other activities. During day, most active users are away from home. Noticeable, the decrease in the evening is not as abrupt as the increase seen in the morning. Here, the different user groups contribute to the number of active users differently: pupils, for example, are expected to return home or pursue other activities during early afternoon while workers are expected to work until early evening. As a result, we see a change back in heavily populated zones to be residential areas. The combination of the user activity with the estimated traffic is expressed in the summarized workload, which is illustrated in Figure 15 above. The peak traffic value does not correspond directly to the peak value of user activity, because the traffic produced from users in different classes is not homogeneous but differs substantially.

As noted earlier, our user and traffic distribution is dependent on the geographical location. Hence, we move from the summarized user density over day to the geographical density distribution to get a clearer picture of mobility induced fluctuations. Figure 16 presents the corresponding density chart of user activity at cell level. The depicted values account for the projected aggregated user activity $\overline{U}_{c,\,active}$ for all active users from 12:00 h to 13:00 h. The activity is given in user-hours. The city center of Darmstadt, which is also located in the center of the investigated area, consists of office and shopping facilities and large parts of the university while the region in the east is mainly covered by industrial areas. It is clearly visible that the city center and the nearby university attract most of the users.

Figure 17 shows the density of active residents $\overline{U}_{c,\,resident}$ for the time period from 20:00 h to 21:00 h. Because most workplaces and shops are closed at this time, we expect most of the residents to be at home. For this period, the main residential areas are clearly visible in a belt surrounding the center of the city and in the west of the city using our model. The center itself is only populated with few residents at the time of the presented snapshot. The visible density differences in Figure 16 and Figure 17 also illustrates the phenomenon observed earlier: the shift in user densities from the city center towards residential areas in the evening. Please note that commuters living outside the investigation area contribute to the active users but not to the active residents.

We are able to obtain additional insights on different granularities because of the clear separation of user classes. While the consumers (see Figure 17) are concentrated in the shopping areas of the city center, the office workers (see Figure 19) spread over the collocated office buildings in the same area and the district south east of the center. The industry workers, which are not represented in a figure, show a different distribution: they mostly populate the area in the north east of the model.

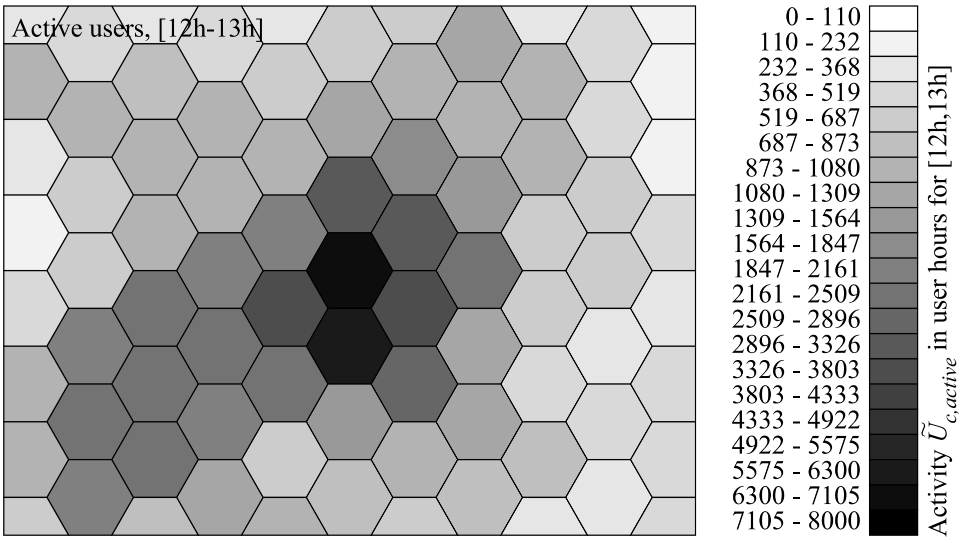


Figure 16: Activity of all active users from 12:00 h - 13:00 h. The active users are the super-group including all other user groups except the inactive ones. The activity level $\overline{U}_{c,\,active}$ is given in user hours.

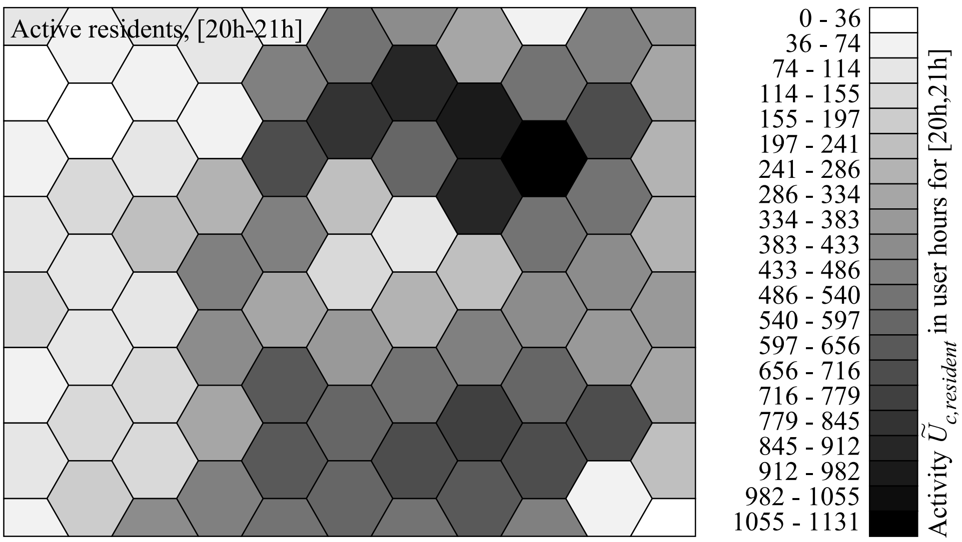


Figure 17: Activity of active residents from 20:00 h - 21:00 h. The activity level $\overline{U}_{c,\,resident}$ is given in user hours.

In contrast to consumers and workers the image for the trainees (see Figure 20) shows a distribution that has two major hot spots: Darmstadt University of Technology, which is located in the city center and Darmstadt University of Applied Science located south east account for most activity. The schools and so the pupils are distributed more homogeneously over the area modeled. They account for the activity besides the university hot spots in Figure 20. In addition to the static results presented at this point, animated visualizations of the resulting user and traffic fluctuations over time can be found at [HSHS04] for the Darmstadt scenario.

From the geographical distribution of activities $\overline{U}_{c,\,b}$, we now move to the absolute number in users $U_b(t)$. Figures 21 to 24 depict the user count of the four main user groups over the modeled day. The number of *residents* is separated into active and inactive residents and shown in Figure 21. During the morning, the inactive (sleeping) residents get up and change their role to active. This can be seen starting from 4:00 h in the morning, the number of *active residents* increases as people get ready for work. Users are expected to be in the state of active residents only shortly until they leave home. From 8:00 h on most people are up but mostly account for activity in other roles than resident. During day, depending on their business, users may re-appear for a short period of time in the role of an active resident before changing their role again. Starting from early afternoon, the number of active residents grows again when people get home and decreases when they go to bed.

The daily distribution of the other investigated behaviors differs in shape as follows: consumers and visitors are expected not to be the early birds (see Figure 22). They appear corresponding to the opening hours of attractions and shopping centers, which are mostly between 8:00 h and 10:00 h. In the evening we see nightly visitors, which perform recreational tasks. Trainees are active earlier compared to consumers but do not

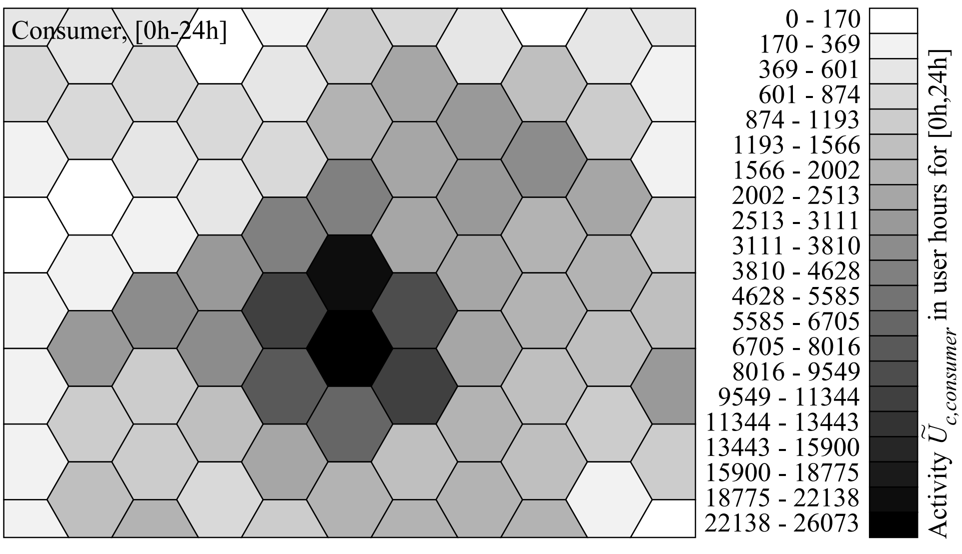


Figure 18: Summarized activity of consumers over 24 h. The activity level $\overline{U}_{c,\,consumer}$ is given in user hours.

stay active as long. We see a significant difference between students and pupils, the latter ones are expected to be only active for much shorter time because of the German education system with school-hours focusing on mornings mainly (see Figure 23). The distribution of workers looks nearly similar for each class of worker. Here, the main difference is reflected in the absolute number of workers per class. Darmstadt traditionally hosts a lot of public and governmental offices. Moreover, in the city center, there is a concentration of office space and the administrative staff of both universities. Hence, we see a clear domination of the office workers in Figure 24.

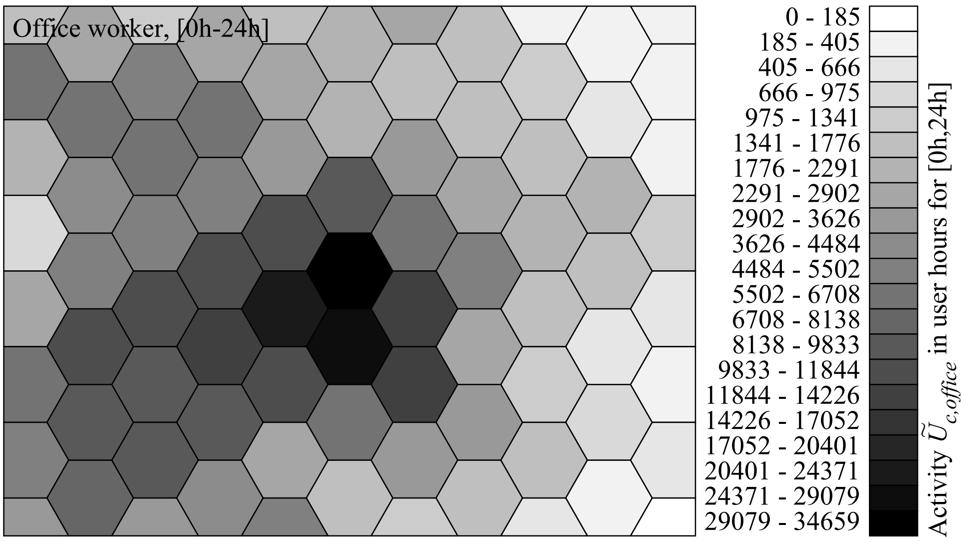


Figure 19: Summarized activity of office workers over 24 h. The activity level $\overline{U}_{c,\,office}$ is given in user hours.

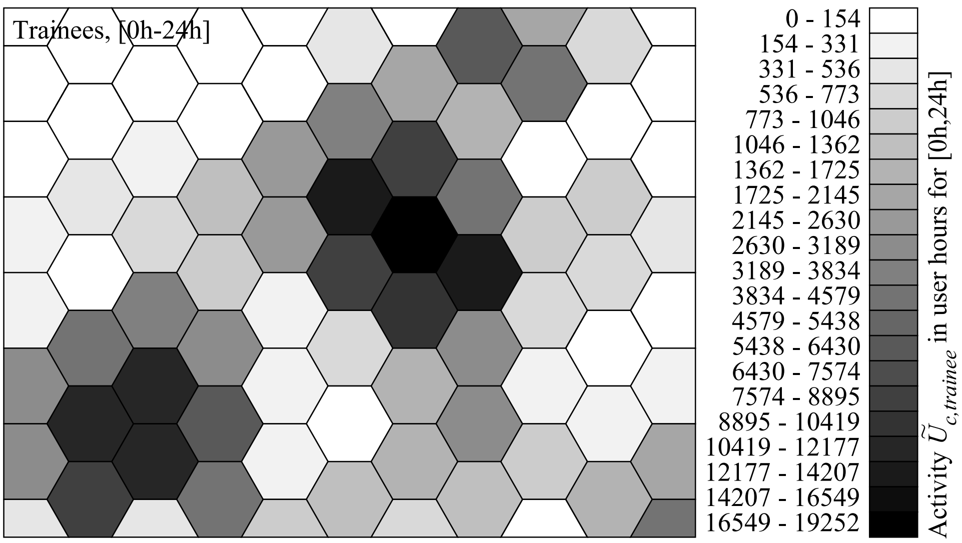


Figure 20: Summarized activity of trainees over 24 h. The activity level $\overline{U}_{c,\,trainee}$ is given in user hours.

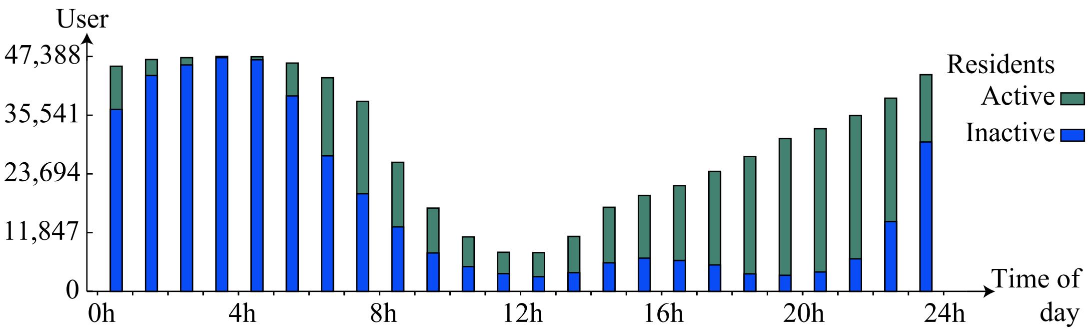


Figure 21: Number of active/inactive residents over 24 hours. The number of users, $U_b(t)$, is given in number of individuals.

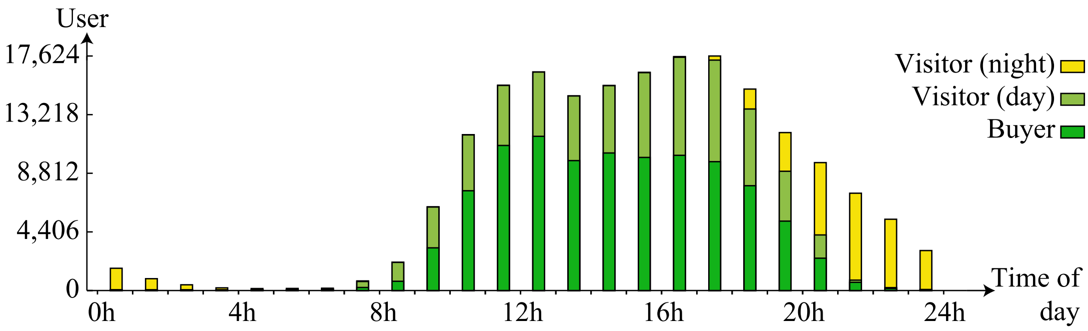


Figure 22: Number of consumers over 24 hours. Consumers are daily and nightly visitors and buyers. We count individual users.

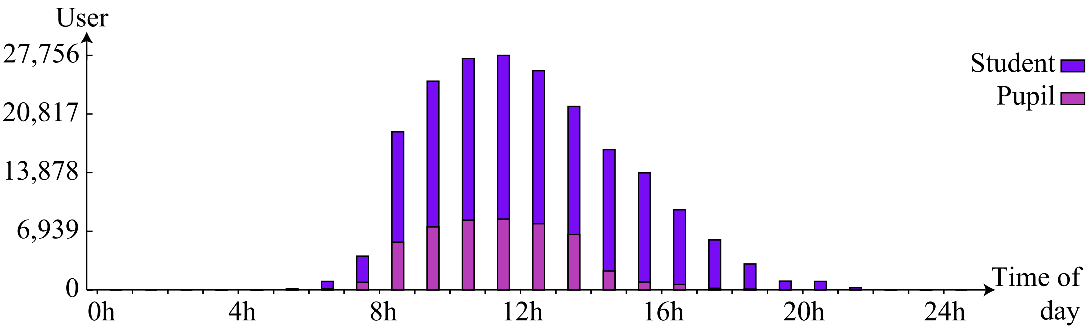


Figure 23: Number of trainees over 24 hours. Trainees are students and pupils. We count individual users.

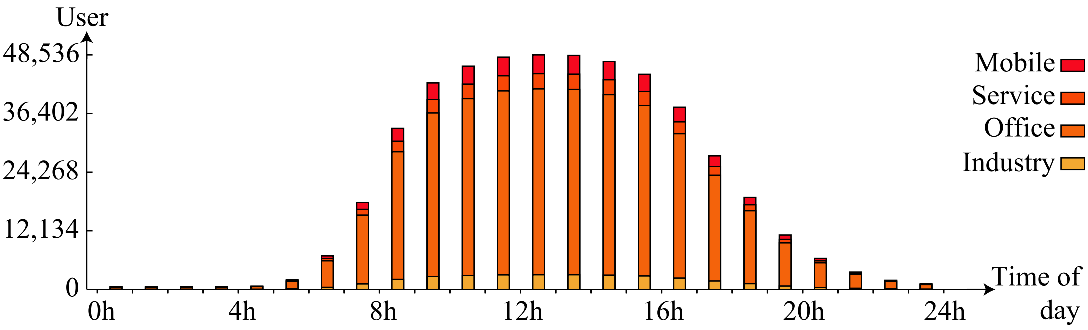


Figure 24: Number of workers over 24 hours. Workers are mobile, service, office, and industry workers. We count individual users.

3.5.2 Evaluation of Model Predictions

Our analysis of the model predictions concentrates mainly on two aspects: first, the overall geographical distribution of users and traffic and, second, the fluctuations of number of users and traffic load over time. Since there exist no empirical models to directly compare our results with, we use existing synthetic state-of-the-art models as yardstick to analyze our model predictions. Our results differ significantly in terms of user and traffic distribution throughout the network compared with the prominent synthetic random walk and random waypoint model.

The work of Yoon et al. [YLN03] as well as the work of Bettstetter and Wagner [BW02] analyze the stationary properties of the random waypoint model [JM96], which can be summarized as follows: the distribution of the location of nodes within a random waypoint model is concentrated near the center of the modeled area. This is because the nodes traveling between uniformly chosen points spend more time near the center than near the edges (see the given references for the closed form and approximations of the geometrical node distribution). The random walk model on the other hand converges to a nearly uniformly distribution of location of nodes in the stationary case, which is insufficient for the intended macroscopic usage, too [Rev90]. Our model does not show a uniform distribution of users and traffic, though. Figure 25 shows the fraction of active users over the fraction of area during the busy hour. Within approximately 20% of the area we expect ~50% of the active users, which cause ~44% of the total network traffic. This result clearly illustrates the presence of hot spots in the modeled area. Moreover, within approximately 50% of the area, we expect ~78% of the active users causing ~76% of the total traffic. In other words, the remaining 50% of the covered area have only to deal with ~24% of the traffic. Also, the calculated rate for the estimated workload varies substantially using our model instantiation of Darmstadt. While the average rate over all cells is approximately 14.9 Mbyte/s during the busy hour, there are cells with a rate as high as ~62.0 Mbyte/s and other cells with a rate of only ~2.0 Mbyte/s. The standard deviation over the 83 cells being 11.00 Mbyte/s.

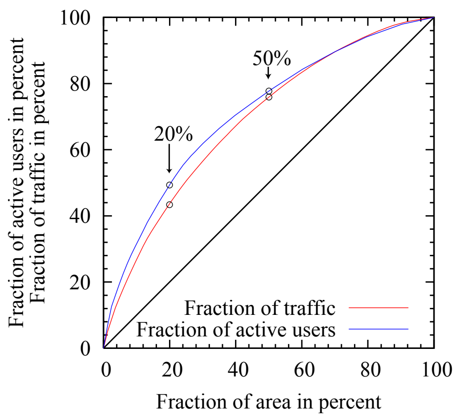


Figure 25: Fraction of active users/generated traffic vs. fraction of area. We show the busy hour of the network. A few cells contribute very much to the overall user activity and traffic generation.

We analyzed the dynamics of user and load fluctuation between cells over day. Our observation was an oscillation of users between residential and non-residential areas for the case of Darmstadt. For the investigation of the dynamic behavior of the instantiated mobility model we measured the coefficient of variance (CoV) of the traffic load over locations (at fixed times) and over time (at fixed locations). The CoV normalizes the standard deviation to the average; the higher the CoV, the higher the fluctuations of load. If we compare the traffic fluctuations over day for the individual cells, we obtain CoVs from 0.53 up to 1.02, the average CoV being 0.63. This shows clearly that all cells have to deal with high dynamic load fluctuations over day while some cells experience extreme fluctuations. To illustrate the CoV measure: the arbitrarily chosen cell 41 in the city center has an average load of ~13.9 Mbyte/s with standard deviation of 12.52 Mbyte/s, the maximum load of the cell is ~33.8 Mbyte/s while over 13 hours in the evening and the early morning the load is constantly below 10.0 Mbyte/s. This translates to a CoV of 0.90. If, in contrast, we regard the fluctuations over cells at fixed points of time, we see that during the peak hours the CoV is as high as 0.99 while the minimum and average CoV are 0.56 and 0.72, respectively. This indicates that the load distribution over locations is unbalanced, too. Here, the extreme conditions can be observed during the busy hour. Our model predicts that there exist hot spots and areas with nearly no activity/users at the same time. Moreover, over day there are heavy fluctuations of load. This clearly contradicts with the distributions of the random waypoint and random walk model. In contrast, our model exhibits multiple local maxima in user population and load, which build up and disappear over day. This heterogeneity in user densities and workload clearly mandates for adequate access technology like, for example, hot spots to provide for high capacity within selected areas for next generation networks to adequately support mobility.

3.6 Summary

In this chapter, we presented a novel approach towards a mobility/workload model covering the macroscopic effects observed within urban areas. Moving from traditional cellular networks to wireless local and metropolitan area networks results in smaller cells and traffic types other than voice. We have shown in this context that the effects of user mobility cannot be neglected and manifest in high traffic load fluctuations on medium timescale. Currently, the design and operation of cellular or ad hoc networks do not consider this influence.

After surveying the current state-of-the-art approaches towards mobility modeling, we formulated a model to fit our special needs, which is a hybrid of an empirical mobility model combined with a synthetic traffic model. We borrowed from models used for transportation planning while emphasizing the aspects needed for investigations of traffic related issues of wireless communication networks. First, the fundamentals of the mobility model and the classification of the locations as well as the user behavior have been described in detail. Second, we augmented the results of the mobility part with traffic predictions to form a workload model. We derived and explained the analytical description of our model and gave detailed definitions and equations to allow for easy instantiation. For further usage, we instantiated the model for the case of Darmstadt, a large city in Germany. Lessons learned with respect to the instantiation of our model include that data not intended to serve as a basis for network traffic analysis needs significant post-processing.

The model predictions have been carefully studied. The insights obtained show major fluctuations of user densities and user traffic, respectively, induced by personal mobility. We see an increasing need to deal with time varying traffic on multiple timescales. This presents a case for new routing architectures and algorithms to support these mobility induced characteristics. At the same time, we need to maintain highly dependable network operation as described in Chapter 2. Accordingly, we describe novel network architectures and routing algorithms in the next part of this dissertation to allow for better adaptation of cellular networks to mobility induced effects resulting in higher performance of the network. Moreover, we use the results obtained in this chapter to perform an experimental analysis of various QoS routing strategies in Chapter 4. The insights gathered in this chapter are also relevant for the area of dependable ad hoc routing studied in Part III of this dissertation. However, due to the immaturity of current technologies, ad hoc networks are not ripe yet to operate on such large scales—here our study of dependability focuses on smaller levels, being one first step to reach out for highly adaptive and scalable ad hoc networks.

"Any sufficiently advanced technology
is indistinguishable from magic"
—(Arthur C.) Clarke's Third Law

Part II
Dependable Routing for Cellular Networks

4 Dependable Routing for Cellular Networks with Static Infrastructure

The establishment of the concept of routing dependability for cellular and ad hoc networks in Chapter 2 and the study of the effects induced in networks by mobility in Chapter 3 form the foundation of this work. Starting with this chapter we investigate cellular networks. In particular, we develop an architecture for a static infrastructure-based radio access network. Traditional cellular architectures are mainly organized in tree hierarchies. Coupled with centralized control mechanisms, these architectures allow for easy administration and management on one hand while, on the other hand, leave only little room for decentralized optimization. We present an architecture, which is neither structured as tree, nor centrally managed. To leverage the capabilities of this architecture, we discuss and study the impact of various routing strategies as well. For this, we perform an extensive simulation study using the workload modeled in Chapter 3. The studied routing strategies include shortest path routing and delay constrained routing as well as various multipath routing variants. Moreover, we investigate different traffic distributions. Our findings are that multipath routing is able to enhance the utility of the network and allows for efficient resource management within wireless metropolitan area networks. However, we see further room for improvement and address the identified shortcomings in Chapter 5, where we demonstrate the power of a novel variability-augmented cellular architecture in combination with novel routing paradigms.

4.1 Motivation

The intersection of two important trends in modern society, namely individual mobility and personal communications, contributed to the development of cellular networks for mobile communications. These networks have steadily matured over the recent decades and currently provide for nearly ubiquitous availability and coverage in the highly developed countries of the world. Initially driven by the need for voice communication, cellular networks evolved to carry data as well. Today, the state-of-the-art in voice and data communication manifests in the third generation of cellular networks (3G) as specified in the International Mobile Telecommunications framework (IMT-2000) (see, for example, Schiller [Sch03], Eberspächer et al. [EVB01], Rappaport [Rap01], or the standardization bodies [3GPP04] for detailed technical and historical information on the subject of cellular networks).

Existing cellular networks up to and including 3G systems are mainly homogeneous with respect to the radio access network and the core network. Our work builds on and extends these concepts. We do not limit our work to homogeneous systems but also regard heterogeneity of intermediate systems and end systems as an important aspect. Thus, we are consistent with recent proposals for *beyond 3G networks* [PKH+02] and the vision of a *wireless Internet* [Ell01], which embrace a heterogeneous set of access technologies. These technologies range as far as from satellite systems over nationwide cellular networks and large scale wireless metropolitan area networks down to wireless local area networks and small scale personal area networks. However, they do not only complement each other but also compete in the market.[8]

The variety in radio access networks resulting from the above envisioned scenario gives rise to several interesting research challenges. The results presented in Chapter 3 indicate that today's hierarchical and mostly tree-structured radio access network topologies as well as static resource management approaches need reconsideration. Induced by user and device mobility—while tightly coupled with usage and application-specific traffic patterns—these radio access networks experience heavily varying loads on different timescales. We regard the investigation of the hereby caused problems as crucial to allow for efficient and dependable control of large scale wireless communication networks.

In this chapter, we approach two particular challenges: first the development of a novel cellular network architecture (the so-called MobQoS architecture) and second the evaluation of novel resource management paradigms within this architecture. Our architecture is designed to be distributed and decentralized and supports variable cell sizes. We identified resource management issues as critical for dependability in infrastructure based networks in Chapter 2. To allow for novel paradigms in this context, our network

8) Currently, this phenomenon manifests in the emergence of medium scale community and metropolitan area networks that are based on inexpensive wireless local area network technology. These networks pose a significant competition to established and upcoming cellular networks.

design resembles tightly meshed routers to interconnect the cells. As a result the architecture is flexible with respect to the technology operating on top and supports, for example, QoS routing strategies to optimize the network performance. Despite the fact that a huge amount of work is performed in the area of quality of service routing, the implications of these mechanisms within the context of cellular networks are not well understood. This is especially true if we consider the transition from mainly voice-oriented (connection-oriented) towards data-centric (connectionless) cellular networks. We further evaluate and analyze the potential of QoS routing mechanisms for resource management in future cellular networks. The chosen scenario for our study is a model of a cellular radio access network covering Darmstadt. Since analytical tractability is infeasible due to the given complexity of the scenario, an experimental analysis is performed. We instantiate the generic cellular architecture for the case of the city center of Darmstadt and are able to use the results from Chapter 3 to serve as workload for our study. The major contributions of this chapter are:

- The development of a *novel cellular network architecture*.
- An *instantiation* of our architecture *for the case of Darmstadt*.
- The *experimental design* to allow for an *exhaustive simulation study*.
- The analysis of *shortest path routing*, *delay constrained routing* as well as of various *multipath QoS routing strategies* with respect to network performance.
- The evaluation of the influence induced by *different traffic distributions*. This includes various degrees of external/internal traffic[9].

4.2 Outline

The remainder of this chapter is organized as follows. In Section 4.3 we motivate and describe a novel static network architecture for cellular networks. The decentralized nature of this architecture compared with the strict tree hierarchies in current systems results in a high gain in flexibility. The architecture is instantiated for the case of Darmstadt likewise the mobility/workload model in Chapter 3 to allow for an experimental analysis. We present the case for QoS routing in cellular network in Section 4.4. To qualify and quantify the performance and dependability of our architecture, we conducted extensive simulation studies of various QoS routing algorithms. We include the description of the experiments and detail the experimental design as well as the selection of factors and workload. The analysis and interpretation of the results is presented in Section 4.5. In Section 4.6 related work is surveyed. Finally, we lead over to Chapter 5,

9) Today's cellular networks are mostly operating in the connection-oriented domain. To carry data traffic, they use an overlay structure, which results in non-optimal routing via a single gateway node only. To explore the gain in flexibility of our proposed architecture, we also investigate the direct routing of local traffic.

where the dynamic case of dependable routing within cellular networks is discussed. The coarse structure of this chapter is shown in Figure 26.

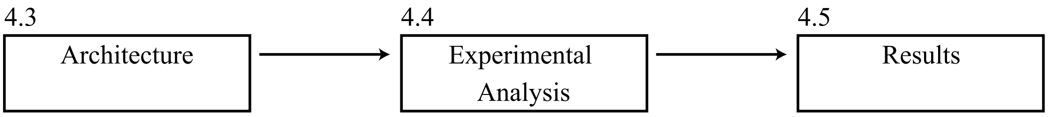


Figure 26: Structure of Chapter 4.

4.3 Architectural Considerations

This section comprises the description of a novel architecture as a basis to support dependable routing for cellular networks. First, we describe the requirements and motivate the restriction to the static case, that is, to a fixed infrastructure. Second, we derive the architecture based on a general radio access network architecture. The principle design goals of our architecture are *scalability* to allow for easy growth of the network as well as *flexibility* in terms of alternate paths to introduce the necessary degree of freedom for decentralized resource management algorithms.

Requirements

Our architectural requirements can be derived from existing cellular networks in combination with the special necessities of our application domain. The most important requirements of existing cellular telecommunication networks can be summarized as follows [EVB01] [Sch03][10]:

- The network allows for *seamless mobility support* for *users* and *end systems*.
- The network allows for *roaming* into networks of different operators if business agreements exist.
- The network supports various *classes of services* (for example, *conversational*, *streaming*, *interactive*, or *background traffic*) with well-defined quality of service parameters.
- The network is *highly available* by means of redundancies in connections and infrastructure components.
- The network is *secure* for *signalling* and *data-communication.*
- The network allows for *metering*, *accounting*, and *billing* of the user service.

These requirements mark the desired features of cellular network architectures and are usually reached on top of the characteristics presented in Section 2.4.1. In particular, the network architecture relies on an infrastructure-based paradigm and is organized strictly hierarchical and follows a tree structure. Network control is centralized in the core of the network. While the access network is a non-routed network, the core network provides

10) See also the draft and standard documents of the respective standardization bodies. For GSM and UMTS the standards are publicly available from the 3GPP organization [3GPP04].

for routing functionality. The network service is mostly connection-oriented and reliable. Moreover, the network is closed, that is, only subscribers participate. Similarly to traditional cellular networks our routing system consists of static infrastructure components; end systems are not part of the radio access network. See Figure 27 for the radio access network architecture of the Universal Mobile Telecommunication System (UMTS), which currently represents the state-of-the-art in cellular telecommunication networks. For our case we relax some of the above outlined requirements to gain flexibility in network design.

Design Principles and Restrictions

While our network design bases on existing infrastructures, we add some design elements, which have been inspired by the work of Perlman [Per99]. Primarily, we want to support a wide range of applications and technologies and solve the problem of future radio access networks as general as possible. We assume that:

- The network is able to integrate a *heterogeneous set of access technologies* for the radio access, for example, various types of hot spots or local area network technologies.
- There is no central instance for network control and resource management but the architecture solely relies on *decentralized mechanisms*.

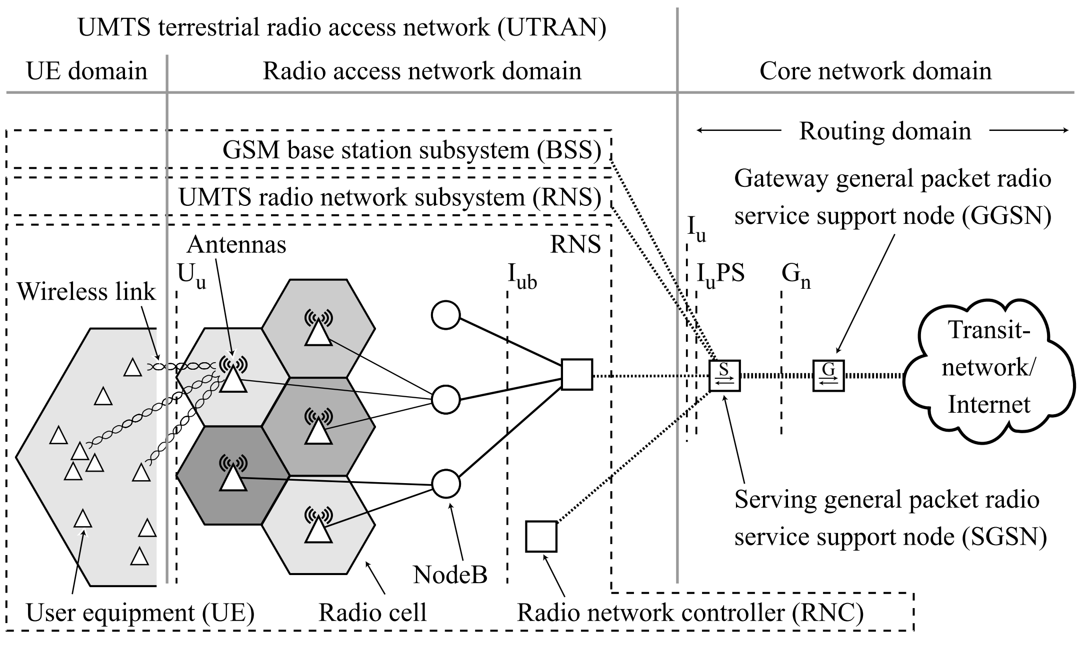


Figure 27: The UMTS radio access network architecture. The architecture separates the UTRAN from the core network. Moreover, there are clear interfaces between the individual tiers of the architecture (U_u, I_{ub}, I_u, I_uPS, and G_n). The RNS is organized in tree structure. One NodeB may service multiple antennas/radio cells. The RNCs control multiples NodeBs. The SGSN operates in the packet-switched/routing domain and also allows for legacy GSM-BSS to be connected. The GGSN is the gateway to other transit-, foreign-, and internetworks.

- The *loose or tight coupling to 3G/B3G* environments should be possible. In reality this may imply that our proposed architecture is directly connected to the packet-switched domain of the 3G/B3G network.
- The routing system follows a *connectionless best-effort* paradigm.
- The infrastructure components are assumed to provide *IP-routing functionality.*

The most severe change in design compared to traditional cellular networks is our choice to adopt Internet paradigms, that is, a connectionless and unreliable routing service. This mandates the distribution and decentralization of the forwarding- and control-logic on one hand. On the other hand, we are able to depart from the strict tree hierarchy of the network towards tightly meshed topologies that provide for alternate paths between arbitrary source/destination pairs. Moreover, current approaches towards traffic engineering mainly rely on centralized components and build upon explicit mechanisms for signaling and admission control. To be consistent with our assumptions of distributed resource management mechanisms and a connectionless service paradigm we have to reconsider these mechanisms, too. We opt that our approach should be able to work without explicit signaling and without a need to operate on the level of individual flows, thus, following a stateless core philosophy.

The routing system is the primary target of our investigation. Based on our basic assumption of a connectionless network service, we define the self-adaptability to changing network and traffic conditions with special emphasis on load fluctuations to be a further characteristic of the routing algorithm. Having introduced the most important design principles we also have to consider some restricting assumptions to provide boundaries for the further analysis of our suggested solution within this chapter:

- The routing mechanism should support exactly *one class of service* likewise available in today's Internet.
- The network is equipped with a sufficient *admission control* to ensure that offered traffic is not accepted if the capacity of the network is exceeded.
- The *mobility support* is out of scope for our work. We expect this issue to be solved using appropriate mechanisms on the network layer such as Mobile IP.

4.3.1 Modeling the MobQoS Architecture

To be able to depart from traditional cellular network design, we need to sacrifice some of the design goals of these networks. Towards a distributed and decentralized architecture we opt to move the "intelligence" from the core of the network towards the edge. In particular, we introduce additional routing functionality at radio access server (RAS) level of the network. To reach flexibility in network design we do not require strictly tree-structured topologies but allow the introduction of interconnections between RASs to allow for resource management starting at this level. By distributing and decentralizing functionality, we also leverage the characteristics of heterogeneity, scalability and adaptability, which should allow for both, network planning and seamless network growth while supporting high-bandwidth pico-cells and hot spots. Next, we develop an novel architecture to fulfil the above discussed assumptions and requirements.

Our MobQoS architecture features an evolutionary approach based on traditional cellular network architectures. Like traditional architectures, we follow a three-tier design depicted in Figure 28. Goal is to open up the strictly tree-based architectures of today's cellular systems. Our architecture is based on concepts from telecommunication systems, especially UMTS technology [3GPP03]. To introduce the necessary flexibility, we integrate concepts discussed within the IETF, namely the architecture underlying Mobile IPv4 [Per02] and Mobile IPv6 [JPA04]. We also include work from the areas of QoS architectures [Hus00] as well as traffic engineering [ACE+02]. The mobile nodes or terminals (MN) of the customers are associated with wireless base stations, the so-called radio access points (RAP), representing the last hop of the provision network. The function of the first tier can be described as radio access. The second tier, the provision radio access network, comprises radio access server (RAS), which are used to attach multiple RAPs. Within the radio range covered by all RAPs attached at one RAS, mobility is supported by appropriate link layer mechanisms such as, for example, handover. Each RAS has built-in IP-Router functionality. The RASs are meshed with their neighboring RASs and, thus, allow the start of resource management at this level. The third tier of the architecture is the core of the radio access network, which is built by radio access routers (RAR), which are fully meshed. Between selected RASs and RARs are uplinks. The transition to other networks, such as the provider backbone or the Internet, is performed by one or multiple edge gateways (EGW). At this point, the technology may be mapped

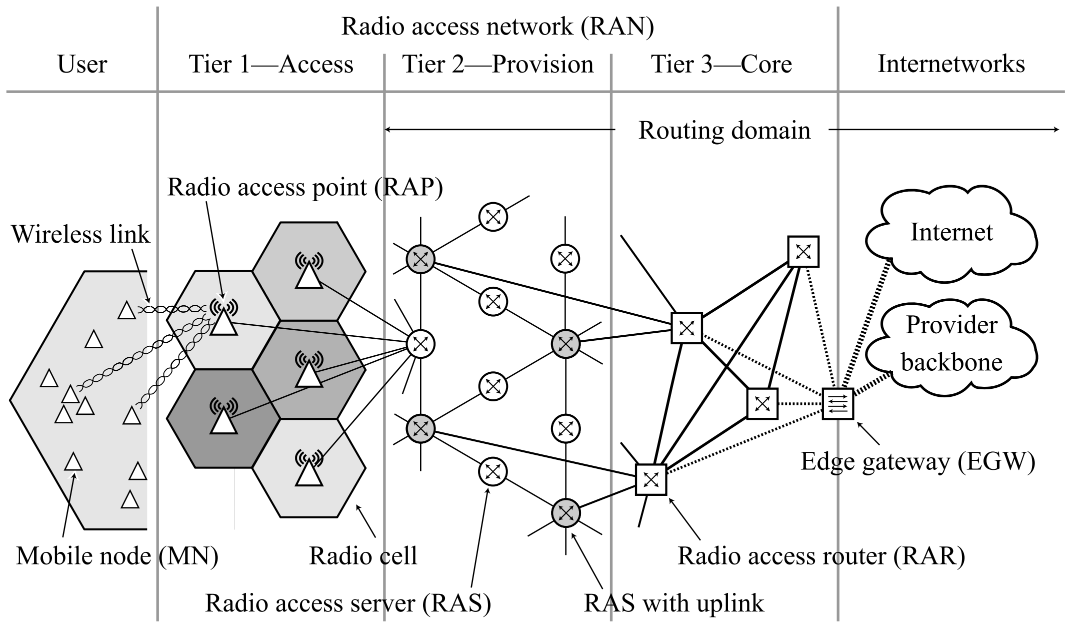


Figure 28: The MobQoS radio access network architecture. The RAN consists of three tiers; access, provision, and core. The RASs are partially meshed and may serve multiple RAPs/radio cells each. Selected RASs provide uplinks to the fully meshed RARs. The EGW connects with the provider backbone (home network) and/or other transit-, foreign-, and internetworks.

to the according mechanisms and strategies of the provider backbone network or the Internet and vice versa. See Figure 28 for a schematic overview of the MobQoS radio access network architecture. Differing from a traditional RAN, the MobQoS architecture is highly flexible with respect to the interconnection of individual RASs or RARs. This flexibility on the provision tier simplifies the reconfiguration and extension of the network, one example being the integration of hot spots.

Instantiation of the Architecture for the Case of Darmstadt

Based on our mobility/workload model described in Chapter 3, we have instantiated our radio access network architecture for the case of Darmstadt to show its applicability for a metropolitan area network. The instantiation is based on current state-of-the-art network component technology to mirror the possibility of realistic implementation. In short, we modeled the full coverage of the area with high bandwidth hot spots. The hot spots are controlled by RAPs and allow for wireless access using 802.11 or similar technology [IEEE99a] [IEEE99b] [IEEE01] [IEEE03]. We expect the hot spots to be sparse or dense according to the traffic demands of the covered area. The resulting cell sizes are expected to be in the range of some 10 m to some 100 m. As a result we obtain a high number of radio cells compared with traditional cellular systems. The RASs aggregate multiple of these cells or hot spots and form the second tier of the access network. To service the large number of radio cells, our topology significantly increases the number of RASs, too. As already mentioned, the RAS component includes routing functionality. We expect the backbone of the RAN, that is, the RAP-RAS connection, to be wired, one possible implementation being an ethernet connection [IEEE02]. The RAS-RAS and RAS-RAR connections are also wired.

For the case of Darmstadt, we modeled a homogeneous distribution of tightly meshed RASs. See Figure 29 for different degrees of interconnections. For our investigation we stick with the high degree of RAS-RAS interconnection. This allows to investigate the implications of a tightly meshed structure in contrast to the strict tree structure of existing cellular networks. It should be noted that in reality the coverage of such a cellular network as well as the degree of interconnection between the individual RASs may be significantly reduced compared to our assumptions. Krop [Kro02] gives more detailed information of the instantiation process and the resulting network topology.

4.4 Experimental Analysis of Selected QoS Routing Strategies

The architecture developed in Section 4.3 introduces flexibility with respect to the routing framework. Compared with traditional cellular architectures, this allows to introduce novel paradigms in network control. In contrast to our distributed and decentralized architecture, most existing approaches towards traffic-engineering in cellular networks rely on centralized components and build upon explicit mechanisms for signaling and resource admission. Our distributed and decentralized network architecture opts for the investigation of network control mechanisms similar in nature.

As a result, the goals to be addressed are the development and investigation of distributed routing algorithms for resource management purposes to ensure resource-efficient network operation while supporting mobile users. We perform a simulation study to compare the performance of selected protocols and strategies. Here, we are particularly interested in the investigation of resource management and load-balancing issues, because these network performance metrics contribute to the perceived dependability of the network. In particular, the system boundaries for our analysis are defined by the MobQoS scenario [HSHS04]. Since the target scenario of cellular networks is of special interest in metropolitan areas, we use the pathological case of a radio access network covering Darmstadt. The topology we investigate is the tightly meshed metropolitan

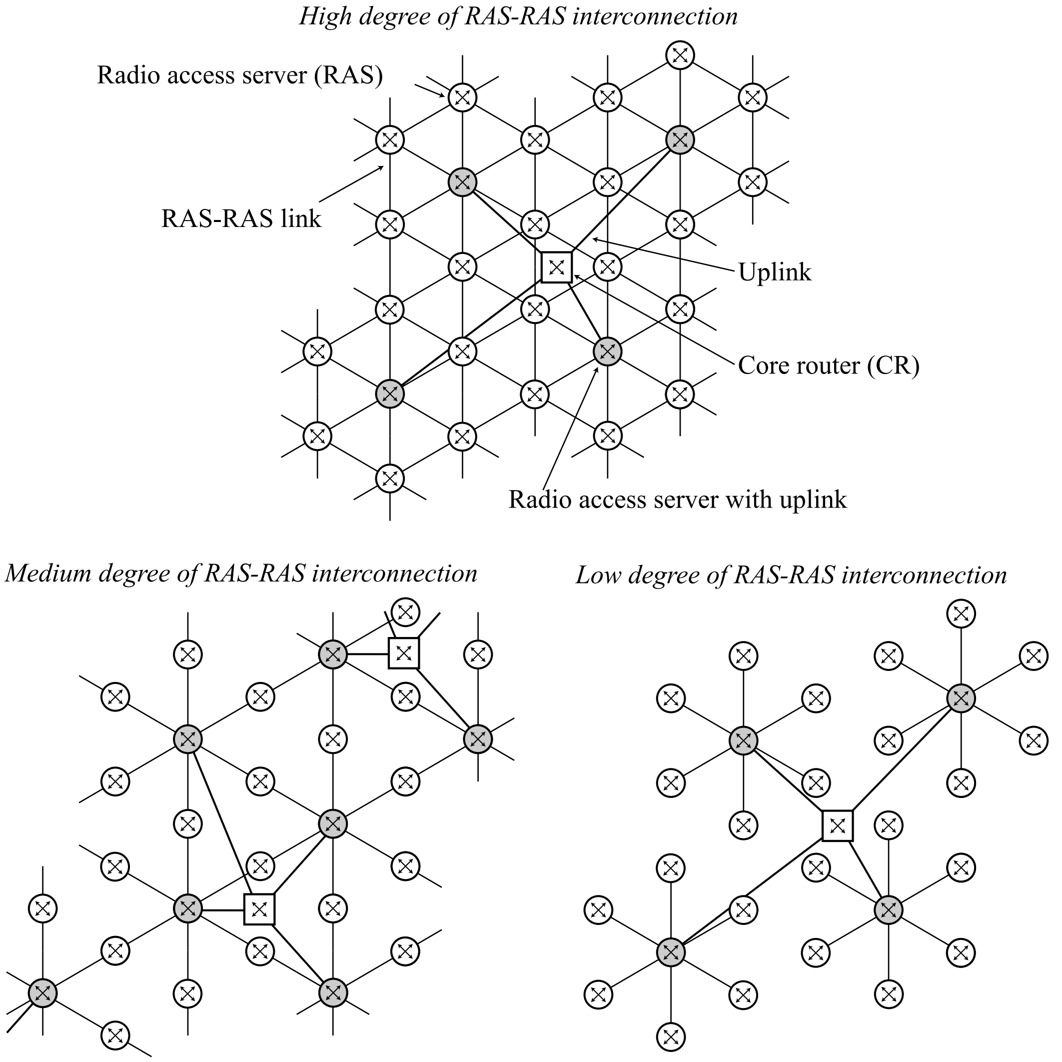


Figure 29: Examples for a regularly meshed radio access network topology. We show samples for tightly, medium, and lowly meshed topologies. All shown links are wired. The node degree on RAS-RAS level is 6, 2, and 1 respectively. Uplinks for the high degree of interconnection cover 1, for the medium degree 4, and for the low degree 7 nodes. Please note that a RAS can be responsible for multiple radio cells and/or sectors in cells. The core network structure as well as the wireless RAPs are not shown.

radio access network infrastructure we instantiated for the case of the city of Darmstadt in the previous section.

Our distributed architecture does not fit well with today's static and centralized resource management approaches. Hence, we consider distributed QoS routing strategies as a replacement (Section 4.6 below details the related work in the area of QoS routing). The special focus of our analysis is now mapped onto the investigation of the performance of selected QoS routing algorithms to deal with the predicted traffic demand within future radio access networks. We expect the mobility induced characteristics as well as the possibilities introduced by novel applications and services within future networks to yield network traffic different from the one that can be observed in today's cellular networks. Within our MobQoS scenario, we concentrate on a transition from voice oriented communication towards multimedia communication. In terms of workload this mirrors a shift from constant bitrate traffic towards a more Internet-like heterogeneous traffic-mix including conversational, streaming, interactive, and background traffic.

The workload for our analysis is formed using our mobility and workload model presented in Chapter 3. The instantiation for the case of Darmstadt captures the mobility-induced effects on the data-plane. Since analytical tractability is out of scope for the given complexity of the problem, we use the methodology of experimental analysis for our investigation [Jai91]. We present the experimental design including the parameters studied. Important trade-offs being the investigation of *single-path vs. multipath operation*, *static vs. dynamic algorithms*, different *link-cost metrics* and *scheduling disciplines*. Thus, we are able to present detailed results for the performance aspect of dependability in cellular networks for our setup.

4.4.1 Experimental Design

This section comprises the description of the experimental design based on the selection of factors and workload. We follow the methodology proposed by Jain [Jai91] for our experimental analysis and adopt the individual steps to suit our case. The methodological steps can be summarized as follows:

1. The *definition* of the *system*, *goals*, and *services*.
2. The *selection* of the *metrics*.
3. The *definition* of the *parameters* to study.
4. The *selection* of the *factors/elements* of the parameter set.
5. The *choice* of the *evaluation technique*.
6. The *selection* of the *workload*.
7. The *design* of the *individual experiments*.
8. The *analysis* and *interpretation* of the obtained *data*.
9. The *presentation* of the *results*.

We define the *system* to represent the above described metropolitan radio access network structure. The *goal* of our evaluation is a comparative performance study of existing QoS

Table 1: Routing Strategies: Factors, Levels, and Descriptions for the RAN-I Topology.

Factor	Description			
	Routing	**Metric**	**Type**	**Scheduling***
Algorithm-*DST*	Static	Distance	Single-path	—
Algorithm-*DLY*	Static	Delay	Single-path	—
Algorithm-*SMP*	Static	Distance	Multipath	RR
Algorithm-*DMP*	Dynamic	Delay	Multipath	RR
Algorithm-*WMP*	Dynamic	Delay	Multipath	WRR

* The multipath scheduling disciplines are round robin (RR) and weighted round robin (WRR).

routing protocols to provide the *service* of dependable routing, which translates to efficient resource management ensuring resource-optimal network operation whilst supporting mobile users. As described in Chapter 2, we are especially interested in the network performance as *metric* for the utility of the network, representing an efficient resource management. This, however, cannot be measured easily within our complex scenario. Thus, we need to derive response variables that represent this metric. We measure the packet loss and load of individual link classes as well as the average path length. Moreover, we investigate the end-to-end delay of individual flows and packets acting as indirect measures for the efficiency of the resource management. In our case the type of routing algorithm is an important *parameter* (*predictor variable*). Since we are interested in the routing performance, we include various different routing strategies in our study. The *factors/elements* of the predictor variables include shortest path algorithms as well as delay constraint algorithms for single-path and multipath operation. We detail the parameter/factor set of our study below. Due to the complexity of our problem, simulation was chosen as the *evaluation technique*. We implemented a ns-2 [FV04] based simulation environment modeling our scenario. The *workload* model implicitly includes the factors of user mobility and user behavior. We use our hybrid workload model as described in Chapter 3. The *experimental design* is detailed below, as well as the *analysis* and *interpretation of data* and the *presentation of results*.

4.4.2 Parameters, Factors, Workload, and Experiment

Parameters and Factors

The parameter of utmost importance for our analysis is the *routing strategy*. Other parameters include the *influence of user mobility* and the *influence of the traffic distribution*. The routing strategies investigated include combinations out of the sets {*static routing*, *dynamic routing*}, {*single-path routing*, *multipath routing*} and {*distance metric*, *delay metric*}. In particular, we include the following routing strategies summarized in Table 1 in our investigation:

- Algorithm-*DST*, a static shortest-distance single-path algorithm. Algorithm-*DST* represents the standard Internet routing strategy, that is, a link-state protocol in combination with Dijkstra's algorithm [Dij59], which operates on a shortest-path metric: $cost = distance$. One example for a protocol of this class is the OSPF (Open Shortest Path First) algorithm [Moy98].
- Algorithm-*DLY*, a static minimum-delay single-path algorithm. Algorithm-*DLY* favors paths with smaller delays irrespectively from the hop count compared to Algorithm-*DST*. Various QoS routing algorithms including quality of service enhanced OSPF (Q-OSPF) [AWK+99] are designed to operate on minimum-delay for individual flows. The link-cost metric is the propagation delay to the destination: $cost = delay$. In our case the algorithm operates on the granularity of all traffic for a specific destination instead of the granularity of individual flows. This causes route flapping in the dynamic variant of Algorithm-*DLY*. Hence, we only investigate the static case for this algorithm.
- Algorithm-*SMP*, a static equal-cost shortest-distance multipath algorithm. In our analysis Algorithm-*SMP* represents the class of widely used equal-cost algorithms including, for example, the OSPF-ECMP (Equal Cost MultiPath) variant described by Moy [Moy98].
- Algorithm-*DMP*, a dynamic minimum-delay multipath algorithm. Multipath scheduling is performed using round robin on packet level. The link-cost metric used for route computation at time t_1, $cost(t_1)$, is the average delay to the destination measured for the time-interval $\Delta T_1 = t_1 - t_0$ from the last route-computation; $cost(t_1) = delay(\Delta T_1)$. In contrast to Algorithm-*SMP*, the adjustment of routes is dynamic in Algorithm-*DMP*, thus, resembling some of the concepts presented in the OSPF-OMP (optimized multipath) described in [Vil02]. We do not model the exact heuristic for load distribution as detailed in [Vil02] but operate on the delay to indicate high load situations. In contrast to [Vil02] multipath scheduling uses a round robin for reasons of simplicity.
- Algorithm-*WMP*, a dynamic minimum-delay multipath algorithm. Similar to Algorithm-*DMP*, we use the metric $cost(t_1) = delay(\Delta T_1)$. Multipath scheduling is performed using weighted round robin on packet level. Algorithm-*WMP* coarsely models the near-optimal algorithm proposed by Vutukury et al. [VGLA99]. Vutukury introduces two timescales: t_l on which the algorithm computes new routes and t_s on which the algorithm computes new weights for the multipath scheduling. Dissimilar from Vutukury we only model the t_l period to reduce the overhead for route computation. Moreover, in contrast to Vutukury, we do not use the incremental delay as a metric but the delay itself.

The initial implementation of Algorithm-*DMP* and Algorithm-*WMP* resulted in oscillations for high load situations, which is described in detail later. To limit this behavior we also implemented two additional variants of these algorithms using different cost metrics. The first variant, denoted as Algorithm-*DMP*1 and Algorithm-*WMP*1, uses an exponentially weighted sum to damp the oscillations. Within the experiments, the exponential weighted sum is calculated to alleviate wrong routing decisions:

$$cost(t) = \frac{delay(\Delta T_1) + delay(\Delta T_0)}{2} \tag{12}$$

A second variant of the dynamic algorithms (Algorithm-*DMP*2, Algorithm-*WMP*2) was implemented to follow a more aggressive strategy, namely to add a correction term to the measured delay that depends on the trend in change of delay and traffic-rate observed over the last interval. The metric for Algorithm-*DMP*2 and Algorithm-*WMP*2 is:

$$cost(t) = delay(\Delta T_1) + \frac{(delay(\Delta T_1) - delay(\Delta T_0))(sending\ rate(\Delta T_1))}{(sending\ rate(\Delta T_1) - sending\ rate(\Delta T_0))} \tag{13}$$

The algorithm and the workload act as predictor variables while other factors of the experiments are kept constant. These secondary factors include the parameterization of the routers and links performed during the instantiation of the architecture for the case of Darmstadt. See Figure 30 for a graphical illustration of the chosen topology of the radio access network (RAN-I network topology) and Table 2 for the capacity and propagation delay of the individual links. Moreover, the buffer space of the routers is fixed to store an equivalent of 5 ms of data at full link speed during all experiments.

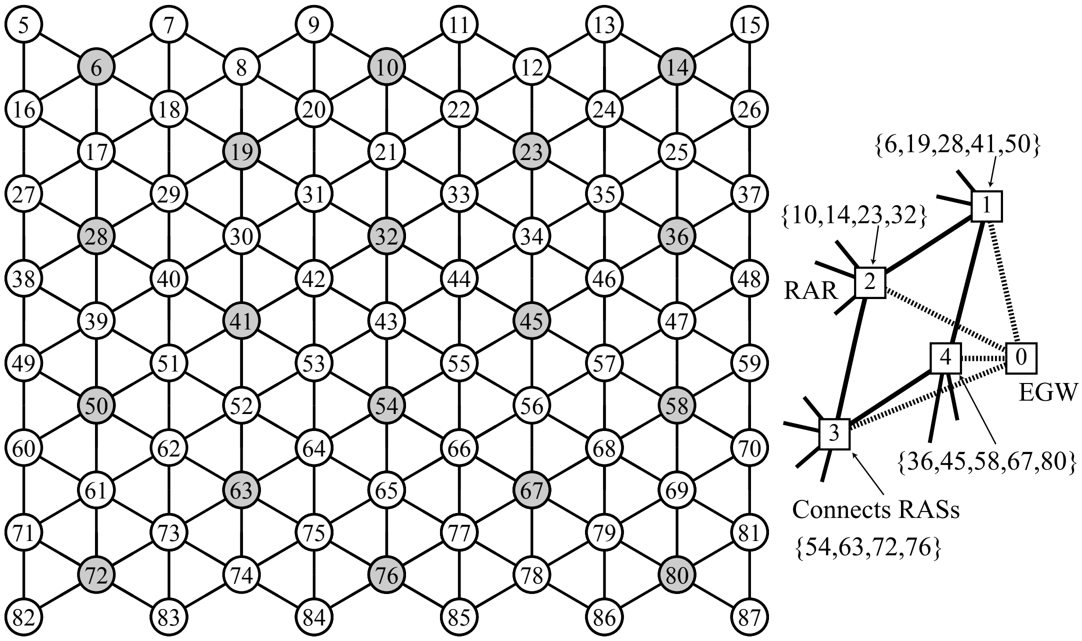


Figure 30: RAN-I network topology. The topology has been designed to cover the city of Darmstadt and fits with the workload model presented in Chapter 3. The regular topology provides for a high number of alternate paths. The links differ in capacity and propagation delay for each tier. See Table 2 for the complete parameterization of the topology.

Table 2: Experimental Configuration of the RAN-I Topology; Test-series (4.*).

Factor	Description
Topology graph	RAN-I (see Figure 30), Graph order $p = 88$, graph size $q = 476$, average degree $deg = 5.41$.
Link characteristics	RAS-RAS link bandwidth = 20 Mbit/s, Propagation delay = 0.2 ms. RAS-RAR link bandwidth = 75 Mbit/s, Propagation delay = 0.3 ms. RAR-RAR link bandwidth = 100 Mbit/s, Propagation delay = 0.4 ms. RAR-EGW link bandwidth = 200 Mbit/s, Propagation delay = 0.5 ms.
Router characteristics	Propagation delay = 0 s, Queue size depends on the link bandwidth, Queue size = equivalent to 5 ms at full line speed.
Workload: Traffic classes, types, and distribution	All tests are conducted for all traffic distributions TrafDist-0 to TrafDist-4. The class distribution of traffic is fixed according to the workload model presented in Chapter 3. We model the traffic as aggregates of the four classes described in Table 3 using UDP as transport protocol, the packet size is 768 bytes. We use CBR as well as self-similar traffic to model the aggregate traffic. All nodes can act as source and destination. The traffic is distributed to the source destination pairs according to the workload model. See [Kro02] for details of the instantiation of the workload model.

Workload

For our investigation, we use our hybrid mobility/workload model described in Chapter 3. Using the generated workload as predictor variable we also investigate the influence of different traffic distributions. The starting point is derived from current radio access network tree structures. This corresponds to only external traffic for the end systems to/from a single edge gateway. To account for more flexible architectures as well, we investigate five different *e/i-ratios* (ratio external/internal traffic) as shown in Table 3. While distribution TrafDist-0 describes the situation of all traffic going to/from the edge gateway, this assumption is stepwise relaxed in TrafDist-1 to TrafDist-4. This reflects various amounts of local traffic, which may arise for future services. The traffic is separated using the four traffic classes {*conversational*, *streaming*, *transactional*, *background*} introduced in [3GPP03] since they provide a good guess of the intended usage of the network in question (see Figure 15 in Chapter 3 for our overall traffic estimate). The *e/i-ratio* is set for each class independently. We model different *e/i-ratio* for the different traffic classes because the expected applications are used and implemented in different forms. The *conversational* traffic is mainly induced by voice and video communication. We know from the area of telecommunication networks that caller and callee in such direct synchronous communication are often in the proximity. Consequently, we model *e/i-ratios* from 100/0 to 0/100 for *conversational* traffic. The origin of *streaming* traffic is likely an audio- or video-server. We reflect this behavior by modeling

Table 3: Traffic Distribution: Factors, Levels, and Descriptions for the RAN-I Topology.

	Description[1]			
Factor	**Conversational**	**Streaming**	**Interactive**	**Background**
TrafDist-0	100/0	100/0	100/0	100/0
TrafDist-1	75/25	90/10	80/20	85/15
TrafDist-2	50/50	80/20	60/40	70/30
TrafDist-3	25/75	70/30	40/60	55/45
TrafDist-4	0/100	60/40	20/80	40/60

[1] The traffic distribution is given as external/internal traffic ratio (*e/i-ratio)*. For example 80/20 means: 80% of total traffic to/from the edge gateway, 20% of total traffic internal.

e/i-ratios from 100/0 down to only 60/40. With emerging capabilities of cellular networks, we expect a change in nature of *interactive* traffic. From today's bulk load to access the web or databases towards more peer-to-peer style applications. The range of the corresponding *e/i-ratio*s from 100/0 to 20/80 reflects this expectation. The forth class modeled is the *background* traffic, which is modeled with *e/i-ratio*s ranging from 100/0 to 40/60. Table 3 provides a summary of the variables and respective levels of the investigated traffic distributions. See Figure 31 for a summary of the experimental design.

Experiment

We have chosen to implement a full factorial experiment [Jai91], since we regard only two predictor variables, namely the routing algorithm and the traffic distribution with

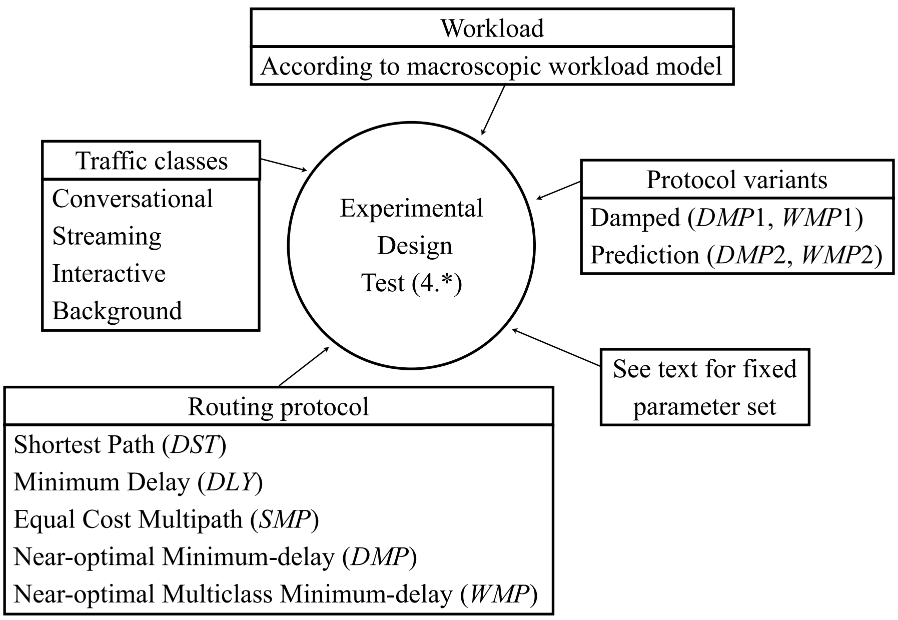


Figure 31: Experimental parameters for test-series (4.*).

five levels each. There are several other factors that are kept constant for all experiments. The experiment was implemented using ns-2 [FV04]. We have incorporated major changes to the ns-2 simulation framework to be able to simulate the scale of our scenario. We have also implemented a flexible multipath QoS routing algorithm that can be parameterized to model all different variants of routing strategies studied. Moreover, we needed to implement methods to transfer the results of our traffic model into the simulation environment. A more detailed description of the changes in ns-2 is given in [Kro02].

The scaling of the experiment to a different timescale is inevitable due to the size and complexity of the modeled scenario. We have chosen the simulation time to be 960 seconds to match the real time of 24 hours. Thus, one second in simulation equals 90 seconds in real time. The amount of traffic is given by our workload model, which inherently includes the mobility of users (see Chapter 3), and was used to parameterize the traffic agents at regular intervals. The updates of the workload were performed every 2 minutes (real time). The routing updates were performed every 6 minutes (real time). See Table 2 for the full description of the RAN-I network topology and Table 4 for an overview of the experimental parameter for all tests performed.

4.5 Results

The main goal of our analysis is the investigation of the performance of QoS routing mechanisms within our novel cellular network architecture. We want to achieve an optimal load-balancing in the network. We investigate various response variables and analyze the results on different aggregation levels of the obtained data, which are *link level*, *flow level*, and *packet/application level*. Besides measuring the exact numbers, we give various visualizations to allow for better interpretation of the results. For most parts, we provide results, which cover the complete 24 hour simulation period. If necessary, we give additionally the results for the busy hour.

The presentation of the results is structured as follows. Results on link level are given in Section 4.5.1. This includes figures to illustrate *load* and *loss* for *tier-2* (RAS-RAS links), *tier-2*/*tier*-3 (RAS-RAR links), and *tier*-3 (RAR-EGW links) for selected algorithms and traffic distributions as well as a summary of *load* and *loss* on link level for selected algorithms. The results obtained on *flow level* are given in Section 4.5.2. This includes results for *path length* as well as *delay per hop* on *flow level*. The results for *delay* of individual packets (*packet level*) are presented in Section 4.5.3. Finally, we give a *summary* of all *application* and *routing* related response variables in Section 4.5.4.

4.5.1 Results on Link Level

The results on link level include the average link load and link loss for all tiers of the network topology. We define the link load to be the ratio of used link capacity (measured in byte/s) to total link capacity (in byte/s) and the link loss to be the ratio of lost data per link (in byte) to total data sent over the link (in byte). The analysis of these metrics

Table 4: Experimental Parameter Set for the RAN-I Topology; Test-series (4.*).

Test	**Test (4.1), Algorithm-*DST***	**Test (4.2), Algorithm-*DLY***	**Test (4.3), Algorithm-*SMP***
Cost-metric	Distance (hop count)	Delay	Distance (hop count)
Multipath scheduler	No multipath scheduling necessary because of the single-path nature of the algorithms.		Round robin
Duration	960 s (24 h real time)		
Routing parameters	Static routing, the routing table is calculated at the beginning of the simulation.		
Workload	All tests are performed for all traffic distributions TrafDist-0 to TrafDist-4.		
Replications	1 for each test (focusing on the macroscopic effects of the workload)		

Test	**Test (4.4), Algorithm-*DMP*** **Test (4.5), Algorithm-*WMP***	**Test (4.6), Algorithm-*DMP*1** **Test (4.7), Algorithm-*WMP*1**	**Test (4.8), Algorithm-*DMP*2** **Test (4.9), Algorithm-*WMP*2**
Cost-metric	Delay	Smoothed delay	Delay prediction
Multipath scheduler	For all Algorithm-*DMP* variants: round robin, for all Algorithm-*WMP* variants: weighted round robin.		
Duration	960 s (24 h real time)		
Routing parameters	Dynamic routing, the routing table is periodically calculated, $T_l = 4s$ (6 min real time).		
Workload	All tests are performed for all traffic distributions TrafDist-0 to TrafDist-4.		
Replications	1 for each test (focusing on the macroscopic effects of the workload)		

allows to study whether the algorithms use the interconnections on RAS-RAS level efficiently. Moreover, we obtain measurements to show the different usage of the individual tiers of the architecture for all algorithms studied. Depending on load and loss we also judge if the performance of the algorithms is acceptable under normal conditions (investigated over 24 hours) and under heavy load (investigated over the busy hour). Due to the large amount of obtained results we select the most interesting results for presentation.

Load on link level

The average load on link level for the traffic distribution TrafDist-0 over 24 h can be seen in Figure 32 for all algorithms studied. The results clearly show a lower load in favor of the static shortest path algorithms including the equal-cost multipath as well. The dynamic algorithms operating on metrics different than hop count try to use lightly loaded links at the expense of longer routes. This increases the overall load. The absolute load per level cannot be used as indicator for the quality of the load-balancing, though. It allows, however, to see at which levels the algorithms act differently. In our example the increased use of RAS-RAS links is visible for all Algorithm-*DMP* and Algorithm-*WMP* variants. This result indicates that the multipath algorithms operating on delay are using additional paths, which may translate into a possible gain in performance within the modeled network if the load is well distributed.

The differences between the algorithms remain visible if we introduce local traffic. However, the individual algorithms show different tendencies under these conditions. We here discuss selected algorithms. Figure 33 to Figure 35 show the average load for

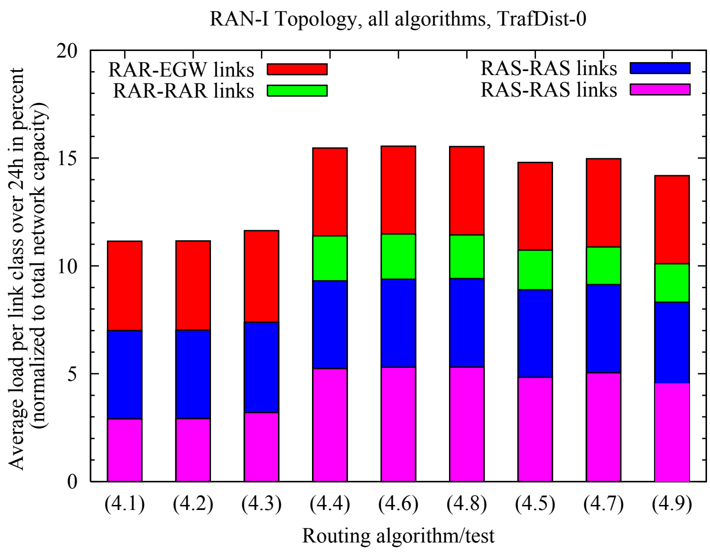


Figure 32: Average load of all algorithms per link class over 24 h. The load is normalized to the total capacity of the network. We show traffic distributions TrafDist-0.

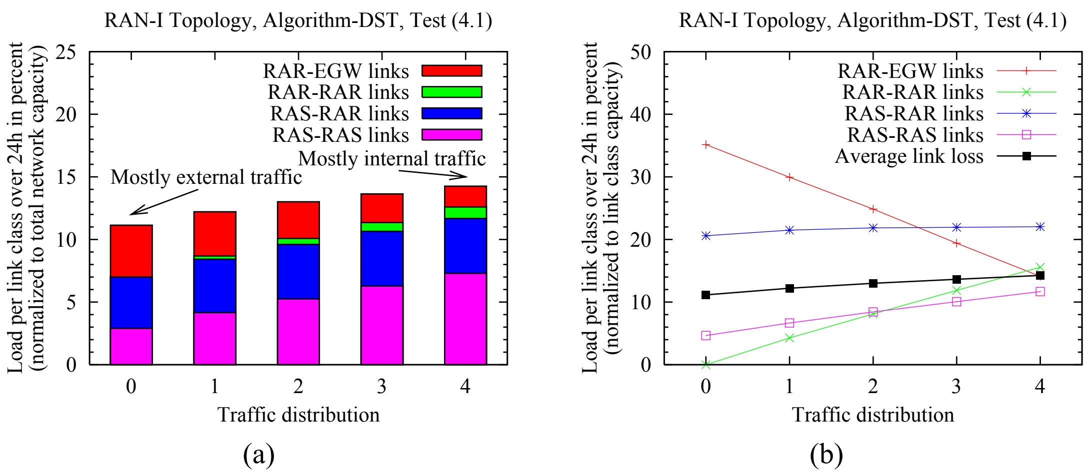
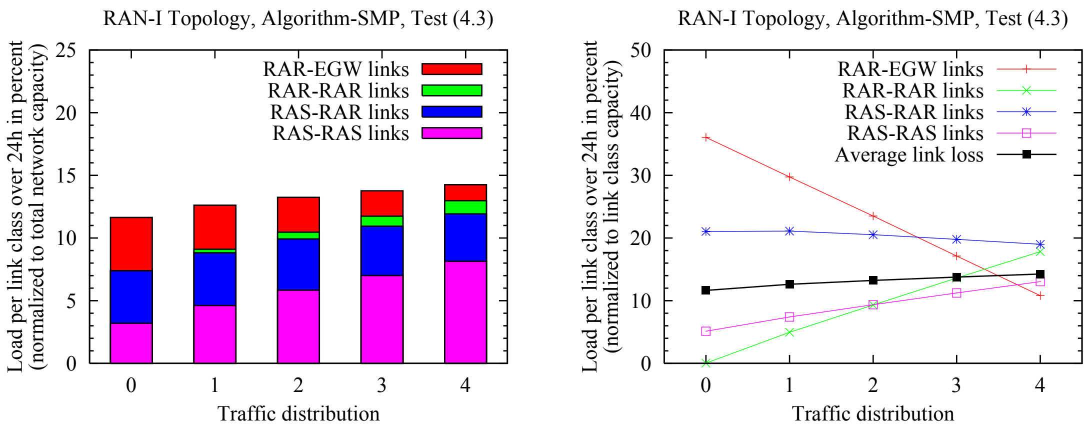


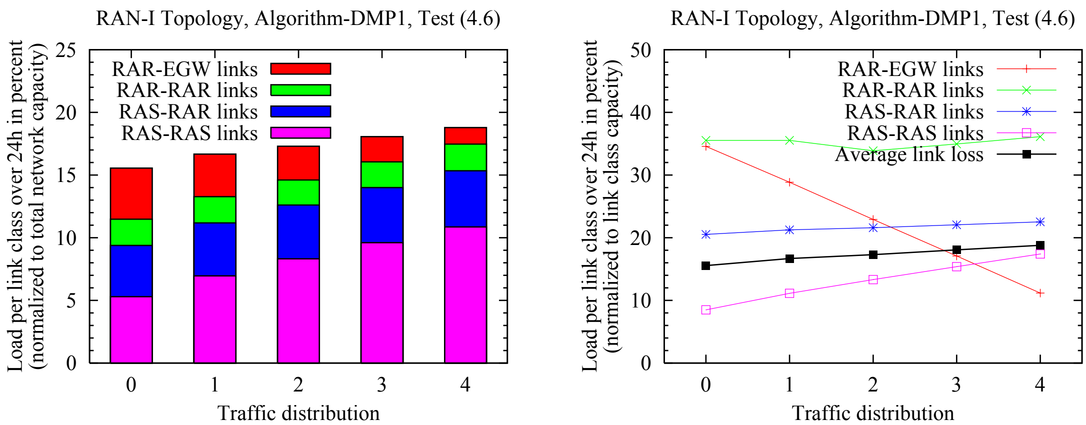
Figure 33: Average load of Algorithm-*DST* per link class over 24 h. The load is normalized to (a) the total capacity of the network and (b) the capacity per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.

Algorithm-*DST*, Algorithm-*SMP*, and Algorithm-*DMP*1 for all investigated traffic distributions. We use two different visualizations to show (a) the influence of the individual link loads vs. the total network load and (b) the trends in load for each link class in our figures. Algorithm-*DST* uses the RAR-RAR links only moderately, even for a high degree internal traffic. Moving towards local traffic reduces the workload of the RAR-EGW links as expected. The load of Algorithm-*SMP* is quite similar compared with Algorithm-*DST*. For the class of dynamic algorithms we depict Algorithm-*DMP*1. The average load for Algorithm-*DMP*1 is higher compared with Algorithm-*DST* and Algorithm-*SMP*. This accounts for the use of alternate paths. We see especially that Algorithm-*DMP*1 exploits the backbone links even for traffic distribution TrafDist-0. A remarkable result for Algorithm-*DMP*1 is the use of the RAR-RAR links for all traffic distributions.



(a) (b)

Figure 34: Average load of Algorithm-*SMP* per link class over 24 h. The load is normalized to (a) the total capacity of the network and (b) the capacity per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.



(a) (b)

Figure 35: Average load of Algorithm-*DMP*1 per link class over 24 h. The load is normalized to (a) the total capacity of the network and (b) the capacity per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.

Loss on link level

The load on link level yields no meaningful interpretation without considering the loss on link level at the same time. Packet loss on link level is an indicator for congested links and areas within the network. Moreover, the hierarchy level at which loss occurs is closely related to the fundamental principles of the underlying algorithms, which interdepends with the chosen topology. In the following we illustrate the loss characteristics for selected algorithms in all investigated traffic distributions. Figure 36 shows the average loss per link class for all algorithms using TrafDist-0. Differences on the overall loss as well as on the level at which loss occurs are clearly visible. Both single-path algorithms are clearly limited by the capacity of the links. Figure 37 shows the loss behavior for all

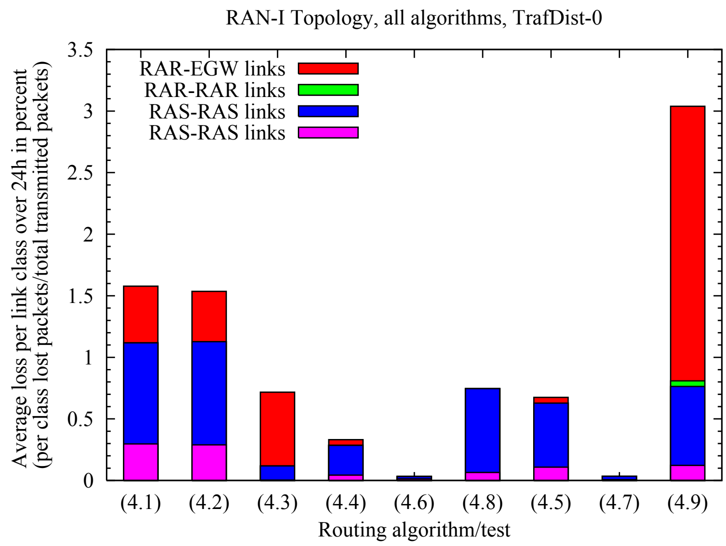


Figure 36: Average loss of all algorithms per link class over 24 h. The per class lost packets are normalized to the total number of packets transmitted. We show traffic distributions TrafDist-0.

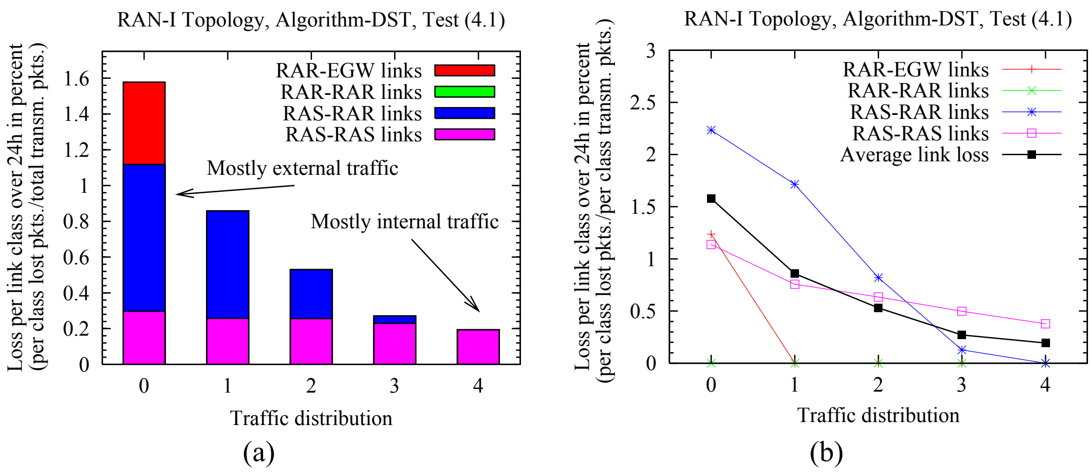
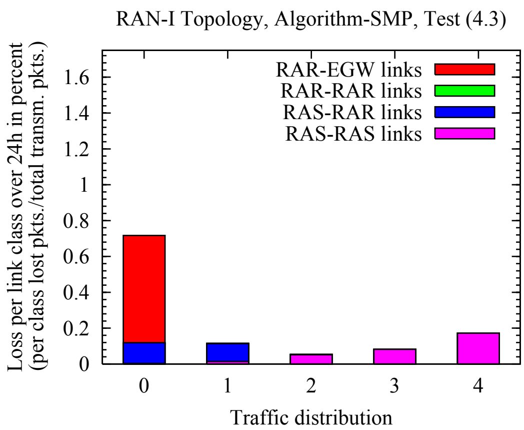
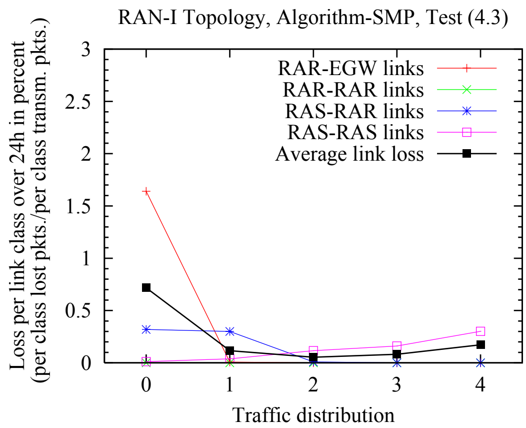


Figure 37: Average loss of Algorithm-*DST* per link class over 24 h. The loss is normalized to (a) the total number of packets transmitted in the network and (b) the number of packets transmitted per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.

traffic distributions for Algorithm-*DST*. The most severe losses are induced on the RAS-RAR links and the RAR-EGW links, though. In contrast, the results for Algorithm-*SMP* in Figure 38 show only minor losses on the downlinks (RAS-RAR) and on the gateway links (RAR-EGW). The losses occur because the equal cost criterion limits the capability of traffic distribution. Hence the traffic to/from the EGW overflows the available multipaths. However, the loss rate of Algorithm-*SMP* is much better compared with Algorithm-*DST*. The results for Algorithm-*DMP*1 shown in Figure 39 are very promising for TrafDist-0 and TrafDist-1. Only few packets are dropped despite the fact that the network load is much higher. However, this behavior changes if the traffic distribution includes higher internal traffic.

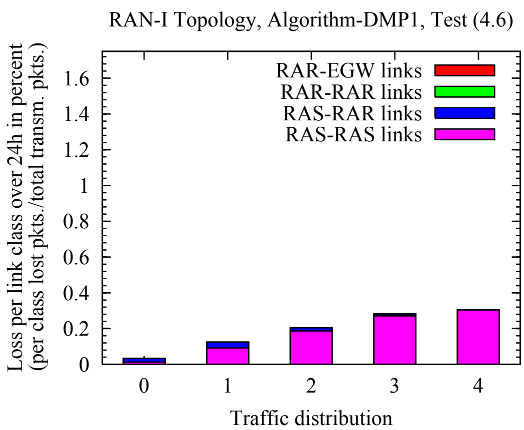


(a)

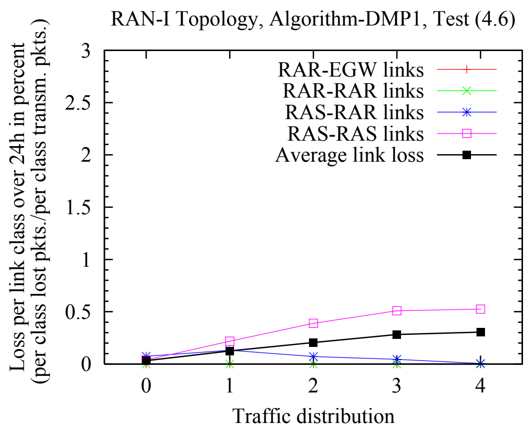


(b)

Figure 38: Average loss of Algorithm-*SMP* per link class over 24 h. The loss is normalized to (a) the total number of packets transmitted in the network and (b) the number of packets transmitted per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.



(a)



(b)

Figure 39: Average loss of Algorithm-*DMP*1 per link class over 24 h. The loss is normalized to (a) the total number of packets transmitted in the network and (b) the number of packets transmitted per link class. All traffic distributions TrafDist-0 to TrafDist-4 are shown. Please note the different y-axis.

Geographical visualization of load and loss

Besides the aforementioned values, we are interested in finding the problem zones of our network. It is obvious that the hot spots of our workload model are correlated to these areas. However, we observed interesting effects on how the different routing algorithms deal with these hot spots. We include Algorithm-*DST* in Figure 40, Algorithm-*SMP* in Figure 41, and Algorithm-*DMP*1 in Figure 42 of our analysis. The plots distinguish between (a) tier-2/tier-3 and (b) tier-2 to allow for better visualization. The busy hour plots emphasize the observed losses and point out different degrees of loss for the depicted algorithms. Moreover, hot spots are clearly visible. The losses for Algorithm-*DST* (single-path) are induced from traffic, which flows from the gateway to the center of the city (see Figure 40). In addition to the large losses on RAR-EGW level and RAS-RAR level, there are minor losses at the RAS-RAS level. Algorithm-*SMP* (multipath) is able to split the traffic and only a small fraction of packets is lost (see Figure 41). Moreover, the load is distributed more evenly resulting in a better link utilization. Algorithm-*DMP*1 (see Figure 42) shows a behavior even better than Algorithm-*SMP*. We can conclude that in our scenario even simple multipath algorithms are able to achieve a better load distribution than standard Internet routing.

Figure 43 and Figure 44 illustrate the second tier for selected algorithms and traffic distributions. We show the load/loss visualization for Algorithm-*SMP* in Figure 43 (a) and Algorithm-*DMP*1 in Figure 43 (b) for TrafDist-4 during the busy hour. We observe that Algorithm-*SMP* is able to achieve a higher load with smaller loss rate than Algorithm-*DMP*1. This behavior can be explained with oscillations of Algorithm-*DMP*1: during the monitored time-interval, periods of high loss alternate with periods of low loss, thus, decreasing the overall load. We explain the source of the oscillations in more detail below. Figure 44 (a) and (b) show the worst case behavior observed for Algorithm-*DLY*

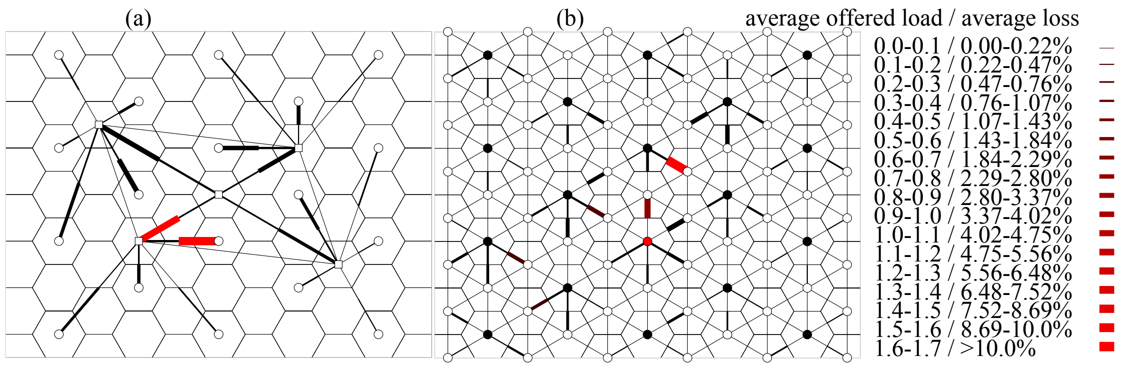


Figure 40: Load/loss visualization for Algorithm-*DST*, TrafDist-0. We visualize the traffic on (a) the links on RAR-EGW level and (b) the links on RAS-RAS level for the busy hour.
Legend: the visualizations have to be interpreted as follows: the width of the edges entering a node depicts the average load offered to the link in direction of the node. The average load is given in relation to the link capacity. The direction is important to model asymmetric traffic. We estimate, for example, more gateway to host traffic than vice versa. The brightness of the edges marks the average link loss. The brighter the red, the more loss is measured.

and Algorithm-*WMP*2, respectively. The shortest delay paths chosen by Algorithm-*DLY* are only suboptimal, because they cannot handle the high load offered. Oscillations are again visible for the Algorithm-*WMP*2 in the mismatch of the loss compared with the load of the links.

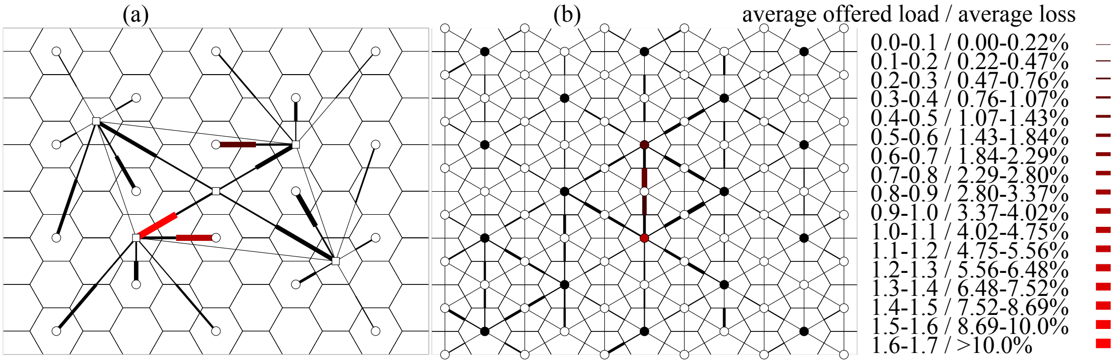


Figure 41: Load/loss visualization for Algorithm-*SMP*, TrafDist-0. We visualize the traffic on (a) the links on RAR-EGW level and (b) the links on RAS-RAS level for the busy hour.

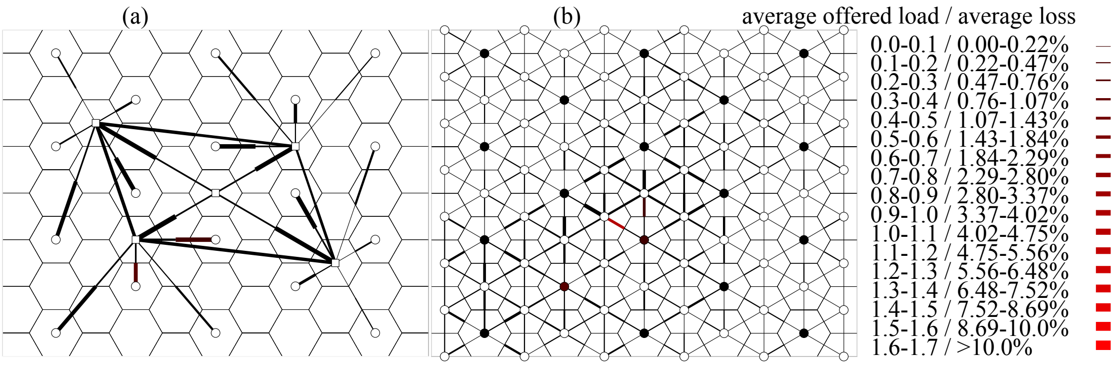


Figure 42: Load/loss visualization for Algorithm-*DMP*1, TrafDist-0. We visualize the traffic on (a) the links on RAR-EGW level and (b) the links on RAS-RAS level for the busy hour.

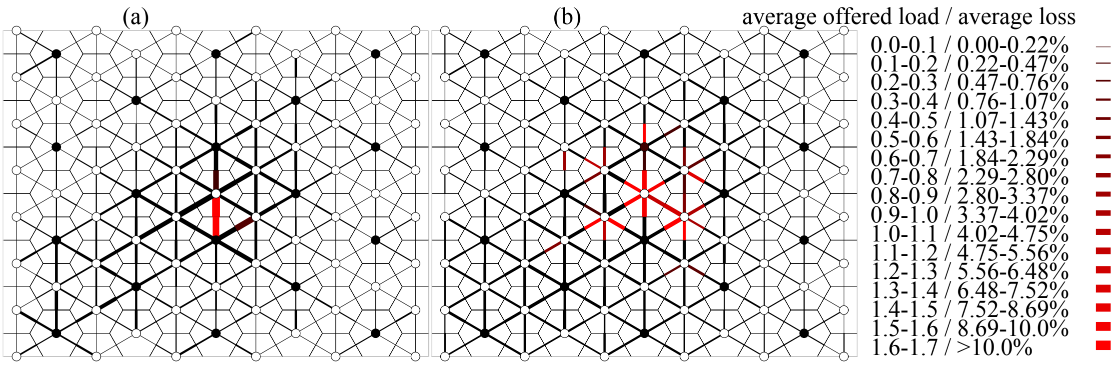


Figure 43: Load/loss visualization for Algorithm-*SMP* and Algorithm-*DMP*1, TrafDist-4. We show the traffic on RAS-RAS level links for (a) Algorithm-*SMP* and (b) Algorithm-*DMP*1 in the busy hour.

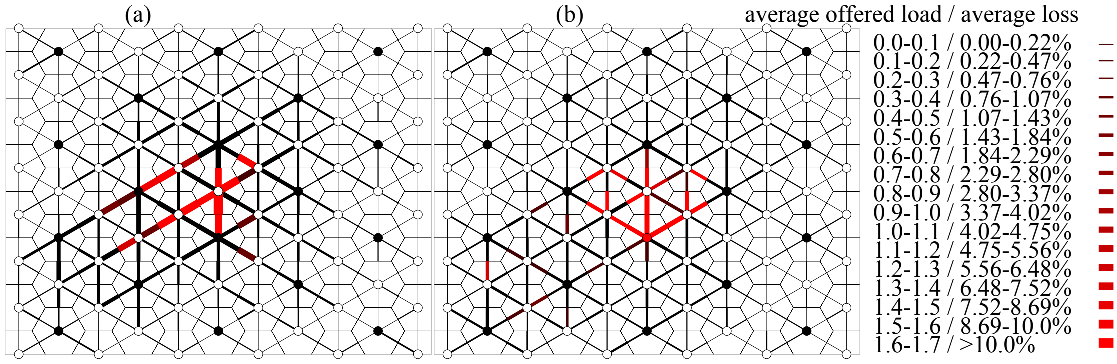


Figure 44: Load/loss visualization for Algorithm-*DLY* and Algorithm-*WMP*2, TrafDist-4. We show the traffic on RAS-RAS level links for (a) Algorithm-*DLY* and (b) Algorithm-*WMP*2 in the busy hour.

4.5.2 Results on Flow Level

Some response variables including delay and path length can only be measured for individual flows, thus, we give these results on flow level. The end-to-end delay acts as a response variable describing the ability of the algorithm to balance the network load. We expect the delay to remain low even under heavy load if the load is well balanced. The path length is an indicator for the characteristics of the multipath algorithms. The flow level results account for the average treatment of flows throughout the entire network while load and loss on link level (see above) account for the treatment of all flows in the corresponding tier of the network.

In the following, we regard the path length and delay per hop on flow level. Figure 45 shows the average path length for all algorithms in hops. Figure 46 gives the results for the delay per hop in milliseconds. The increased path length of the Algorithm-*DMP* and Algorithm-*WMP* variants are due to the usage of longer routes (in terms of

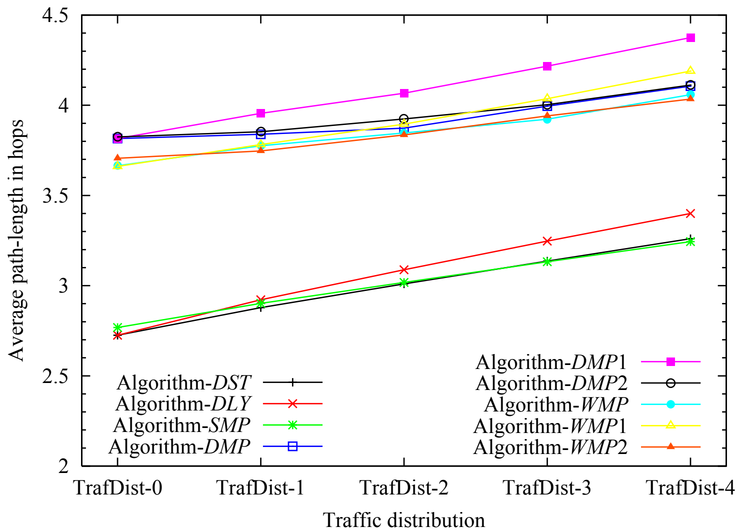


Figure 45: Average path length for all algorithms over 24 h. All traffic distributions are shown.

hops but not necessarily in terms of delay) for load-balancing. On the other hand, the normalized delay per hop is smaller for *Algorithm-DMP*1 and Algorithm-*WMP*1 compared to the single-path algorithms. Here, the delay trend along the traffic distributions is also of interest. This reduction in normalized delay per hop shows that these algorithms are able to make use of the alternate paths. The figures show the measurements for all traffic distributions. See Appendix A for additional results on flow level.

4.5.3 Results on Packet and Application Level

The trace of the treatment of individual packets allows to give results on packet level. We injected some measurement flows into the network, which were closely monitored. The flows were injected in both directions between measurement points to account for the asymmetric behavior because of non-symmetric traffic distributions. From a total of eight measurement flows, which were present in each run of the experiment, we only provide the most important insights while all traces can be found in [Kro02]. The individual measurement flows have been set up as constant bitrate flows using UDP, 128 kbit/s each. In particular, we select four traces of interest. The first selected trace captures the flow from the edge gateway to a RAS located in the center of the network (EGW—RAS53), see also Figure 30 for the corresponding locations in the RAN-I topology. We expect the results to be of interest, because during the busy hour this area of the network is under heavy load. The second closely monitored flow starts from RAS83 and is destined to RAS43 (RAS83—RAS43). This flow crosses the backbone if routed on the shortest path or may use diverse routes if the backbone is under heavy load conditions for delay sensitive algorithms. We present these traces for a selected subset of the algorithms and selected traffic distributions. Special behavior of some algorithms is illustrated using selected traces.

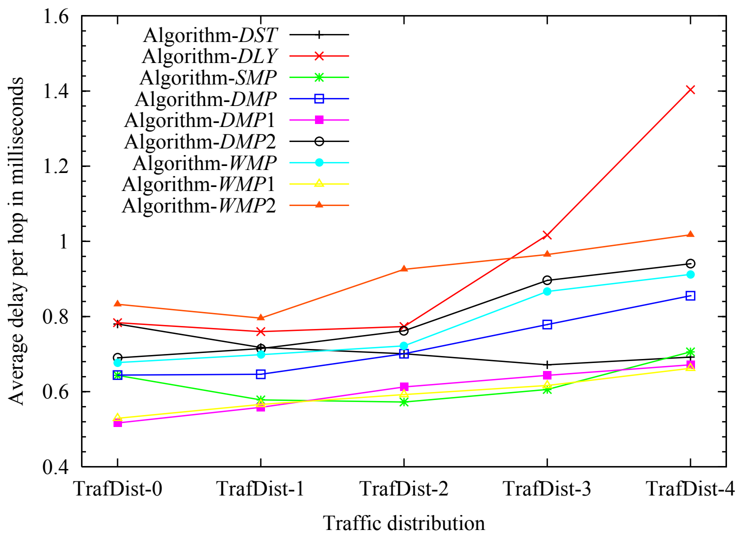


Figure 46: Average delay per hop for all algorithms over 24 h. All traffic distributions are shown.

Delay of Packets for an Edge Gateway to City Center Connection

Because of the modeled traffic distribution, we expect the connection from the edge gateway to the city center as primary bottleneck. Here, the asymmetry of the traffic is mainly caused by the transactional web traffic, which is expected to yield much bigger responses than queries. Moreover, this traffic mostly departs at the edge gateway, which acts as gateway to the Internet from a simulation perspective. We have chosen to represent this connection using the measurement flow between the EGW and RAS53. The results on packet level are shown in Figure 47 to Figure 51 for traffic distribution TrafDist-0 and for the routing strategies Algorithm-*DST*, Algorithm-*SMP*, Algorithm-*DMP*1, Algorithm-*DMP*2, and Algorithm-*WMP*1. All scatter-plots show an increase in delay for the high load situations during noon (see also Figure 15 above, which shows the overall workload for the simulation). The algorithms differ significantly in the length of the time period during which the delay is heightened. Moreover, they show a different behavior during this time period, which is visible in the shape of the scatter-plot. It should be noted that different scales are used on the y-axis for the individual algorithms. We describe the observed algorithm behavior in detail in the following.

Algorithm-*DST* in Figure 47 shows some increases in delay if the path chosen is congested. Between 9:00 h and 17:00 h this increase is clearly visible. The delay for Algorithm-*DST* rises as high as 19ms and we see some packet drops during the period of high delay. In contrast, Algorithm-*SMP* in Figure 48 constantly remains below 8ms because of the multiple paths in use. We see two different phases of algorithm behavior: from 8:00 h to 10:00 h the delay is slightly higher than during the low workload situations, from 10:00 h to 17:00 h the increase in delay is higher and some loss occurs. Algorithm-*DMP*1 shows a good performance most of the time, the maximum delay being around 8 ms (see Figure 49). However, there are few short periods where the algorithm fails, and as a result some losses and an increasing delay up to around 20 ms can be observed. The periods of high delay and loss are due to the oscillatory behavior of the algorithm, which is investigated in detail below. For some periods, the delays of the packets appear grouped together and form parallel horizontal lines. This grouping illus-

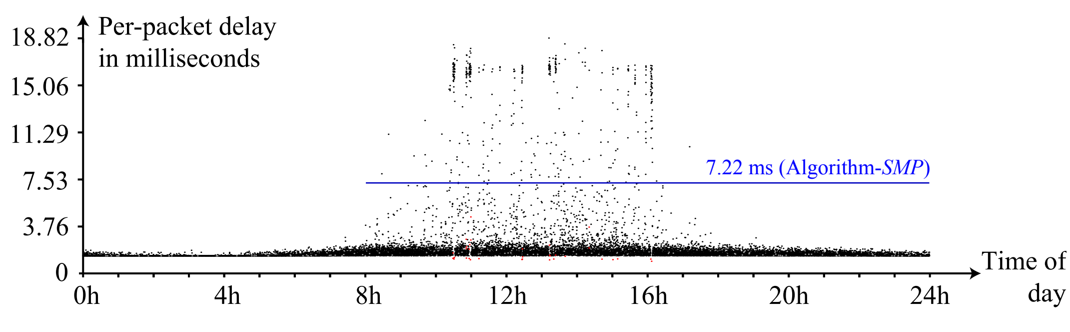


Figure 47: Per-packet delay for Algorithm-*DST* from EGW-RAS53 over 24 h, TrafDist-0.
Legend: the scatter-plots give the end-to-end delay for the individual packets in black. If packets are dropped due to congestion, the time delay until their drop is counted and the point is drawn in red.

trates the different delays of the multiple paths computed by Algorithm-*DMP*1. The per-packet delay of Algorithm-*DMP*2 is shown in Figure 50. We see similar parallel horizontal lines. The oscillations of the algorithm are clearly visible. Periods of good delay/loss ratio are followed by periods where loss occurs and the achieved delay increases significantly. With the current parametrization, this variant of the routing algorithm does not provide a stable performance. The delay rises as high as 34 ms. Moreover, the packet loss is very high between 8:00 h and 19:00 h for the observed flow. Figure 51 shows Algorithm-*WMP*1. Again the multipaths are visible. The delay observed for Algorithm-*WMP*1 remains below 8 ms most of the time. However, between 12:00 h and 15:00 h there are also some visible oscillations. Nevertheless, the algorithm is able to gracefully support the high load in the network with only few losses.

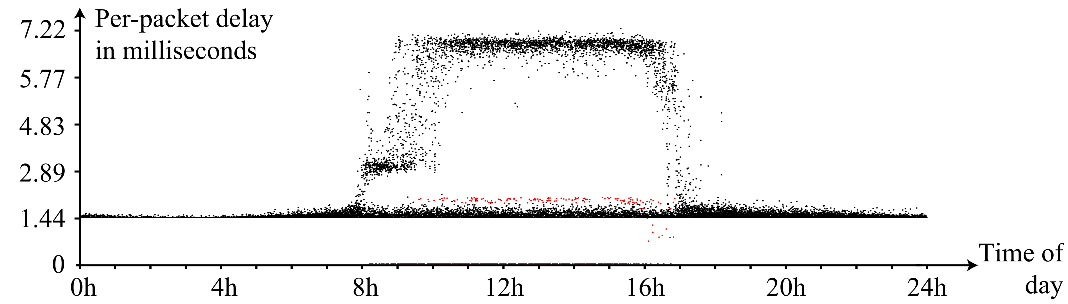


Figure 48: Per-packet delay for Algorithm-*SMP* from EGW-RAS53 over 24 h, TrafDist-0.

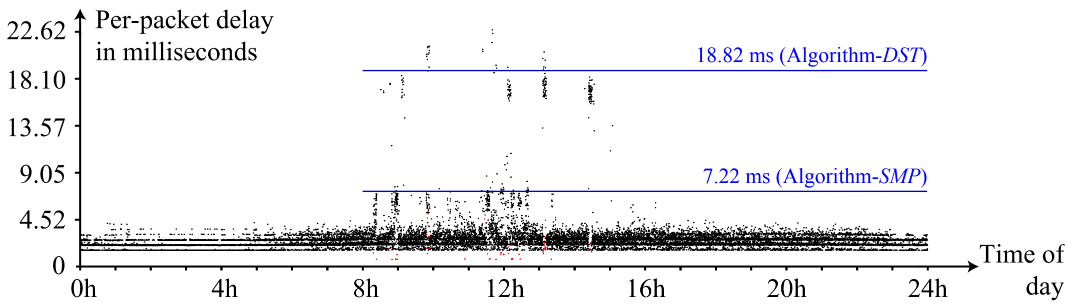


Figure 49: Per-packet delay for Algorithm-*DMP*1 from EGW-RAS53 over 24 h, TrafDist-0.

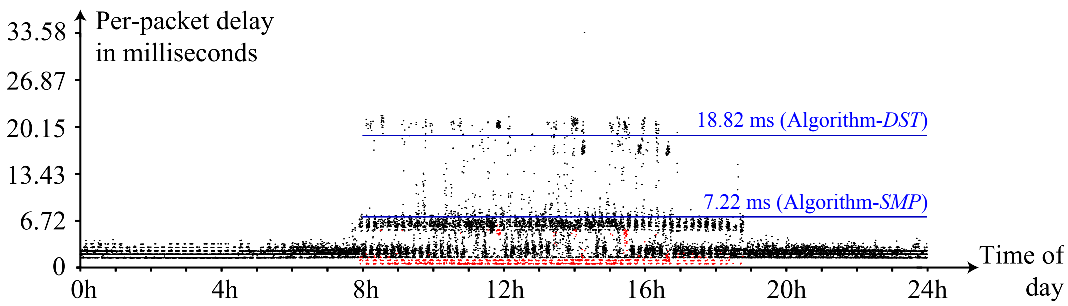


Figure 50: Per-packet delay for Algorithm-*DMP*2 from EGW-RAS53 over 24 h, TrafDist-0.

For traffic distribution TrafDist-2 we depict the behavior of Algorithm-*DST*, Algorithm-*SMP*, and Algorithm-*DMP*1. Because of a higher amount of localized traffic, the delay for the investigated flow is reduced. This can be seen for Algorithm-*DST* in Figure 52 and Algorithm-*SMP* in Figure 53, which show reduced delays compared to TrafDist-0. However, Algorithm-*DMP*1 in Figure 54 shows only a poor behavior. The oscillations lead to some periods where unsuitable paths of high loss and extreme delay are chosen, while during the other periods the loss and delay look very promising. Finally, we illustrate the behavior of Algorithm-*SMP* for traffic distribution TrafDist-4 in Figure 55. The scatter-plot clearly shows smaller delay compared to Figure 48 due to the shift from EGW traffic to localized traffic, which effectively reduces the load on the monitored flow. In Figure 55 the delay does only increase marginally even for the times of high load, the maximum delay being around 4 ms for the investigated scenario.

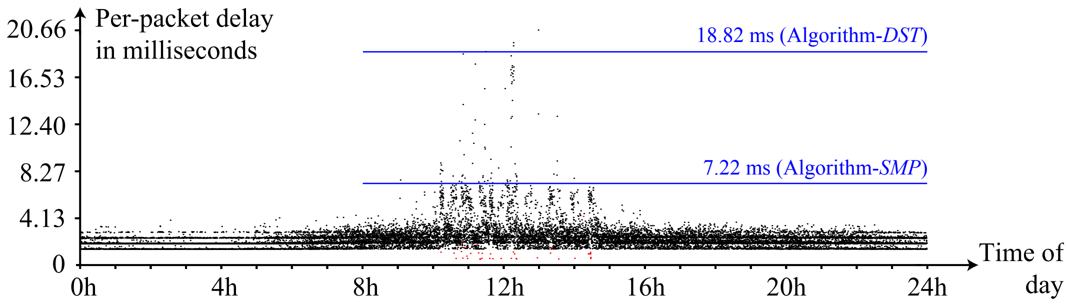


Figure 51: Per-packet delay for Algorithm-*WMP*1 from EGW-RAS53 over 24 h, TrafDist-0.

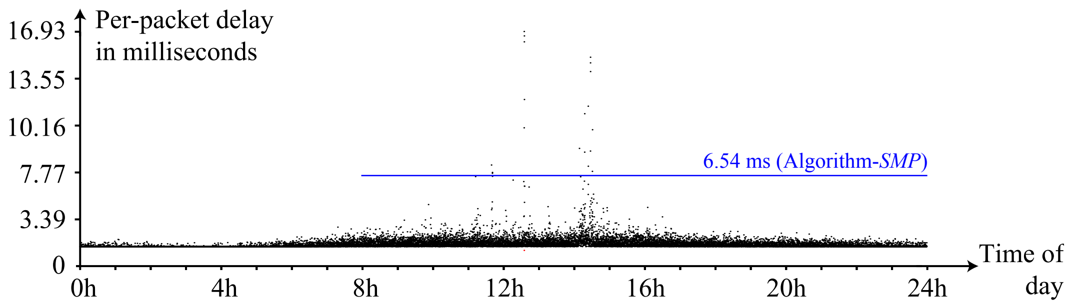


Figure 52: Per-packet delay for Algorithm-*DST* from EGW-RAS53 over 24 h, TrafDist-2.

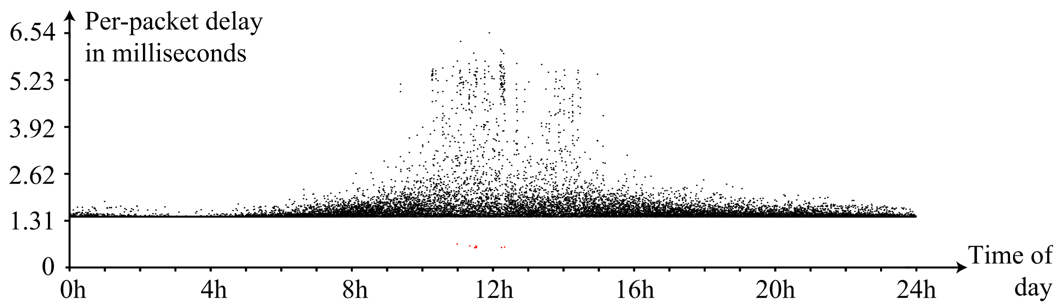


Figure 53: Per-packet delay for Algorithm-*SMP* from EGW-RAS53 over 24 h, TrafDist-2.

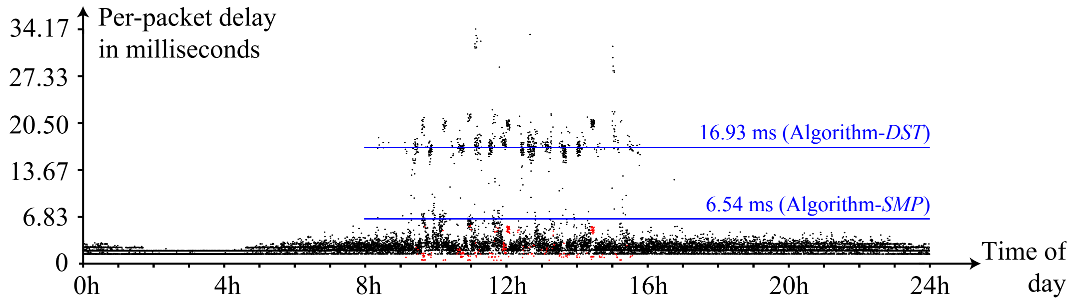


Figure 54: Per-packet delay for Algorithm-*DMP*1 from EGW-RAS53 over 24 h, TrafDist-2.

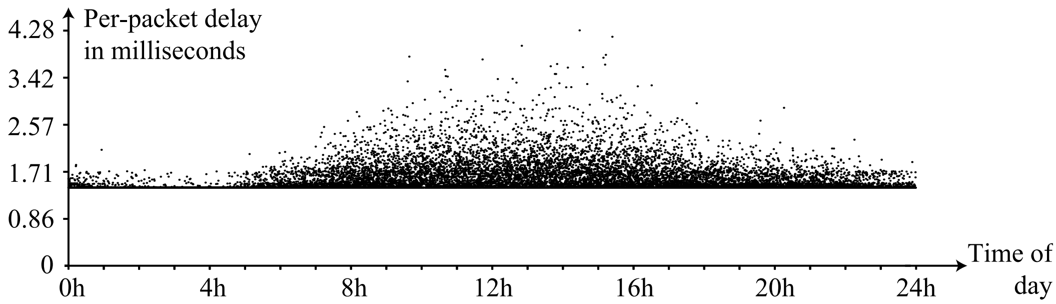


Figure 55: Per-packet delay for Algorithm-*SMP* from EGW-RAS53 over 24 h, TrafDist-4.

Delay of packets, RAS83 to RAS43

The results for the measurement flow between RAS83 and RAS43 are of great interest because the shortest path in terms of hops goes via the backbone while the shortest path in terms of delay goes directly through RAS-RAS level links. We depict the behavior of Algorithm-*DST*, Algorithm-*SMP*, and Algorithm-*DMP*1 for TrafDist-0 and TrafDist-2.

See Figure 56 for the end-to-end delay for Algorithm-*DST* using TrafDist-0. We see a stepwise increase of delay, which is due to the different tiers of the topology. We also see two different levels on which loss occurs. The loss increases and the delay is as high as 21 ms. The rise is relatively abrupt and the decline is smoother. This is due to the traffic load, which is similarly shaped. In summary, the single-path behavior of Algorithm-*DST* leads to problems if the backbone is under heavy load. Algorithm-*SMP* has only a shorter period of time where the increase in delay and loss is heightened (see Figure 57). The increase in delay starts approximately 2 h later in time and the decrease around 3 h earlier in time as observed with Algorithm-*DST* (see Figure 56). This shows the much better handling of congestion using multipath algorithms, which are able to split the traffic. The overall performance of Algorithm-*DMP*1 shown in Figure 58 is quite promising. Delay as well as loss do only marginally increase even under high link loads. Moreover, the we observe only minor losses. However, due to the oscillations of the algorithms, we see again some periods of time with only suboptimal paths.

For TrafDist-2 with a higher amount of local traffic, the measured end-to-end delays between RAS83 and RAS43 for Algorithm-*DST* shown in Figure 59 and Algorithm-*SMP* shown in Figure 60 behave like expected. Again the increase in loss and delay occurs later for the multipath algorithm compared with the single-path algorithm. While the increase in delay and loss of Algorithm-*DMP*1 as depicted in Figure 61 occurs approximately during the same time-period as for Algorithm-*SMP*, the algorithm behavior is dominated by heavy oscillations.

Figure 62 shows Algorithm-*DST* for TrafDist-4 for the reverse direction RAS43 to RAS83. The increase as well as the decrease of delay is much smoother because the offered load does not rise that sharp. One interesting phenomena can be observed for the packet-loss, which takes place directly at the source node RAS43. This indicates that the offered load exceeds the bandwidth available.

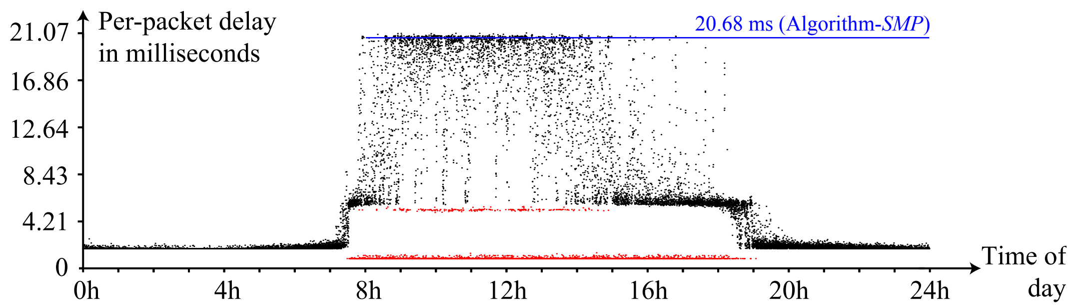


Figure 56: Per-packet delay for Algorithm-*DST* from RAS83-RAS43 over 24 h, TrafDist-0.

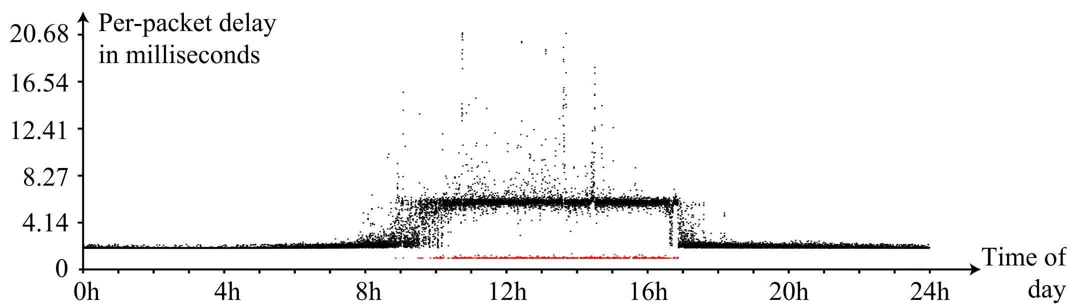


Figure 57: Per-packet delay for Algorithm-*SMP* from RAS83-RAS43 over 24 h, TrafDist-0.

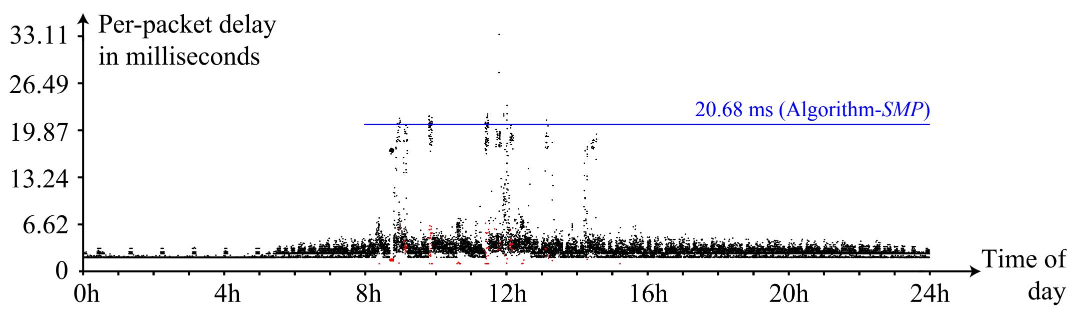


Figure 58: Per-packet delay for Algorithm-*DMP*1 from RAS83-RAS43 over 24 h, TrafDist-0.

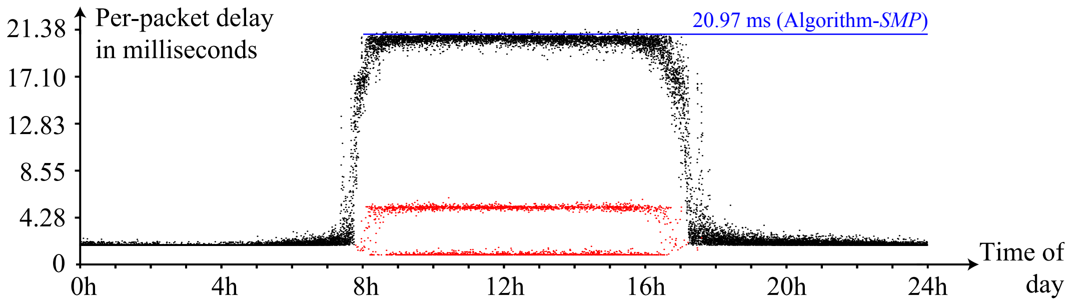


Figure 59: Per-packet delay for Algorithm-*DST* from RAS83-RAS43 over 24 h, TrafDist-2.

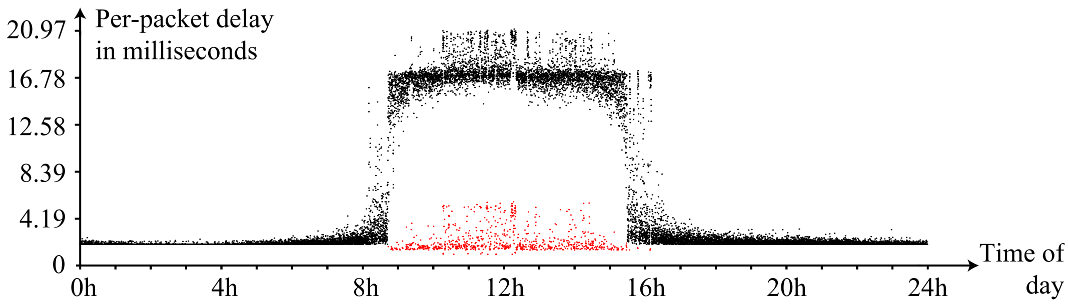


Figure 60: Per-packet delay for Algorithm-*SMP* from RAS83-RAS43 over 24 h, TrafDist-2.

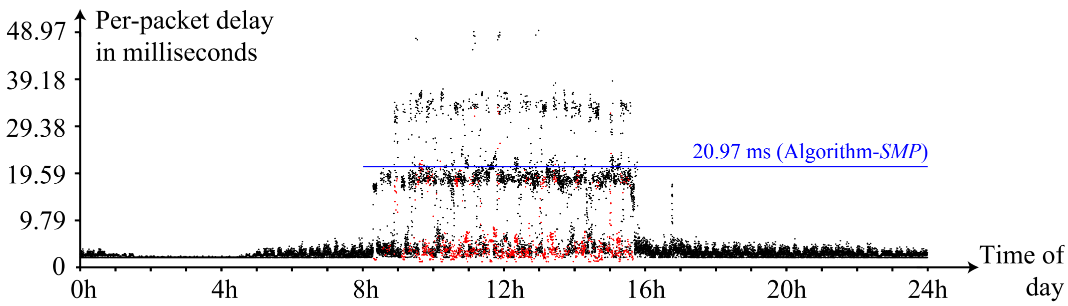


Figure 61: Per-packet delay for Algorithm-*DMP*1 from RAS83-RAS43 over 24 h, TrafDist-2.

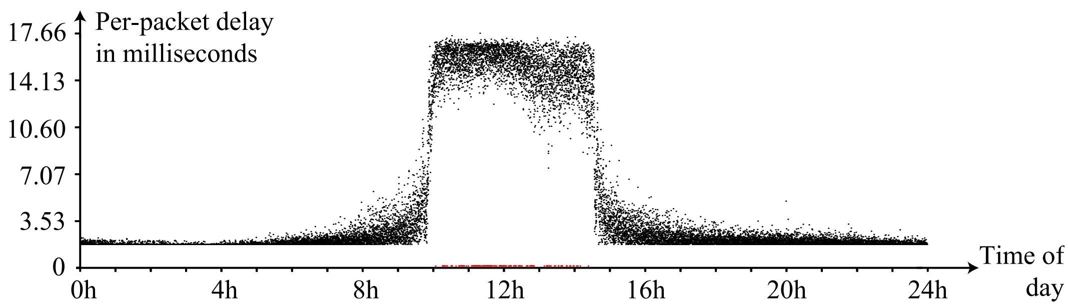


Figure 62: Per-packet delay for Algorithm-*DST* from RAS43-RAS83 over 24 h, TrafDist-4.

To further illustrate the oscillating behavior observed for the class of dynamic multipath algorithms, Figure 63 shows the busy hour for Algorithm-*WMP*2, which exhibits the worst oscillations among the tested algorithms. The oscillations are clearly visible:

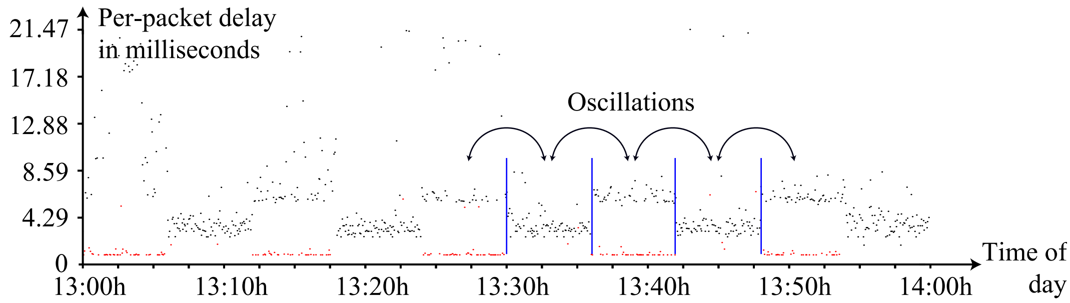


Figure 63: Oscillations in the per-packet delay from RAS83-RAS43 for Algorithm-*WMP*2. The measurement shows the busy hour for traffic distribution TrafDist-0.

periods of good performance (moderate loss, low delay) alternate with periods of bad performance (heightened loss, high delay). The boundaries between these periods are determined by our re-routing interval of 6 minutes. We refer to Section 5.4.4 below for a precise description of Vutukury's algorithm. In short, the oscillations observed are a consequence of the periodic computation of the routing graph. Here, the algorithm removes fully congested links, even if these are part of the optimal path under normal load conditions. At the next re-routing interval, the now nearly empty paths are included in the routing graph again. Only after finishing the experimentation and analysis of the results, we realized that a simplification of our implementation of Vutukury's algorithm contributed to the observed oscillations under extreme traffic-load conditions, additionally. To keep the complexity of the route computation reasonable, we omitted the mechanism for load adjustments at each interval t_s (short interval) between the re-routing intervals t_l (long interval) as specified from Vutukury et al. in [VGLA99]. We are studying the influence of this simplification in the next chapter in more detail. Nevertheless, even with the implementation of the t_s mechanism, the algorithm removes nodes out of the successor set under overload conditions while reconsidering them in the next period. In the next chapter we also discuss some mechanisms to avoid these oscillations.

4.5.4 Summary of Results

Table 5 gives a summary of the overall results obtained for the fixed *e/i-ratio* of 100/0 for all classes (TrafDist-0) while Table 6 summarizes the results for TrafDist-2 with varying *e/i-ratio*. The overall performance of the network can be approached using the application related parameters of average loss and average end-to-end delay over all flows traversing the network. Additional insights on the operation of the algorithms can be obtained using the routing related parameters of average load on different tiers of the network design.

For TrafDist-0 the multipath algorithms Algorithm-*SMP*, Algorithm-*DMP*1, and Algorithm-*WMP*1 are able to achieve a much better delay compared to the single-path algorithms while minimizing the packet loss. Of particular interest is the performance of the dynamic algorithms: only with the variant using the smoothed delay as link cost

Table 5: Summary of Results for Test-series (4.*); RAN-I Topology, TrafDist-0.

Algorithm	Application related results		Routing related results		
	Average end-to-end delay (σ)	**Average loss**	**Path length**	**Average load RAS-RAS (CoV)**	**Average load RAS-RAR (CoV)**
Algorithm-*DST*	2.13 ms (0.0027)	4.42%	2.73	4.65% (0,424)	20.59% (0,048)
Algorithm-*DLY*	2.14 ms (0.0028)	4.13%	2.73	4.68% (0,422)	20.60% (0,048)
Algorithm-*SMP*	1.78 ms (0.0013)	1.96%	2.77	5.12% (0,296)	21.05% (0,044)
Algorithm-*DMP*	2.46 ms (0.0022)	1.26%	3.82	8.39% (0,148)	20.45% (0,039)
Algorithm-*DMP*1	1.97 ms (0.0010)	0.12%	3.82	8.49% (0,146)	20.52% (0,039)
Algorithm-*DMP*2	2.64 ms (0.0026)	2.83%	3.82	8.51% (0,158)	20.62% (0,040)
Algorithm-*WMP*	2.48 ms (0.0026)	2.46%	3.66	7.74% (0,156)	20.43% (0,039
Algorithm-*WMP*1	1.94 ms (0.0010)	0.13%	3.66	8.07% (0,162)	20.59% (0,039)
Algorithm-*WMP*2	3.09 ms (0.0030)	10.52%	3.71	7.27% (0,166)	18.99% (0,038)

Table 6: Summary of Results for Test-series (4.*); RAN-I Topology, TrafDist-2.

Algorithm	Application related results		Routing related results		
	Average end-to-end delay (σ)	**Average loss**	**Path length**	**Average load RAS-RAS (CoV)**	**Average load RAS-RAR (CoV)**
Algorithm-*DST*	2.11 ms (0.0028)	1.59%	3.01	8.42% (0.246)	30.09% (0,033)
Algorithm-*DLY*	2.39 ms (0.0033)	1.60%	3.09	11.80% (0.209)	23.96% (0.042)
Algorithm-*SMP*	1.73 ms (0.0017)	0.16%	3.02	9.36% (0.194)	29.73% (0.034)
Algorithm-*DMP*	2.71 ms (0.0033)	1.12%	3.87	12.03% (0.133)	34.61% (0.029)
Algorithm-*DMP*1	2.49 ms (0.0029)	0.83%	4.07	13.31% (0.109)	36.12% (0.028)
Algorithm-*DMP*2	2.99 ms (0.0039)	1.41%	3.92	12.40% (0.134)	35.05% (0.029)
Algorithm-*WMP*	2.78 ms (0.0034)	2.07%	3.85	12.12% (0.125)	35.25% (0.028)
Algorithm-*WMP*1	2.31 ms (0.0024)	0.76%	3.90	12.43% (0.117)	35.76% (0.028)
Algorithm-*WMP*2	3.55 ms (0.0049)	3.25%	3.84	11.95% (0.134)	35.82% (0.028)

metric, these algorithms are able to deliver a steady performance. The global delay optimization of Vutukury's algorithm is visible in the average end-to-end delay as well as the standard deviation of the end-to-end delay for Algorithm-*DMP*1 and Algorithm-*WMP*1. The routing related values include the average path length, which is of particular interest for the multipath algorithms. For the equal cost algorithm we see nearly the same average path length compared to the single-path variants. The slight increase is due to the nature of our network, which offers more equal cost paths for longer distances. The average path length of the dynamic multipath algorithms shows a significant increase compared to the static shortest-path algorithms. This behavior is due to the optimization of delay, which may lead to longer paths in terms of hop counts.

As routing related response variables we also show the load in combination with the load variation for different hierarchy levels. The coefficient of variance (CoV) normalizes the standard deviation of the variable to its average and allows for a direct comparison of the achieved quality of load-balancing for the algorithms if applied to the individual loads measured. We see a smaller CoV of load for the multipath algorithms and especially for the dynamic multipath variants on RAS-RAS level. This behavior indicates that the global optimization of delay in fact optimizes the load distribution in the network by utilization of the RAS-RAS links introduced in the MobQoS architecture. For TrafDist-0, Algorithm-*WMP*2 shows a very poor behavior: the loss rate is as high as 10.52% caused by oscillations of the algorithms. For the packets not lost, Algorithm-*WMP*2 is able to achieve a load-balancing, which can be compared to the other multipath algorithms, though.

The overall picture changes for traffic distribution TrafDist-2 shown in Table 6. Algorithm-*SMP* is still able to show an excellent performance in terms of delay and loss. Moreover, the average load on RAS-RAS level is well balanced. The single-path algorithms yield a better delay but a higher loss compared to the dynamic multipath algorithms, Algorithm-*DMP*1 and Algorithm-*WMP*1. A comparison of Algorithm-*DMP*1 and Algorithm-*WMP*1 with Algorithm-*SMP* shows, however, that the static algorithm provides for better loss and delay. Again the path length is significantly increased for the dynamic multipath algorithms. The load on RAS-RAR level has a comparable CoV for all algorithms, the load of the dynamic multipath algorithms being slightly increased.

4.6 Related Work

The survey of directly related work for this chapter includes related work on architectures for beyond 3G networks on one hand and related work in the area of QoS routing on the other hand. For the former, we investigate architectures, which introduce alternate paths and/or decentralized control into the radio access part of the network. For the latter, we give a broad overview of QoS routing in general and highlight some algorithms that are closely related to the Internet routing paradigm while allowing for load-balancing to achieve gains in network performance.

Cellular Network Architectures

The standardization of 3rd generation cellular networks is an evolutionary process. The 3GPP organization [3GPP04] as well as vendors and network operators drive this process. As of now, most of the proposals for future architectures stick to the paradigm of a non-routed radio access network [Usk03] [NSA+04], which is in contrast to our assumptions. However, similar to our work, these recent proposals depart from strictly hierarchical and tree-structured topologies, and allow for interconnections at the edge of the network to introduce flexibility in resource management. The European project MIND [MBH+02] [MLM+02] aims at pure IP radio access networks. MIND has many concepts in common with our work but differs slightly in the research focus. In particular, the two main goals addressed in [MBH+02] and [MLM+02] are mobility support and QoS. The former goal is addressed by enhancing Mobile IP to support fast and seamless handovers. The latter goal is pursued with well known mechanisms from infrastructure-based networks, namely DiffServ [BBC+98] and IntServ [BCS94]. This is in contrast to our work, where we do not introduce explicit QoS mechanisms but consider the distributed routing mechanisms themselves as the solver for the QoS problem.

QoS Routing

The area of QoS routing has attracted a lot of research over time. Traditionally, QoS routing is applied to find constrained paths within the context of fine-grained resource reservation. The constraints may be of additive, concave, or multiplicative type as described by Wang and Crowcroft [WC96]. The complexity of finding a feasible path through the network depends on the number of constraints and their type as well as on the nature of the algorithm (centralized, decentralized, hierarchical). Constraints may be, for example, delay, capacity, jitter and loss-ratio. In [WC96] it was shown that the problem of finding feasible paths with two independent types of constraints is NP-complete. Current QoS routing strategies can be divided into source routing, distributed routing, and hierarchical routing. An overview to the field of constraint-based routing and QoS routing can be found in Chen and Nahrstedt [CN98] as well as Paul and Raghavan [PR02]. For an algorithmic treatment of constraint-based routing algorithms see Kuipers et al. [KKKM02], where the authors simulated several algorithms in order to determine their worst case complexity. Last, Sobrinho [Sob01] gives an algebra for QoS path computation and describes hop-by-hop QoS routing mechanisms including multipath algorithms as well.

The use of QoS routing mechanisms for load-balancing has been discussed for multiple protocols. However, most of the work was done for source routing algorithms such as the Q-OSPF protocol by Apostolopoulos et al. [AWK+99]. Since Q-OSPF was designed for QoS routing in the sense of finding hard-QoS constrained paths, the use for resource management/traffic engineering inherits some restrictive assumptions. This includes the operation on flow-granularity and the need for a surrounding framework for admission and reservation of resources. There are other algorithms such as “shortest path first” algorithms optimized for computation of minimum congestion paths. Such algorithms, as given in Park et al. [PSL01], are limited by the source routing paradigm, too. In summary, source routing strategies are well-suited within reservation-based or connection-oriented systems. However, they impose unnecessary and unwanted complexity in connectionless systems that rely on hop-by-hop operation, because intermediate systems need to keep per-flow state in the network.

If we regard connection-oriented networks, it is quite natural to use multiple paths (one per flow) between source-destination pairs and, thus, follow a multipath paradigm. In our context, namely connectionless networks, the class of multipath routing algorithms is often overlooked, however. The use of multiple paths at the same time allows for dispersion of the traffic on different granularity and, thus, alternate paths can help improving network performance significantly. Work in the area of traffic dispersion on packet or sub-packet level range from the early work of Maxemchuk [Max75] to more recent approaches. Examples for multipath routing algorithms include works by Moy [Moy98], Villamizar [Vil02], and Vutukury et al. [VGLA99] [VGLA01]. Hereby, the first three are based on a link state protocol and operate on global imprecise network state. The last algorithm is a distance vector based algorithm operating on information provided from neighboring nodes.

The OSPF variants ECMP (equal-cost multipath) described in [Moy98] and OMP (optimized multipath) described in [Vil02] are heuristics, which are targeted to better distribute load over multiple paths. OSPF-ECMP, for example, distributes the load over multiple equal-cost paths using simple round robin mechanisms and limits the possible load-balancing to this restricted set of paths. The equal distribution among these paths may result in non-optimal performance, too. OSPF-OMP optimizes this distribution using heuristics to predict which path to use. The necessary load information of the network is hereby distributed using OSPF opaque link state advertisements. In contrast to the intuitive heuristics given in [Moy98] and [Vil02], the recent work from Vutukury and Garcia-Luna-Aceves approaches the problem from a more theoretical perspective. Based on Gallager's optimal routing [Gal77] and early work of Cantor and Gerla [CG74], they formulate the properties a distributed multipath routing protocol must follow to achieve near-optimal routing. The application for traffic engineering is described in [Vut01], too. For multipath algorithms it is crucial to efficiently deal with routing loops, because the number of feasible paths largely depends on the strategy chosen to handle loops. However, it is not necessary for the paths to be disjoint if the algorithm is carefully designed. The analysis of the promising performance properties of multipath algorithms in [Sob01] and the robustness of such algorithms are of high interest.

There exist other QoS routing protocols that do not follow link-state paradigms. For example, in [CN99] Chen and Nahrstedt propose a distributed QoS routing scheme that is based on tickets to probe the available bandwidth or delay in ad hoc networks. Designed for highly dynamic environments, the algorithm is able to deal with variability in topology as well. However, the algorithm operates on flow level and thus does not fulfill our requirements (no per-flow state) for an candidate algorithm for our architecture. Menth in [Men04] developed a "self-protecting" multipath routing scheme in combination with various admission control schemes. The results presented in [Men04] show that the performance of the network can be enhanced without compromising the robustness. However, the algorithms presented in [Men04] are based on circuit-switched paradigms such as MPLS and, thus, cannot directly be compared with our work in the connectionless domain.

What currently misses for all of the aforementioned QoS routing protocols are investigations of their performance within wireless metropolitan area networks, our target scenario. Most of the protocols have been designed and evaluated for use in backbone topologies. However, in our scenario we expect heavily varying loads on different timescales induced by user and device mobility. The network structure is likely to be more regular than in backbone networks while the degree of interconnection is expected to be higher. We are not aware of comparative investigations of QoS routing algorithms within metropolitan radio access networks. This is not surprising, because so far these have not been routing networks. While there exists few work to study large scale wireless LAN topologies, including the work of Tang and Baker [TB02] and the work of Kotz and Essien [KE02], the investigation of underlying network characteristics or specialized routing strategies are beyond the scope of these works.

4.7 Summary

In this chapter, we have presented our work in static infrastructure-based radio access networks. Departing from the shortcomings of today's state-of-the-art cellular networks we have designed a powerful novel network architecture. Moreover, we have evaluated the network performance of various routing strategies within this architecture. By means of a simulation study, we have been able to present a proof-of-concept for our architecture. Moreover, we presented a case for resource optimization using decentralized QoS routing mechanisms.

As a first step, our investigation covered architectures for cellular networks. We surveyed the state-of-the-art in cellular network architectures and identified the most important characteristics and requirements for the radio access part of cellular networks. We found traditional architectures to be mostly coupled with centralized control mechanisms, which allow for easy administration and management on one hand, while, on the other hand, leave only few room for optimization by means of decentralized and distributed mechanisms. Based on the results of Chapter 2, we introduced various design principles and assumptions for the structure of our architecture. We challenged the mainly tree-structured architectures of traditional telecommunication networks and introduced a tightly meshed network design, the so-called MobQoS architecture, which can be seen as an evolutionary next step for static infrastructure-based radio access networks. The architecture is flexible to integrate a heterogeneous set of access technologies and does not rely on central instances for network control and resource management. The MobQoS topology provides for alternate paths beginning in the second tier of the radio access network. Most importantly, we follow a connectionless best-effort routing paradigm and assume the infrastructure components to include IP-routing functionality.

As a second step, we have discussed and studied the impact of various routing strategies to leverage the capabilities of our architecture. We have performed an extensive simulation study using the workload modeled in Chapter 3. We gave a concise description of the experimental design. This included the definition of predictor variables, the most important of which are the routing strategies to compare. The strategies include shortest path routing and delay constrained routing as well as various multipath quality of service routing variants. Moreover, we investigated different traffic distributions. The simulation study served as a proof-of-concept for our MobQoS architecture: we have been able to show that network design as well as resource management for cellular networks can be approached using Internet paradigms. Moreover, the simulation results aided in solving the quest of the network performance aspects of dependable routing. A detailed analysis of the results has been performed to qualify the efficiency in resource management. We have employed the loss and load of individual link classes as well as of the entire network to represent this overall utility. Moreover, we investigated the delay and delay variation of individual streams and packets. Our findings are that multipath routing is able to enhance the utility of the network significantly and allows for efficient resource management within wireless metropolitan area networks. However, we have not been able to tap the full potential of dynamic multipath routing algorithms based on the minimum-delay metric, because the variant chosen caused heavy oscillations under high load situations. Hence, we perceive that there is further room for improvement, which is addressed in Chapter 5, where we leverage the power of a novel variability-augmented cellular architecture with novel routing paradigms.

5 Dependable Routing for Cellular Networks with Variable Infrastructure

We have studied routing dependability in the context of cellular networks with static topology in Chapter 4. Our proposed architecture was able to perform well in combination with advanced routing mechanisms but also showed room for improvement. In this chapter we discuss novel architectural designs as well as algorithmic extensions to tap the full potential of distributed quality of service mechanisms for resource management in cellular networks. We borrow concepts from the domain of variable topology networks to deal with the effects of mobility as described in Chapter 3. These concepts represent a powerful basis for the algorithmic optimization that follows. We propose a novel near-optimal multiclass multipath routing algorithm that builds on existing work in the area of optimal routing. The algorithm is able to operate on top of our variable infrastructure and shows very promising capabilities for the scenario studied. We find performance gains of our architecture and algorithm. They enhance the utility of the network significantly, thus, contributing to the dependability in cellular networks according to our definition in Chapter 2. This chapter concludes our quest for dependable routing in cellular networks. In the next chapters, Chapter 6 and Chapter 7, we move towards entirely variable networks, namely ad hoc networks.

5.1 Motivation

Our findings of Chapter 4 show the importance of the network architecture as well as of the routing strategy for the achievable performance region in the radio access part of cellular networks. Today's mainly static resource management approaches are not well suited within highly dynamic environments such as cellular networks to support user and end system mobility. Our simulation results obtained in Chapter 4 clearly indicate that moving from tree-structured towards more meshed cellular infrastructures results in a network performance gain. Here, the use of QoS routing algorithms may enhance the performance significantly, if the network architecture provides for alternate paths, thus, increasing the sustained dependability of the network. We believe, however, that we have yet to realize the full power of future cellular network architectures. We envision that not only the routing mechanism in a cellular network is able to adapt to the effects induced by mobility and resulting fluctuations in workload, but the infrastructure itself can contribute to the process of network performance optimization by means of adaptation.

Departing from the results of Chapter 4, we realize that there are few solutions in terms of performance and standardization of routing architectures for highly dynamic cellular environments, which maximize capacity and performance or give dependability assurances. We believe that the mobility induced dynamics should not be easily condemned but may be exploited to stimulate network dependability. Likewise, the mobility of users contributes to the problem of high fluctuations in network load. Hence, we consider the concept of variability in infrastructure as part of the solution to this problem. In our work, we utilize concepts from variable topology networks such as ad hoc networks, where under special conditions the mobility can increase the capacity of the network as shown by Grossglauser and Tse [GT01]. Their approach makes impractical assumptions about the delay tolerance of the data to be transmitted and, thus, cannot be easily adopted for our problem. Nonetheless, related work such as [GT01] and [Fal03] bears the interesting concept to let mobility increase the performance of the network. For our case we need to transfer the mobility of network components into a variability of topology, which suits infrastructure-based cellular network architectures—thus, introducing a new degree of freedom compared to existing cellular architectures. We do not consider the integration of ad hoc networks with cellular networks at this point, because we cannot control the actions of the end systems. Hence, we investigate the variability in infrastructure-based cellular networks.

Since the mobility of infrastructure components in cellular networks is quite restricted compared to ad hoc networks, we imagine another form of network variability to aid our problem: the variability of links between radio access servers. If we assume that wireless technologies act as a replacement for wireline interconnections in our radio access network, we are also able to imagine these connections to be established on-demand. The evolution of beyond 3G networks to include variable point-to-point networks or mesh networks to extend the existing infrastructure is another step in this direction. Addressing the shortcomings of existing work in this area, we first develop our own architecture augmenting today's cellular architectures with variable topology concepts. This includes on-demand use of relay routers, which are integrated in multihop fashion. We label our architecture as *"variability-augmented cellular architecture"*. Second, we are interested in highly dynamic, decentralized, and self-adaptive QoS routing strategies for resource management. We develop a "*near-optimal multiclass minimum-delay routing algorithm*" to operate in our architecture. To combine both areas, in this chapter, we thoroughly investigate the advantages and disadvantages of our novel architecture integrated with our powerful routing technology.

The general direction of our research is set by the requirements and assumptions described in Chapter 4. This includes the capability of gracefully supporting user mobility throughout the network while providing ultra-high capacity in certain areas. Another key issue is the ability to perform resource management in decentralized and distributed manner even under challenging traffic conditions. Our most important contributions in the area of routing architectures and algorithms for cellular networks are as follows:

- We extend the *requirement analysis* from Chapter 4 to include the additional aspects of *variability in infrastructure*.
- We augment concepts from *variable topology networks* to our *cellular network architecture* developed in Chapter 4 to form the novel "*variability-augmented cellular architecture*".
- We discuss the *influence* of our *architecture on routing algorithms*. From the obtained results, we derive the *requirements* for prospective *routing algorithms*.
- We *design* a novel "*near-optimal multiclass minimum-delay routing algorithm*". The algorithm is *analytically formulated* and *analyzed* with respect to important characteristics.
- As a *proof of concept*, the developed routing algorithm is *implemented* in the *ns-2 simulation environment*.
- A *comparative experimental performance analysis* of our algorithm with *selected state-of-the-art algorithms* is performed in *representative topologies* under *various traffic conditions*.

5.2 Outline

This chapter is structured as follows. We design a powerful novel network architecture, which is described in detail in Section 5.3. There, we briefly outline our basic assumptions and the requirements for our architecture. Thereafter, derived from these assumptions and partially based on related work, our architecture is described in detail. We also describe the possible merits of our architecture and discuss the consequences for routing technology. To leverage the full potential of the routing system, we design a near-optimal multiclass minimum-delay routing algorithm, which is able to operate on top of our architecture. We describe the formulation of this algorithm in Section 5.4 in detail. As a proof of concept we implemented our algorithm in the ns-2 simulation environment.

Using this implementation, in Section 5.5, we perform a simulation study of our algorithm in various realistic and artificial topologies. The results are discussed in detail in Section 5.6. In Section 5.7 we survey related architectures and routing algorithms in literature. We summarize our findings in Section 5.8, which also marks the end of our discussion on dependability in cellular networks. The coarse structure of this chapter is shown in Figure 64.

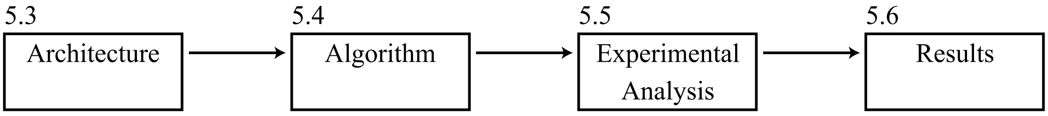


Figure 64: Structure of Chapter 5.

5.3 Variability-augmented Cellular Architecture

In Chapter 4 we extended "traditional" cellular architectures to allow for seamless integration of hot spots as well as to provide the necessary degrees of freedom to adapt to highly dynamic changes in traffic conditions. We now propose to include another dimension into cellular architectures: the variability of infrastructure components themselves. Apart from existing variable networks such as mobile ad hoc networks, we tailor our solution to fit the special needs of cellular networks. In our architecture we specially address the factors scalability and adaptability that have been judged to be of utmost importance during the course of the MobQoS project [HSHS04]. To illustrate this decision we briefly provide one example scenario underlying the MobQoS project:

> *"Future cellular networks are expected to provide nearly ubiquitous coverage while, at the same time, ultra-high bandwidths should be accessible in urban hot spot areas. Thus, we imagine to complement 3G and B3G cellular networks with micro-cells, pico-cells, and hot spots based on the technologies available at that point in time. This translates to an architecture, which allows for extension and reconfiguration of the infrastructure on a short-term scale (to fit actual needs) and a long-term scale (to integrate new network components). The end systems as well as the network should be able to handle context aware handovers while the effects of mobility should be mitigated as well. Both, network components and end systems are to be seen as integral parts of a system, which allows for self-configuration and autonomous operation."*

Motivated by the results of Chapter 3, which provides some insights into the load patterns observed in radio access networks, we observe that the load distribution may fluctuate heavily on a medium timescale. In the case of Darmstadt, we observe a shift in load towards the city center in the morning and back to residential areas in the late afternoon. Traditional cellular network architectures provide only limited adaptability to these load fluctuations. The more flexible MobQoS architecture devised in Chapter 4 allows to compensate these load fluctuations in parts if the available alternate paths are utilized by means of adaptive multipath routing algorithms, which act as highly dynamic online traffic engineering mechanisms. However, the observed dynamics of fluctuation call for a more powerful mechanism. The introduction of variable topology components allows us to utilize approaches from the network design domain. We propose to reconfigure the infrastructure on-demand, compared to the much longer timescale of the traditional network design process. We now move forward to discuss the basic assumptions and requirements for our framework.

The basic requirements for our variability-augmented cellular architecture stem from the area of cellular networks as well as from Internet technology and have already been discussed in the context of the static MobQoS architecture in Section 4.3. We also stick to our earlier introduced design principles—which are adherence to the connectionless Internet paradigm, support for heterogeneity, and decentralized and distributed operation of the routing framework—to move towards a smart network edge. At the same time we relax some of the discussed assumptions and adapt the objectives accordingly for the case of variability in infrastructure:

- The network architecture should be able to deal with *variability/dynamics in topology* such as introduced by variable point-to-point connections.
- The variable infrastructure components should allow for *scalable on-demand extension* of the existing network. These components form the RAN while end systems do not contribute to the variable components.
- The resulting *network topology* should be *self-adaptive* to *changing network and traffic conditions* including load fluctuations, node failures, etc.
- The algorithms used/developed should be capable of *delivering sufficient quality of service* to allow for a wide range of applications.

It should be noted that we focus on wired and wireless point-to-point connections, which form the infrastructure part of the radio access network. We do not consider the influence of interference or of the underlying radio technology, though.

5.3.1 Infrastructure Components to Introduce Topology Dynamics

We propose to enhance our static architecture introduced in Chapter 4 to include variable network components, which may react to load-fluctuations as well as link failures. In particular, we introduce two types of relays, which are able to introduce variability in the network topology: wireless relays (WR) and wireless relay routers (WRR). WR and WRR are not mobile but stationary. However, they allow to establish wireless links between different network nodes on demand and enable a variable topology.

Wireless relays (WR) are targeted to reduce the call blocking/dropping rate and allow to balance load between neighboring radio access points (RAP). Moreover, for heavily loaded cells, the WRs allow for support of additional users. The WRs work on the first tier of our RAN architecture. The functionality of the WR is similar to that proposed for the ad hoc relaying stations in the work by Wu et al. [WQDT01]. WRs operate on link-level only. See Figure 65 for the function of the wireless relays. We model WRs as presented in [WQDT01] and assume that they can support all the different relaying modes outlined in that work. However, we do not consider the WR to be an end user device but a dedicated infrastructure component only. This is to avoid related problems that arise with respect to security and reliability when end systems are involved.

Wireless relay routers (WRR) are relays with routing functionality. They operate on the second tier of our proposed architecture. We do not restrict the wireline interconnections between RASs but provide additional links by introducing WRRs, the so-called virtual wireless links (VWL). We see two main purposes of WRRs as depicted in Figure 66. First, they can be used to provide additional bandwidth between neighboring RASs. Second, they allow to form a direct or multihop path to connect two non-neighboring RASs. For our work, we assume that RASs have wireless transmission/reception capabilities. Moreover, flow control/medium access mechanisms are assumed to be implemented by suitable data link layer protocols. AWRR is able to establish alternate links in the network, thus, a variable topology of the RAN is introduced. However, it does not need to sustain wireless links to all neighbors simultaneously. For the remainder of this document, we assume the WRRs to be capable of having at most two links to neighboring WRRs/RASs. This does not limit the generality of the concept of WRRs, however.

The WRRs are not limited to fixed infrastructure components but may also be mobile. From a provider perspective a mobile WRR can be placed in strategic locations to increase the connectivity of the network and provide relief to heavily congested areas. As discussed in [Kle76], the process of network planning can be used to optimize network throughput. With our proposed architecture such an optimization process can be performed online, that is, a cellular network in combination with WRRs can modify its topology in order to achieve predefined optimization goals. We do, however, focus on the general mode of operation of the infrastructure components in this dissertation and leave this optimization to future work. For seamless operation, the WRRs need support from the routing algorithm, which is detailed in Section 5.4 below.

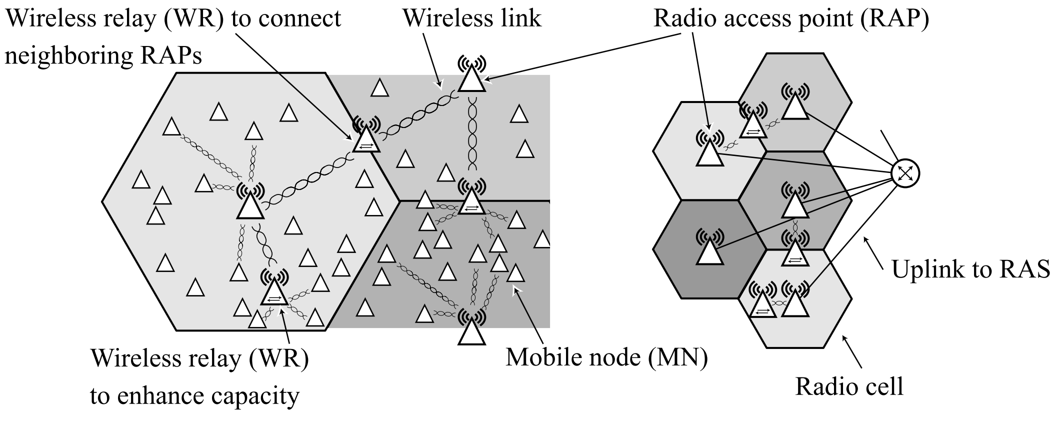


Figure 65: Use cases for wireless relays. We show the possible operation of wireless relays. WRs are introduced in the radio access point (RAP) tier and do not participate in the routing network. They enhance radio access capacity and allow for load-balancing between neighboring radio cells.

Since we are interested in routing networks, we limit the following discussion to the functionality of the wireless relay router, which is unique in our architecture compared to related work. Once deployed, the function of the WRR is to adapt the network topology to aid routing. This includes temporarily adding bandwidth to existing links and connecting RASs on demand like, for example, in case of link-failures. We distinguish two modes of operation of the WRRs namely *pull-mode* and *push-mode*. The WRR operating in the pull-mode requests for data packets to be sent to it based on congestion at neighboring routers and passes on some load to other parts of the network. Here, the WRR uses the status information of the neighboring routers to determine its configuration. In contrast, in the push-mode the WRR behaves like any other RAS and data packets are pushed towards it by the neighboring routers as directed by the routing tables. Because we are mainly interested in routing systems, we restrict our further investigation to the push-mode only. The seamless integration of the WRR in push-mode into the RAN is obvious: the WRR can act like an ordinary RAS (router) if the routing algorithm can deal with the variability in topology. While this tight integration of the WRR into the routing system is appealing, it commands for a relatively complex WRR component that decides autonomously, which virtual wireless links to activate. A more lightweight variant of the WRR can also be implemented in push-mode, though. Here, the choice of the links of the WRR is triggered by information obtained by neighboring RASs. This comes at the expense of a more complicated RAS mechanism. The routing framework has to be able to integrate the characteristics of the wireless links, which are likely to differ from wired link characteristics. In normal operation the RASs avoid the path through WRRs if possible. However, in case of overload situations, the WRRs should be included in the packet forwarding process.

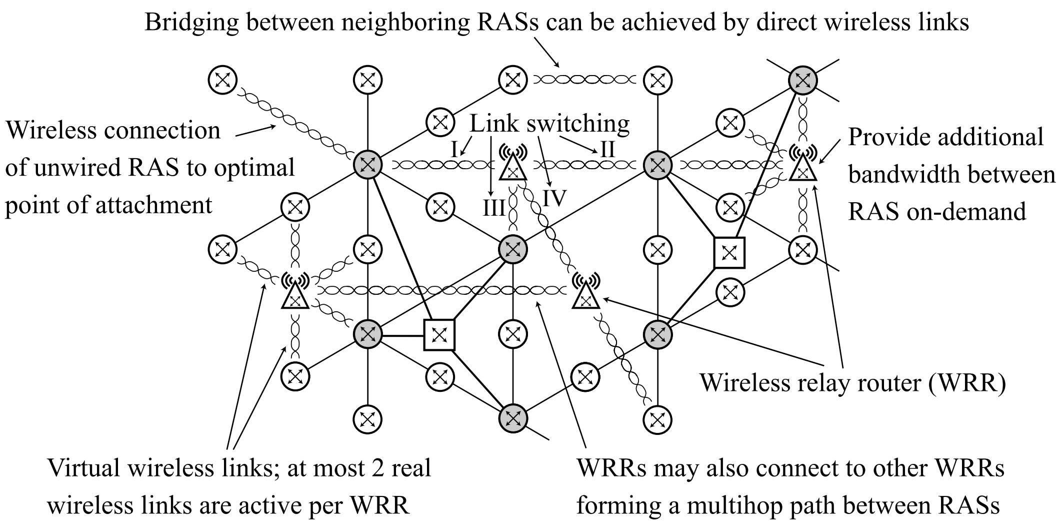


Figure 66: Use cases for wireless relay routers. WRRs introduce variable (switched) links on-demand and include both, link-layer and network layer functionality. They are able to connect RASs directly or in multihop fashion to provide additional bandwidth and allow for dynamic resource management.

5.3.2 Operation of the Variability-augmented Cellular Architecture

To describe the operation of our novel architecture, we outline the functions of the WRR and describe the necessary steps allowing for a basic implementation of the WRR functionality. The physical and electrical aspects of the construction of the WRR are not considered in this work, though. The WRR is a combination of a smart wireless relay as well as a router. Hence, we need strategies to activate the relay functionality between selected links and routing algorithms on top of these. Here, we present a simple heuristic for the activation of the relay functionality that allows to implement a very lightweight WRR, which does not need to maintain routing table entries for all the nodes in the network and give the most important steps in sequential order. More advanced heuristics and the corresponding algorithms for the WRR operation are discussed in [Mog04].

1. *Activation and configuration of a WRR.* For bootstrapping, the WRR executes a neighbor discovery and configures the virtual links and virtual interfaces, respectively. Neighbors are nodes such as other WRRs or RASs, which have a suitable wireless interface to communicate with the WRR and are also within radio range. A simple implementation of the discovery procedure can be performed using broadcast control messages from the WRR. The neighboring nodes reply to the WRR using unicast control messages. The representation of the virtual links between neighboring nodes in the routing table needs to reflect their special role in the network. Since these links should only be activated in case of congestion, the links costs should be set sufficiently high to avoid the routing of regular traffic on these links.
2. *On-demand activation of virtual links.* The virtual links are activated upon requests from RASs. Overloaded RASs send requests for the activation of the link. Depending on the availability of radio resources and the criticality of the request, the WRR activates the respective virtual links.
3. *Seamless integration of virtual links in the routing framework.* Routers have to deal with virtual interfaces and virtual links. For the envisioned multipath algorithms we need to provide for mechanisms to redistribute the load if the virtual link is activated/deactivated.

If we assume the WRRs to have full router functionality and/or to support more than two interfaces simultaneously, these simple heuristic can be enhanced or even integrated with the routing algorithm. We stop this discussion on heuristics for the interaction of the WRRs here and leave this question to future research.

5.3.3 Dependable Routing for Variable Network Infrastructures

The introduction of smart relays yields the capability to modify network topologies on demand and according to traffic and routing needs. Here, we capitalize on non-permanent links that can be switched between the infrastructure components of our network. To exploit the full capabilities of our variability-augmented cellular architecture, we have to use appropriate routing mechanisms. However, routing algorithms available today do not suit our requirements as is shown in Section 5.7 below. There exist various algorithms to deal with network variability in the area of ad hoc communication networks. However, only few of them regard the special characteristics of cellular architectures such as performance guarantees. In addition, our primary design goals for the variability-augmented cellular architecture include decentralized operation, self-adaptability, and robustness. Our routing mechanism must adhere to these goals, too.

We are, at the time of writing, not aware of any routing algorithms that on one hand, sufficiently supports network variability in cellular networks while, on the other hand, optimizes performance and dependability. We find the algorithms discussed in combination with related work in the area of novel cellular architectures to be mostly derived from the simple best-effort ad hoc routing protocols discussed today, namely the *Ad Hoc On-demand Distance Vector Routing (AODV)* by Perkins et. al [PBRD03] and the *Dynamic Source Routing (DSR)* by Johnson et al. [JMH04]. While both protocols are able to react quickly to changing network topologies, they do neither adequately support the scalability requirements for wired infrastructures nor provide for quality of service mechanisms. See Section 5.7 below for more details on routing in the proposed next generation cellular network architectures. A subset of the QoS routing algorithms for the domain of infrastructure-based networks we investigated in Section 4.6 does provide the necessary QoS capabilities. However, the adaptability to variable infrastructures has not been investigated sufficiently yet. Figure 67 summarizes our findings regarding the suitability of existing routing strategies for our architecture.

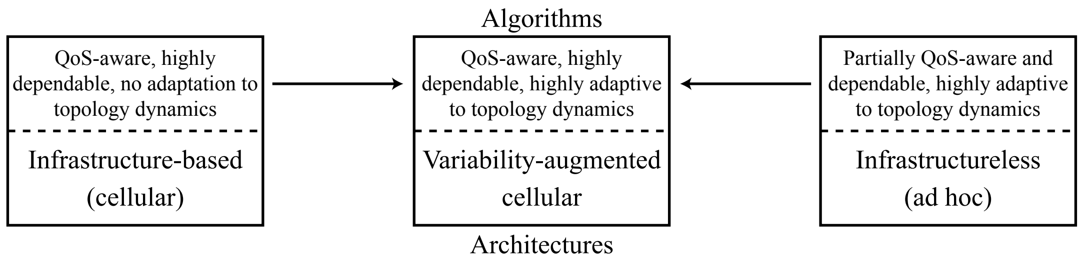


Figure 67: Routing strategies for the variability-augmented cellular architecture.

In the following, we develop an algorithm, which fulfils the necessary requirements. We depart from algorithms in the infrastructure-based domain, which we adapt to fit such a flexible network architecture. In particular, we start from Vutukury's near-optimal algorithm [VGLA99] to solve the minimum-delay routing problem described by Gallager in [Gal77]. However, the solutions in [Gal77] and [VGLA99] cannot easily be applied to real networks with dynamic topology. In particular, our algorithm is addresses the following problems of these algorithms:

- Our algorithm should operate in highly dynamic environments while the optimal solution in [Gal77] is available for static traffic only and Vutukury's solution

[VGLA99] performs best for quasi-static environments with slowly changing traffic conditions.

- The degree of self-adaptability should be increased compared to the algorithm presented in [VGLA99] while the resulting load distribution should approach the optimal (but computationally infeasible) solution of [Gal77].
- The algorithm given in [VGLA99] suffers from severe oscillations under high load situations. Our algorithm should reduce these oscillations.

Despite the fact that Vutukury's work was in part performed in the context of traffic engineering, it assumes best-effort traffic. Within cellular networks, we have to support more complex traffic criteria and differentiate different types of traffic [3GPP03].

A more general issue for the integration of our architecture with a routing algorithm is the choice of a suitable timescale for switching of wireless links. To allow for stable operation, this period should not be too short, because else we may observe a rapid switching of links that may even increase the bottleneck situation we want to resolve initially. If the period for deciding which virtual wireless links to activate is too long, we are not able to respond adequately to congestion caused by temporary bursts of traffic. Here, some interdependencies with the implemented routing algorithm exist. Vutukury's algorithm as well as most routing algorithms in traffic engineering/resource management operate on two different timescales: a longer interval t_l to adapt the routing tables to provide suitable but stable alternate paths and a shorter interval t_s to dynamically re-assign the load to the paths computed in the first step.

Let us briefly explain the possible drawbacks of coupling switching of wireless links with t_l: here the WRR loses the possibility to react to instantaneous congestion or traffic bursts at neighboring RAS until the next re-routing period t_l has expired. If we assume a conservative setting of t_l to match the default settings suggested for the OSPF protocol, this period may be as long as 30 minutes. Link switching on such a coarse granularity in time is suitable for network extensions to cover hot spots caused by planned events such as, for example, concerts or sport events. However, the daily load-fluctuations mandate a much faster reaction time of the topology adaptation. Switching the wireless links faster than t_s leads to problems with the adaptation of the load distribution among the existing paths. The reason is that the adaptation mechanism may not be able to include the virtual wireless link sufficiently. We conclude that the lower bound of the timescale for switching the virtual links lies in between t_s and t_l. This, on one hand, allows the WRR to capitalize on the long-term routing decisions and, on the other hand, to respond fast to congestion caused by temporary bursts of traffic.

5.4 Near-optimal Multiclass Minimum-delay Routing

To fit our scenario of highly dynamic and smart edge networks, we propose a novel routing algorithm, the so-called *near-optimal multiclass minimum-delay routing algorithm* (Algorithm-M^2DR). Our algorithm is specially designed to utilize the variability-augmented cellular architecture developed in Section 5.3 above. The scenario is characterized by wired or wireless point-to-point communication links. To meet our requirements, Algorithm-M^2DR follows the unreliable and connectionless paradigm of Internet routing. The most important characteristics of the algorithm regard this paradigm and are as follows:

- The support of *multiple traffic (priority) classes*.
- The *route computation* is *distributed*.
- The algorithm is capable of *multipath routing*.
- The algorithm is *self-adaptive to changing traffic conditions.*
- The algorithm supports *variable topologies*.
- The routing is *near-optimal* with respect to the *minimum-delay routing problem*.
- The algorithm's *computational complexity* is *suitable for today's routing hardware*.

Next, we introduce basic definitions and present the minimum-delay routing problem as formulated by Gallager [Gal77]. We explain the work of Vutukury and Garcia-Luna-Aceves [VGLA99] [VGLA01] [Vut01], which presents a near-optimal solution to the work of Gallager. Finally, we extend the work of Vutukury et al. and formulate a near-optimal multiclass minimum-delay routing algorithm.

5.4.1 Basic Definitions

Let a communication network be described as *graph* $G = G(V, E)$, which consists of a finite set V of $|V| = p \geq 1$ *vertices* (nodes, routers), together with a set E of $|E| = q \geq 0$ unordered pairs of individual vertices. G is said to have the order p and size q. We refer to G as a (p, q)-graph. The elements of E are the *edges* (links) of G. An edge $\{i, k\}$ is said to join i and k in G. The link $\{i, k\}$ is said to be *incident to* the nodes i and k, which are *adjacent* nodes, so-called *neighbors*. A node is called *isolated* if no edge is incident to it. We define the set of neighbors, N^i, of node i to be

$$N^i \equiv \{k | \{i, k\} \in E\}. \tag{14}$$

The *degree* of a node i, $deg(i)$, gives the number of edges incident to it. In the following we assume links to be unidirectional with possibly differing cost in each direction. We speak of a directed graph or *digraph*. The digraph is defined as a graph $G = G(V, E)$ with edges consisting of *ordered* pairs of distinct nodes. In the case of a directed graph we also write $l = (i, k) \in E$ with the special meaning of a link *from* i *to* k. The opposite link $l' = (k, i)$ needs not necessarily to be in E if l is in E. We speak of G as *symmetric* digraph if $\forall l, l' \in E$, and of G as *asymmetric* if this never is the case. For a given node $i \in V$ the *indegree* of i, $indeg(i)$, is defined to be the number of nodes $k \in V$ for which $(k, i) \in E$. The *outdegree* of i, $outdeg(i)$, equals the number of links in E that go from i to some other node $k \in V$.

5.4.2 Problem Formulation

The minimum-delay routing problem (MDRP) was formulated by Gallager [Gal77] and builds upon the optimal routing problem described earlier by Cantor and Gerla [CG74]. Here we present the key ideas as far as they are necessary for the understanding of our work. Additional details can be found in [Gal77]. For the problem formulation we assume links to be unidirectional and that if $l = (i, k)$ exists so does $l' = (k, i)$. Let the rate of traffic entering the network at router i for destination j be denoted as $r_j^i \geq 0$. r_j^i is measured in bit per second. Further, let t_j^i be the total traffic at router i destined for router j in bit per second. t_j^i can be obtained by summation of r_j^i and the traffic arriving at i from its neighbors for destination j. Figure 68 shows a sample graph to visualize the basic variables.

Let the routing parameter ϕ_{jk}^i be the fraction of all traffic t_j^i that leaves the router i over the link (i, k) for destination j (that is, k is the next-hop from i to j). If the network does not lose any packets, we have

$$t_j^i = r_j^i + \sum_{k \in N^i} t_j^k \phi_{ji}^k . \tag{15}$$

We define f_{ik} to be the expected traffic on link (i, k). f_{ik} is measured in the unit bit per second. $t_j^i \phi_{jk}^i$ describes the traffic destined for router j on link (i, k), and accordingly the summation over all adjacent links to router j gives us f_{ik} to be

$$f_{ik} = \sum_{j \in N} t_j^i \phi_{jk}^i . \tag{16}$$

f_{ik} is upper bounded by the capacity C_{ik}, measured in bit per second, of the link (i, k): $0 \leq f_{ik} \leq C_{ik}$. The parameters have to satisfy the following conditions for each router i and destination j:

$$\phi_{jk}^i = 0 \text{ if } (i, k) \notin E \text{ or } i = j , \tag{17}$$

$$\phi_{jk}^i \geq 0 \text{, and} \tag{18}$$

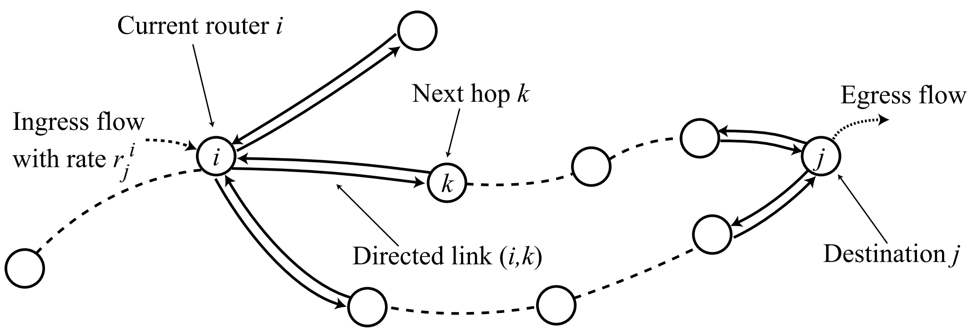


Figure 68: Visualization of a sample graph.

$$\sum_{k \in N^i} \phi_{jk}^{\ i} = 1\,. \tag{19}$$

We define D_{ik} as the expected number of packets (messages) per second transmitted on link (i, k) multiplied by the expected delay per packet. The delay per packet includes the queueing delay on the link. Moreover, D_{ik} does only depend on flow f_{ik} through link (i, k), which can be described using its link characteristics such as propagation delay and capacity. $D_{ik}(f_{ik})$ is increasing and convex and tends to infinity as the flow f_{ik} approaches the link capacity C_{ik}. The total expected number of packet arrivals per second multiplied by the total expected delay per packet is given by (see, for example, [Kle76] for the reasoning)

$$D^T = \sum_{(i,k) \in E} D_{ik}(f_{ik})\,. \tag{20}$$

The router traffic flow set $t = \{t_j^{\ i} | i, j \in V\}$ and the link flow set $f = \{f_{ik} | (i, k) \in E\}$ can be obtained from $r = \{r_j^{\ i} | i, j \in V\}$ and $\phi = \{\phi_{jk}^{\ i} | (i, k) \in E, j \in V\}$. Therefore, we can express D^T as function of r and ϕ using Equation (16) and Equation (20).

The minimum-delay routing problem can now be described as follows: Assume a given fixed topology and input traffic flow set $r = \{r_j^{\ i} | i, j \in V\}$, and delay function $D_{ik}(f_{ik})$ for each link (i, k). The minimization of the total expected delay D^T can now be achieved by computing the routing parameter set $\phi = \phi_{jk}^{\ i}$ appropriately.

5.4.3 Optimal Solution to Minimum-delay Routing

In [Gal77], Gallager's theorem states the necessary and sufficient conditions to be satisfied for the solution of the minimum-delay routing problem. Here, we briefly summarize these conditions because they are vital for the understanding of our work. The partial derivatives with respect to r and ϕ of the total expected delay D^T serve as the vehicle to formulate and solve the MDRP and are given as

$$\frac{\partial D^T}{\partial r_j^{\ i}} = \sum_{k \in N^i} \phi_{jk}^{\ i} \left[D'_{ik}(f_{ik}) + \frac{\partial D^T}{\partial r_j^{\ k}} \right] \text{ and} \tag{21}$$

$$\frac{\partial D^T}{\partial \phi_{jk}^{\ i}} = t_j^{\ i} \left[D'_{ik}(f_{ik}) + \frac{\partial D^T}{\partial r_j^{\ k}} \right], \text{ respectively.} \tag{22}$$

$D'_{ik} = \partial D_{ik} / \partial f_{ik}$ denotes the so-called *marginal delay* or *incremental delay* while the *marginal distance* from router i to router j is represented using $\partial D^T / \partial r_j^{\ i}$. Thus, Equation (21) shows the relationship between the marginal distance of one router via its neighbors towards one particular destination. Gallager's theorem [Gal77] derives the necessary condition to obtain the minimum of D^T with respect to ϕ for all $i \neq j$ and $(i, j) \in E$ to be

$$\frac{\partial D^T}{\partial \phi_{jk}^{\ i}} = \begin{cases} = \lambda_{ij} \text{ for } \phi_{jk}^{\ i} > 0 \\ \geq \lambda_{ij} \text{ for } \phi_{jk}^{\ i} = 0 \end{cases}, \tag{23}$$

λ_{ij} being a positive number. The sufficient condition to minimize D^T with respect to ϕ for all $i \neq j$ and $(i, j) \in E$ is then

$$D'_{ik}(f_{ik}) + \frac{\partial D^T}{\partial r_j^{\ k}} \geq \frac{\partial D^T}{\partial r_j^{\ i}} . \tag{24}$$

Equations (22) to (24) can be used to obtain the routing parameter set ϕ to yield *minimum overall delay* inside the network, which indeed maximizes the throughput of the network. Adhering to the definitions introduced in Chapter 2 and the scenario presented in Chapter 4, we maximize the utility of the network by minimizing the overall delay, thus, increasing its performance/dependability.

Let the set of neighbors through which a router i forwards traffic towards the destination j be denoted as successor set $S_j^{\ i}$ of i for destination j. The marginal distance through neighbors in the successor equals the marginal distance of the router itself for perfect load-balancing of the traffic towards a particular destination. The marginal distances through routers not in the successor set are higher than the marginal distance of the router itself. Figure 69 shows the successor set for node i.

Let $D_j^{\ i}$ denote the *marginal distance* from i to j, that is, $\partial D^T / \partial r_j^{\ i}$. Let moreover the *marginal delay* $D'_{ik}(f_{ik})$ from i *to* j be denoted by $l_k^{\ i}$, the so-called cost of the link (i, k) from router i to router k. Using the results of Gallager's work we can now solve the minimum-delay routing problem by determining the routing parameters $\phi_{jk}^{\ i}$, $S_j^{\ i}$, and $D_j^{\ i}$ at each router i for all destinations j such that the following five equations are satisfied:

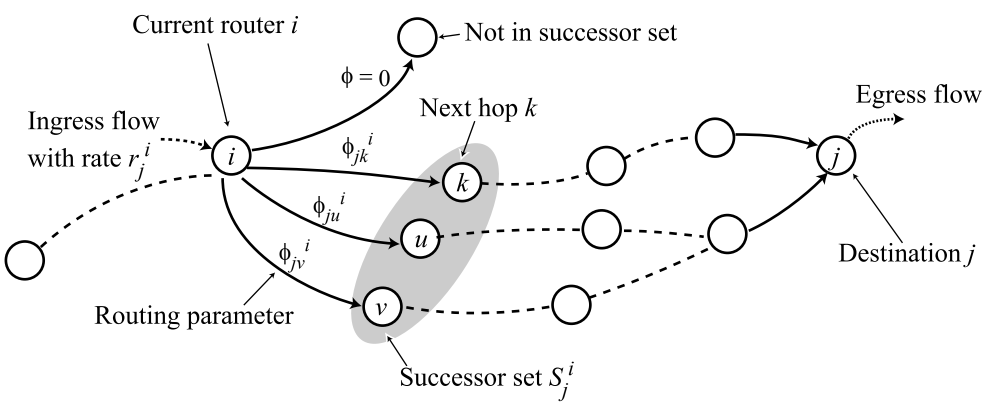


Figure 69: Routing related variables and successor set.

$$D_j^i = \sum_{k \in N^i} \phi_{jk}^i (D_j^k + l_k^i) \tag{25}$$

$$S_j^i = \{k | \phi_{jk}^i > 0 \text{ and } k \in N^i\} \tag{26}$$

$$D_j^i \le D_j^k + l_k^i \text{ for } k \in N^i \tag{27}$$

$$(D_j^u + l_u^i) = (D_j^v + l_v^i) \text{ for } u, v \in S_j^i \tag{28}$$

$$(D_j^u + l_u^i) < (D_j^v + l_v^i) \text{ for } u \in S_j^i, v \notin S_j^i \tag{29}$$

More intimate details of Gallager's algorithm for a distributed solution of Equations (25) to (29) can be found in [Gal77]. While the resulting optimal routing is very desirable, the implementation of the algorithm is hindered by various problems. First, the calculation of the routing parameters set ϕ needs to be initiated at the destination. Multiple iterations, depending on the size of the network, need to be performed until the optimal solution of the problem is found and the number of routing messages needs to be proportional to the number of links. Second, the algorithm depends on a global constant for its convergence [Gal77]. This variable is dependent on the network traffic conditions at each point in time. Optimizing the network with respect to fast convergence is coupled with the risk of never converging to the minimum-delay, however. As a consequence the algorithm is only suitable for static or quasi-static environments with respect to traffic and topology. In summary, the algorithm of Gallager presents a theoretical sound framework, which in practice cannot be implemented easily.

5.4.4 Near-optimal Solution to Minimum-delay Routing

Vutukury et al. [VGLA99] [VGLA01] [Vut01] have developed a near-optimal solution to minimum-delay routing. Their work builds on the concepts described in the previous subsection; the general framework being minimum-delay routing. In contrast to the work of Gallager [Gal77] the algorithms provided in [Vut01] can be practically implemented in distributed fashion. Other properties of the algorithms include instantaneous loop freedom, multipath forwarding capability, and no need for global synchronization or global constants. In this subsection we give a concise description of Vutukury's work.

The basic idea of Vutukury is to split up the combined optimization process of Gallager into multiple independent stages. First, multiple loop free paths are established using global (end-to-end) long-term delay information. Second, local short term delay information is used to adjust the routing parameters along the precomputed multipaths. Gallager's Equations, (25) to (29), have been transformed into the following three equations by Vutukury:

$$D_j^i = min\{D_j^k + l_k^i \,|\, k \in N^i\} \tag{30}$$

$$S_j^{\,i} = \{k|D_j^{\,k} < D_j^{\,i} \text{ and } k \in N^i\} \tag{31}$$

$$\psi(k, A_j^{\,i}, B_j^{\,i}) \text{ for } k \in N^i = \phi_{jk}^{\,i} \tag{32}$$

with $A_j^{\,i} = \{D_j^{\,u} + l_u^{\,i} \big| u \in N^i\}$ and $B_j^{\,i} = \{\phi_{ju}^{\,i} \big| u \in N^i\}$.

The function ψ is used to allocate the flows over the paths and can be implemented using a heuristic. The meaning of Equations (30) to (32) is that, at any time t, the successor sets of all nodes define a routing graph, $SG_j(t)$, for a particular destination j to be

$$SG_j(t) = \{(m, n)|n \in S_j^{\,m}(t), m \in V\}. \tag{33}$$

While Equations (30) to (32) provide a tool to solve the MDRP in distributed fashion, they are based on the assumption that the routing information is consistent throughout the entire network at each point in time. Since this is not true for real networks, Vutukury introduces loop free invariant (LFI) conditions, which ensure that routing loops are omitted during the computation of the paths. We take the definition of the loop free invariants from Vutukury [VGLA99] to be:

$$FD_j^{\,i} \leq D_{ji}^{\,k} \text{ for } k \in N^i \tag{34}$$

$$S_j^{\,i} = \{k|D_{jk}^{\,i} < FD_j^{\,i} \text{ and } k \in N^i\} \tag{35}$$

According to Vutukury's definition "any routing algorithm designed such that the two Equations (34) and (35) are always satisfied, automatically provides loop free paths at every instant, regardless of the type of routing algorithm being used". Here, $D_{jk}^{\,i}$ represents the value of $D_j^{\,k}$ reported by router i to its neighbor k. $FD_j^{\,i}$ denotes the *feasible distance* of router i for destination j. $FD_j^{\,i}$ equals $D_j^{\,i}$ in steady state but is allowed to differ from it during state transitions of the network, being an estimate of $D_j^{\,i}$ until reaching steady state again. Further details of Vutukury's work including the proof of his theorems and pseudo-code for his implementations of the devised algorithms can be found in [VGLA99] and [Vut01].

5.4.5 Multiclass Approach to Minimum-delay Routing

We extend the work of Vutukury et al. and formulate a near-optimal *multiclass* minimum-delay routing algorithm (Algorithm-M^2DR). Until now we regarded the optimization parameter, namely the total expected delay D^T, to be measured without distinction of various classes of traffic. In the case of the near-optimal routing algorithm described in Section 5.4.4, this results in increased delays in some flows compared to a single-path algorithm, while the average delay over all flows is better compared to single-path routing. We believe that in multiservice networks such as cellular networks not all classes of traffic should be treated equally. In the case of Vutukury's work only one class of service is available and may lead to high priority traffic having a worse performance than lower priority traffic for the same source destination pair.[11] To be applicable for the envisioned environment the goal of our work is to devise a solution for the multicommodity routing problem, which

- provides for *near-optimal routing* and, at the same time,
- ensures that the performance experienced by *higher priority traffic* is not worse than the performance experienced by lower priority traffic.

We want to maintain the distributed and connectionless nature of Vutukury's algorithm while we approach a multiclass algorithm in which packets can be classified into a finite number of classes $m \in M$, $|M| = g$. This alleviates us from keeping per flow state in the network, too. Classes may be introduced with respect to delay tolerance to conform with usual traffic models such as proposed in [3GPP03]. Without loss of generality, we assume that the assignment of individual packets to particular priority classes is done at the end systems.

Concept

We can formulate the problem as an optimization problem where the aim is to optimize the overall delay in the network while preserving the priorities of the individual classes. The objective function considers that the traffic classes m are not treated equally:

$$D^T = \sum_{(i,k) \in E, m \in M} \omega_m \times D_{ik_m}(f_{ik}). \quad (36)$$

We introduce a term ω_m to describe the relative contribution of the delay experienced by packets of the respective priority class m to the overall delay. Unfortunately, the minimum-delay routing for the multiclass case as outlined in Equation (36) cannot be solved easily. To proceed, we need to obtain the delay functions for each class, which give the delay experienced by packets of one class on a particular link as a function of the existing flow on that link. It is obvious that the interactions among the flows of different classes of traffic on a link are complex, and traffic of each class on a link has an impact on the delay experienced by the traffic of other classes on the link. Online methods for estimating the delay gradient like perturbation analysis for single class networks are given by Segall [Seg77], Ho [Ho87], and Cassandras et al. [CAT88]. However, they cannot be transferred to the case of a multiclass network because they fail to consider the aspects of intra-class and inter-class dependency simultaneously. In such a scenario, it is not feasible to obtain the delay gradients for the individual flow classes, and also the gradient of the total delay cannot be easily obtained as shown by Cassandras in [CAT88].

Moreover, it has been shown in an example of Netter [Net72] that multiple equilibrium solutions exist for a multiclass routing problem. The reached solution depends on

11) In the area of public transport networks, the work of Dafermos [Daf72] targets a related problem area, namely the assignment of flows for a multiclass transportation network. However, the approach is centralized and requires prior knowledge of the traffic demands.

the initial state of the network as well as the parameterization of the classes. Existing algorithms assume that the delay functions are monotone and also differentiable over the entire solution set. In particular, Kobayashi and Gerla [KG83] show, for the case of a closed queueing network that the optimal routing algorithm they consider converges to differing local delay-minima depending on the initial conditions of the network. Nonetheless, the algorithm requires a priori knowledge of the flow requirements and assumes that the flows as well as the topology are static. In real implementations this algorithm would also suffer from a slow convergence.

For our variability-augmented cellular architecture to support user mobility, these strict requirements are not desirable and cannot be satisfied. Given the fact that the near-optimal routing approach provided by Vutukury in [VGLA99] relaxes most of the above restrictions, our approach is to extend his work to be applicable in the case of multiple traffic classes. Multipaths are orthogonal to class-based services in multiclass networks. Nevertheless, both dimensions may benefit from each other. In developing our routing framework, we strive to obtain better traffic engineering and resource management capabilities of the network than shown by Vutukury for his work in [Vut01]. At the same time we do not want to introduce per-flow reservation to keep the state of the network reasonable and stay with a distributed algorithm of low computational complexity. In the following the concepts and algorithms of our routing framework are described and a queue management approach to implement sufficient priority mechanisms is proposed.

5.4.6 Near-optimal Multiclass Minimum-delay Routing Algorithm

It has been shown that the near-optimal multiclass minimum-delay problem cannot be solved with as simple solutions as in the single-class case. The reason is mainly the arbitrary cross-influence of traffic of one class on the delay experience by flows in other classes. Here, not only higher priority classes affect lower classes traffic but also vice versa. As stated earlier, we assume the traffic of our network to be separated into g different priority classes, where class 1 denotes the highest priority traffic. To approach our problem, we virtualize the network, that is, we logically separate the routing network for the individual classes. This results in g logical separate networks, one for each traffic class. By introducing a strict priority queueing mechanism we are able to simplify the interdependency between classes, that is, the relationship between the delay experienced by a particular flow of the network to other flows of differing classes.

The asymmetric relationship between higher and lower classes provides a tool to the network designer/operator to specify the quality a higher class flow should experience compared with the perceived quality of a lower class flow. The near-optimal delay bears a different interpretation for each class, because the routing network is class-dependent. The benefit of separating the network into g logical networks is the ability to approach our multiclass problem as a series of single-class problems that can be solved using any algorithm for the single-class routing problem. If we assume a strict distinction of priority classes, that is, packets enqueued into one priority class are never transferred into another priority class, some proven properties of the single class algorithm also hold for our multipath multiclass algorithm. This includes for example loop freedom.

Our algorithm for near-optimal minimum-delay multipath routing in a multiclass network, Algorithm-*M²DR* (Algorithm-*MulticlassMinimumDelayRouting*), makes use of a distributed algorithm, which provides a solution to the near-optimal minimum-delay routing problem for a single class network, Algorithm-*MDR* (Algorithm-*MinimumDelayRouting*). In Figure 70 we give pseudo code for Algorithm-*M²DR*. We apply Algorithm-*MDR* to compute the routing parameters separately for each class iteratively. This single-class algorithm can be any near-optimal routing algorithm to solve the minimum-delay routing problem. Algorithm-*MDR* computes the multipath routing table MRT for one class at each router. The link costs are given in the link state map LSM of the graph G. LSM contains the topology table for the network for all priority classes and is maintained at all nodes using link-state updates.

In order to solve the multiclass problem, Algorithm-*M²DR* and Algorithm-*MDR* rely on a multiclass capable cost function: Function-$getLinkCost(m)$ (see Figure 72). In a single-class real network, analytical methodologies like perturbation analysis given by Cassandras et al. [CAT88] can be applied to compute the gradients of the delay on links with respect to the flow. However, for the case of a multiclass network the results may be highly biased. To be able to exploit the asymmetric relationship between the flows of different classes in order to achieve the near-optimal multiclass minimum-delay routing, we need to compute the class-dependent link costs. In our case these costs are given by the incremental delay $D'_{ik} = \partial D_{ik} / \partial f_{ik}$, which differs for each logical copy of the network. We use the following equation for the link delay D_{ik} as a function of the flow f_{ik} entering the system based on a M/M/1 approximation of the link.[12]

$$D_{ik}(f_{ik}) = \frac{f_{ik}}{(C_{ik} - f_{ik})} + f_{ik} T_{ik} \tag{37}$$

The incremental delay is given as

$$D'_{ik}(f_{ik}) = \frac{C_{ik}}{(C_{ik} - f_{ik})^2} + T_{ik}. \tag{38}$$

Function-$getLinkCost(m)$ operates on the offered flows ${f_{ik}}^m$ and the perceived capacities $C_{perceived}$. For calculation, Algorithm-*MDR* passes the arguments extracted from the link state map LSM, which are updated after the calculation. The process is performed for all links $l \in E(G)$ of the router i in question. To calculate the cost of a link $l = (i, k) \in E$ for class m, the Function-$getLinkCost(m)$ is realized as described in the pseudo code given in Figure 72.

12) We perceive the M/M/1 approximation to be sufficient to model the traffic characteristics in our target environment of a radio access network. The use of more general approximations complicates the analysis significantly.

```
Algorithm MulticlassMinimumDelayRouting      // M^2DR
//   Solves the multiclass minimum-delay routing problem at router i,
     i ∈ V of the graph G = (V, E) for all classes of a multiclass system
     with g classes m ∈ M. Function-getLinkStateMap retrieves the
     relevant link-state information. Algorithm-M^2DR provides
     multipaths as solution set.
Definitions
     LSM denotes the link-state map containing the cost information of
          all links;
     MRT[m] denotes the multipath routing table of node i for class m;
     t_l denotes the interval for calculation of the successor set;
     t_s denotes the interval for calculation of the routing parameter φ;
     t_wait denotes the pause time between the calculations for the
          individual classes;
     Algorithm-MinimumDelayRouting is a solver for the single-class
          minimum-delay routing problem;
Begin
     Each t_l For all priority classes m Do
          Get and store LSM ← getLinkStateMap(m);
          Calculate MRT[m] ← MinimumDelayRouting(LSM, i);
          Wait (t_wait) for routing changes to take effect;
     EndFor
     Each t_s For all priority classes m Do
          Get and store LSM ← getLinkStateMap(m);
          Redistribute load in MRT[m];
          Wait (t_wait) for load redistribution to take effect;
     EndFor
End
```

Figure 70: Pseudo code for Algorithm-*M^2DR*.

In particular, the above outlined algorithm computes possibly different routing graphs for the same destination depending on the priority class. At each node the packets for this class are routed accordingly to the directives of the local class-specific multipath routing table $MRT[m]$. This mechanism, that is, the segmentation of the network into different logical networks, in combination with a strict priority mechanism, enables us to guarantee better performance for high priority flows compared with lower priority flows. In contrast, a naive multipath extension of Vutukury's work without modification of the minimum-delay routing algorithm would operate on the same routing graph for all priority classes. Here, the lower priority classes suffer from the priority mechanism in place, because they are bounded by the routing graph of the highest priority class.

Figure 71 gives an example of different routing graphs for a network with fictive link costs and traffic demands. For the highest priority class the link costs are calculated using the capacity of the respective links and yield the routing graph for priority class 1. In our example two paths have been chosen, which are assumed to provide the best quality. For the second priority class, the already admitted ingress flow of class 1 decreases the perceived capacity and, thus, increases the costs of the links. As a reaction of the changing link costs, the routing graph for priority class 2 differs in our example from the graph for class 1. The same is true for the graphs of class 3 and 4. The routing graphs for the individual classes may even utilize the same link in different direction depending on the network and load characteristics, which is not shown in our example, though. It should be noted that the routing graphs do *not* necessarily differ for the individual classes.

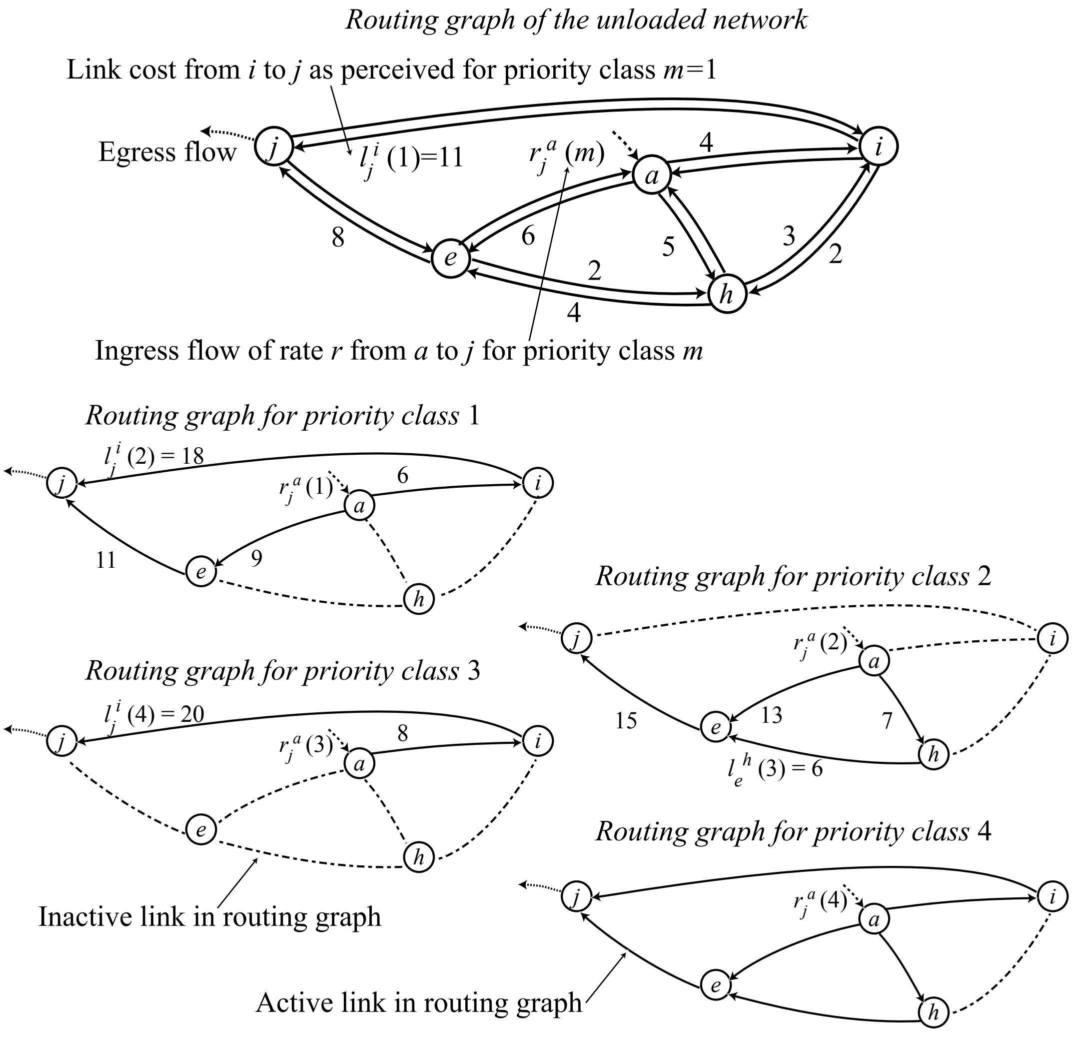


Figure 71: Example of the virtualized routing graphs for multiple priority classes. The individual routing graphs are shown for fictive link costs.

Function $getLinkCost(m)$
// Calculates the cost, $Cost(m)$, of one link, $l = (i, k) \in E$,
depending on the priority class m. The incremental delay,
$D'_{ik} = \partial D_{ik} / \partial f_{ik}$, as perceived by packets belonging to class m
acts as cost metric. A M/M/1 approximation of the link delay, D_{ik}, is
used for the computation (see Equation (38)).
Definitions
$Cost$ denotes the link cost;
$C_{perceived}$ denotes the capacity as perceived by the flow of class m;
C_{total} denotes the total capacity of the link;
α, β, γ denote modeling parameters;
D' denotes the incremental delay;
$T_{propagation}$ denotes the propagation delay of the link;
Begin
$f_{higher} \leftarrow 0$; $f_{lower} \leftarrow 0$;
For all priority classes i above m **Do**
Sum-up flows $f_{higher} \leftarrow f_{higher} + f[i]$;
EndFor
For all priority classes i below m **Do**
Sum-up flows $f_{lower} \leftarrow f_{lower} + f[i]$;
EndFor
Calculate $C_{perceived} \leftarrow C_{total} - \alpha \cdot f_{higher} - \gamma \cdot f_{lower}$;
Calculate $D' \leftarrow \dfrac{Capacity_{perceived}}{(Capacity_{perceived} - \beta \cdot f[m])^2} + T_{propagation}$;
Set $Cost \leftarrow D'$;
Return $Cost(m)$;
// α, β, γ are used for adaptation of the logical copy of the
network, for example, to the chosen scheduling discipline.
α is used for priority classes above m, β for class m, and γ
for priority classes below m. The values can be in range [0;1].
For a strict priority mechanisms we use $\alpha = 1, \beta = 1, \gamma = 0$.
// $\beta < 1$ reduces the oscillatory behavior of *Algorithm-M²DR*
in high load situations. A possible feedback mechanism is to
change β depending on the load or loss rate: for high loss rate
chose $\beta \approx 1$ else $\beta < 1$.
End

Figure 72: Pseudo code for Function-*getLinkCost*(*m*).

Complexity

The complexity of our algorithms can be assessed with the knowledge of the complexity of the underlying Algorithm-*MDR*, in our case taken from [Vut01]. The *storage complexity* of our algorithm is the table space needed at each node. Similar to [Vut01] we have neighbor tables to each of the neighbors in N^i. Moreover, the network state contains the long-term delay information to all other nodes in V. The routing table per class can grow to at most $|N^i|$ times $O(|V|)$. Hence, assuming $|M| = g$ classes, the storage complexity at each node is $O(|M||V||N^i|)$. However, while V and N^i can grow, M has a constant size.

The *computational complexity* of our algorithm is determined by the number of classes and neighbors in combination with the underlying routing algorithm (Dijkstra's algorithm). The former is $O(|M||V||N^i|)$, because for each class $m \in M$ the update of routing information to form the MRT has to be performed to each possible destination in V for at most all neighbor nodes in N^i. With the complexity of Dijkstra's algorithm in our case being $O(|N^i||V|\log(|V|))$, we obtain the complexity to be $O(|M||V||N^i| + |N^i||V|\log(|V|))$ for our algorithm. Again M has constant size.

Per-node Queue Management and Priority Mechanisms

In our analysis we require the algorithm to make use of a strict priority mechanism to justify our assumptions. One prominent scheduling mechanism to fulfil this requirement is the pure strict priority queueing. Strict priority queuing is easy to understand and implement and widely available in today's routers[13]. In reality, since strict priority queueing is known to starve lower priority classes if the highest class is allowed to take all available bandwidth, the network operator may introduce additional control mechanisms. This may include an admission control, which, depending on the parameterization, allows for flexibility in class bandwidth shares with respect to the providers needs. The preparation of a so-called "service menu" is one possible approach. The concept of service menus is described in more detail in [SHHS03]; the investigation covers an admission control scheme, which provides per-flow delay and bandwidth guarantees based solely upon simple class-based strict priority queueing.

Another possibility for queue management is the reservation of fixed bandwidth shares on the links. This allows for contention (using our class-based strict priority approach) for the non-reserved bandwidth while the reserved bandwidth shares are exclusively available for the respective traffic classes.

Numerical Example for Strict Priority Queuing without Fixed Bandwidth Shares

To illustrate the operation of Algorithm-M^2DR we give a numerical example of the Function-$getLinkCost$. The function returns the cost of one link and is called from the algorithm to compute the multipath routing table. See Figure 73 for the resulting queue-

13) For example, most Cisco routers implement strict priority queueing.

Table 7: Numerical Example of the Function-*getLinkCost* in Algorithm-M^2DR.

	Parameter								
	Case-1			Case-2			Case-3		
Class	f_{ik}^{m}	C_{ik}^{m}	$Cost(m)$	f_{ik}^{m}	C_{ik}^{m}	$Cost(m)$	f_{ik}^{m}	C_{ik}^{m}	$Cost(m)$
Class-1	2.4	10.0	10.17	8.2	10.0	13.09	4.0	10.0	10.28
Class-2	2.4	7.6	10.28	1.4	1.8	21.25	4.0	6.0	11.50
Class-3	2.4	5.2	10.66	0.33	0.4	91.63	4.0	2.0	∞ *
Class-4	2.4	2.8	27.50	2.0	0.07	∞ *	4.0	0.0	∞ **

The table displays the link costs, $Cost(m)$, in milliseconds in a four class network. The expected flow, f_{ik}^{m}, and the perceived capacity of the link, C_{ik}^{m}, are given in kbit/s. The available capacity, C_{ik}, of the link is 10 kbit/s, the propagation delay, T_{ik}, is 0.01 seconds.

* The flow can only be transferred with extreme loss since the requested bandwidth of the flow exceeds the remainder of the capacity.

** The flow cannot be transferred because the remaining bandwidth share is 0.

ing system at the outgoing link of router i. We use separate queues of fixed length for each class per link to prevent head-of-the-line blocking. The server process is non-preemptively serving the four classes on the outgoing link following a strict priority directive. It should be noted that because of the strict priority scheme, lower classes may suffer from starvation if the input exceeds the capacity of the server (link).

Table 7 shows the calculated link costs for an example setup in some pathological situations. In Case-1 the link is nearly fully loaded resulting in an increase in cost for the lower priority classes. This may cause a change in the multipath routing table for the constrained classes. Case-2 shows an overbooked link. The increase in cost is evident because the highest priority class occupies most of the capacity and Class-2 and Class-3 get to see only the small remainder of the bandwidth. Link conditions for the upper classes are appropriate in Case-3. However, Class-3 tries to overbook the link and suffers from a severe degradation in service. Class-4 is deferred because there is no capacity left over. Summarized, it can be seen that with the capacity of the flow approaching the per-

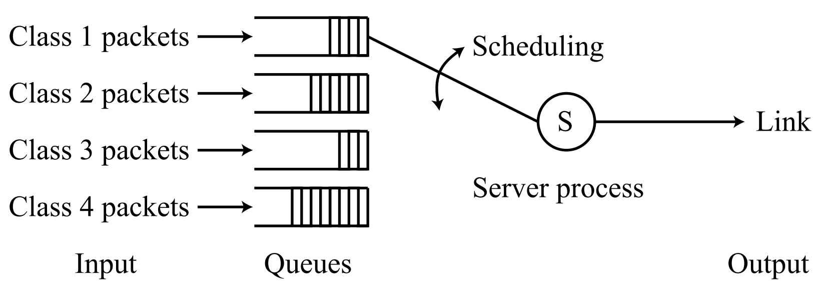


Figure 73: Link queueing model.

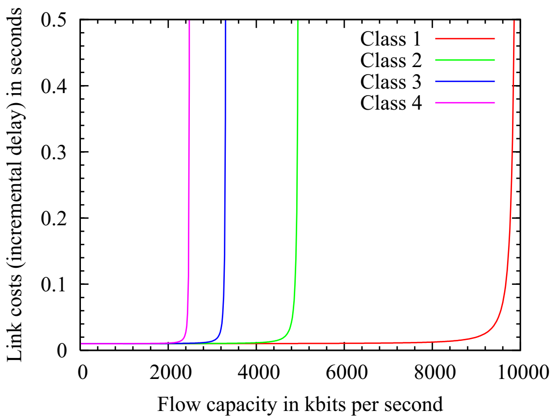


Figure 74: Link costs *Cost(m)* for strict priority queueing without fixed bandwidth shares. The x-axis depicts the estimated flow capacity in kbit/s, which is increased for each class simultaneously. The y-axis shows the costs, that is, the incremental delay D'_{ik} in milliseconds for the expected flows. The available capacity C_{ik} of the link is 10 kbit/s, the propagation delay T_{ik} is 0.01 seconds.

ceived capacity the delay increases significantly. Flows exceeding the available bandwidth cannot be delivered at all or at most with possibly severe degradation in service.

Figure 74 visualizes the Function-$getLinkCost$ and backs our observations from the pathological cases in Table 7. We notice that the link costs for all classes remain nearly constant under moderate load situations. If the flow f_{ik}^{m} approaches the perceived capacity of its class, we see a significant increment in cost. For further analysis we select an increase above 10% of the unloaded state to mark the start point of this behavior while hitting a ten times increased incremental delay compared with the initial delay defines our second measurement point. If we assume that a flow is admitted even if it exceeds the perceived capacity and tolerate the resulting extreme packet loss, we reach the third point of interest: the point of complete starvation of a class. We give the respective x-axis intercept as fraction of the total link capacity C_{ik}.

For the given setup, in Figure 74, the first point can be observed at $f_{ik}^{4} = 0.2 \cdot C_{ik}$ for class 4, at $f_{ik}^{3} = 0.26 \cdot C_{ik}$ for class 3, at $f_{ik}^{2} = 0.37 \cdot C_{ik}$ for class 2, and at $f_{ik}^{1} = 0.68 \cdot C_{ik}$ for class 1. The second point should hardly be reached in normal operation if other feasible paths are available: it marks a point where the high delay for the class extremely threatens service. This second point can be seen in Figure 74 at $f_{ik}^{4} = 0.249 \cdot C_{ik}$ for class 4, at $f_{ik}^{3} = 0.331 \cdot C_{ik}$ for class 3, at $f_{ik}^{2} = 0.497 \cdot C_{ik}$ for class 2, and at $f_{ik}^{1} = 0.99 \cdot C_{ik}$ for class 1. At this point the traffic demand for the class nearly exceeds the perceived (available) capacity; for the limit of f_{ik}^{m} approaching C_{ik}^{m} the costs tend to infinity. Simultaneously the service is degraded and packet loss for the respective priority class is inflicted until the point where, in a heavily congested network, the class gets hardly any service. The third point, that is, the complete starvation of the flow of a class corresponds with the second point of the next higher priority class. To provide one example, Class-3 suffers starvation at $f_{ik}^{3} = 0.497 \cdot C_{ik}$ and goes out of service. At the same time all lower priority classes are completely doomed, to.

For the implementation of our routing algorithm the delay gradient may be upper bounded by a reasonable value, which does not restrict the choice of useful paths unnecessarily. From a networks perspective, the higher classes are always treated beneficially because their perceived capacity ${C_{ik}}^m$ provides for acceptable costs, which translate into feasible paths.

5.5 Experimental Performance Analysis

For the experimental analysis of our near-optimal multiclass minimum-delay routing algorithm, we follow the same methodological steps described in Section 4.4. We only briefly cover the similarities and concentrate on highlighting the differences in experimental design.

5.5.1 Experimental Design

We define the studied *system* to represent a routing network. The *Goal* of our evaluation is again a comparative performance study of our novel algorithm with a representative set of existing state-of-the-art routing algorithms. The algorithms studied should provide the *service* of efficient resource management ensuring resource-optimal network operation. Our *metric* is the achievable network performance, which is investigated using application related metrics such as end-to-end delay, goodput, and loss. Moreover, the coefficient of load variation and the individual link loads have been measured. We keep the type of routing algorithm to be the most important *predictor variable*, because we want to study the performance of our novel algorithm. We also include additional parameters in our study to obtain more general insights of the algorithms' behavior. This includes the investigation of our algorithm in three topologies of different size and of different graph characteristics. Moreover, the modeling parameters of our algorithm are varied in a subset of the tests. The secondary factors that are kept constant include the parameterization of the routers and links: these do only differ for the individual topologies studied. We give the corresponding factors in detail below. Similar to Chapter 4, we use simulation as *evaluation technique*. We refined our ns-2 implementation [HKS+03] to model our near-optimal multiclass minimum-delay algorithm and also added the feature of Vutukury's algorithm omitted earlier: the re-balancing of load at the short t_s intervals in combination with the metric of incremental delay. The changes in ns-2 are documented in [Mog04]. Since we examine topologies different from the RAN-I topology we depart from our realistic workload model and generate synthetic traffic. The *workload* is chosen to be synthetic and contains elastic and inelastic traffic. The detailed *experimental design* is described below, as well are the *analysis* and *interpretation of data* and *presentation of results*.

Table 8: Routing Strategies for the RAN-II Topology: Factors, Levels, and Descriptions.

Factor	Description		
	Routing	**Metric**	**Type**
Algorithm-*DST*	Static	Distance	Single-path
Algorithm-*SMP*	Static	Distance	Multipath
Algorithm-*MDR*	Dynamic	Incremental delay	Multipath
Algorithm-M^2DR	Dynamic	Incremental delay	Multiclass Multipath

Algorithms

We include four routing strategies in our analysis. As reference algorithms, we again use a shortest path algorithm and an equal cost multipath algorithm, both of which produce static routing tables and operate on a distance metric. We also include a full implementation of Vutukury's near-optimal minimum-delay routing as well as our own near-optimal multiclass minimum-delay routing algorithm. In particular, we include the following routing algorithms in our investigation (see Table 8 for a summary):

- Algorithm-*DST*, the static shortest-distance single-path algorithm introduced in Section 4.4.
- Algorithm-*SMP*, the static equal-cost shortest-distance multipath algorithm introduced in Section 4.4
- Algorithm-*MDR*, a near-optimal minimum-delay multipath algorithm. Algorithm-*MDR* is modeled to exactly match the algorithm proposed by Vutukury et al. [VGLA99] and can be seen as an extended version of Algorithm-*WMP*. We use the incremental delay as cost metric and include both time intervals for routing updates: t_l on which the algorithm computes new routes and t_s on which the algorithm computes new weights for the multipath scheduling.
- Algorithm-M^2DR, our near-optimal multiclass minimum-delay routing. Algorithm-M^2DR is a multipath algorithm and uses Algorithm-*MDR* as a solver for the minimum-delay routing problem. The incremental delay acts as cost metric and, in contrast to Chapter 4, both update mechanisms at t_l and t_s are implemented.

While there exist other promising QoS routing algorithms such as proposed in [NW02] or [CN99], we here limit our investigation to distributed link-state algorithms that do not keep per-flow state in the intermediate routers.

5.5.2 Topologies, Workload, and Experiment

We performed simulations in three different topologies with varying degree of alternate paths. The Cairn topology [CAI96] represents a realistic network topology with only a small number of alternate paths (see Figure 76 and Table 9). Vutukury's topology

[Vut01] has a higher number of alternate paths compared to the Cairn topology (see Figure 77 and Table 10). The RAN-II topology is an artificial topology providing for a high number of alternate equal cost paths (see Figure 78 and Table 11). The details of the individual topologies as well as the goals of the tests performed are described below. A summary of the experimental design is presented in Figure 75.

Using the *Cairn topology* tests we perform a thorough comparison of the performance of the studied algorithms for differing traffic elasticity. The workload has been chosen accordingly with various degrees of *elastic* (responsive to congestion) and *inelastic* (non-responsive to congestion) traffic. We parameterized Test (5.1) to Test (5.5) to use a different number of sources for elastic traffic (TCP) and inelastic traffic (UDP), respectively. Test (5.1) represents a TCP-only scenario, which is of particular interest to study the influence of the class based mechanisms on the TCP performance. In Test (5.2) and following, the degree of inelastic traffic is stepwise increased. See Figure 76 for a graphical illustration of the Cairn topology and Table 9 for a detailed description of all tests performed using this topology.

There are several other factors that are kept constant for all Cairn experiments. Link bandwidth and propagation delay of all links are set to be equal. We model the processing delay of the routers to be 0 s. For the single-class algorithms, the queue size at each router equals 80 packets. Algorithm-M^2DR is class-aware and uses 4 virtual queues each of size 20 packets instead of one queue. The simulation time has been chosen to be 310 s with the first and last 5 seconds serving as initialization period to start and stop the traffic flows. The class distribution of the traffic is symmetric, that is, each of the four classes carries 25% of the total traffic. We fixed the source/destination pairs of the individual flows as performed in [Vut01]. The traffic rates of the individual flows have been chosen to be randomly distributed between 100 kbit/s and 1000 kbit/s.

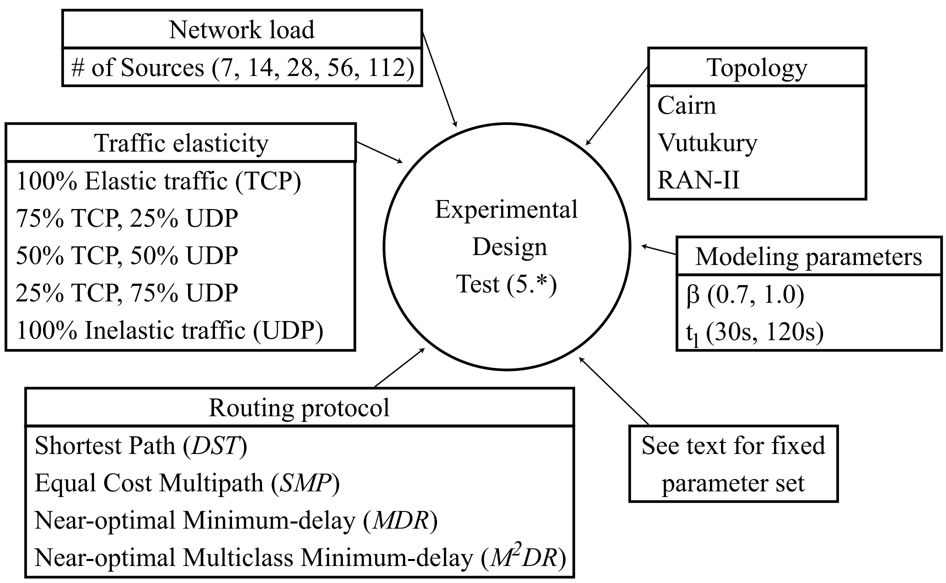


Figure 75: Experimental parameters for test-series (5.*).

Table 9: Experimental Parameter Set for the Cairn Topology Tests; Test-series (5.*).

Test	Test (5.1)	Test (5.2)	Test (5.3)	Test (5.4)	Test (5.5)
Topology	Cairn (see Figure 76), Graph order $p = 12$, Graph size $q = 34$, Link bandwidth = 3,000 kbit/s, Propagation delay = 0.2 ms.				
Fraction of traffic:					
inelastic traffic (UDP)	0%	25%	50%	75%	100%
elastic traffic (TCP)	100%	75%	50%	25%	0%

The common parameters for the Cairn topology tests are as follows.
Simulation: 20 replications per experiment, 310 s simulation time.
Router: Processing delay = 0 s, Queue size = 80 packets, Algorithm-M^2DR uses 4 virtual queues each of size 20 packets instead of one queue with 80 packets. The routing parameters are set to: $t_l = 30s$, $t_s = 3s$, $\alpha = 1, \beta = 1, \gamma = 0$.
Traffic: The source/destination pairs are fixed. We use (source;destination): (4;9), (5;7), (3;11), (0;8), (1;6), (2;8), (6;1), (7;5), (8;0), (9;2), (11;3). The traffic rates are chosen randomly and are equally distributed between 100 kbit/s and 1000 kbit/s. We use a fixed symmetric distribution of 25% for each traffic class $m \in \{1, 2, 3, 4\}$. The packet sizes are: 1000 byte + TCP-Header/UDP-Header + IP-Header.

The second topology used in our tests, *Vutukury's topology,* can be seen in Figure 77. The topology is used in [Vut01] for the simulation study of Algorithm-*MDR*. In contrast to the Cairn topology tests, we are especially interested to study different routing parameters during this setup. In Test (5.6) we start with a set of standard parameters: $\alpha = 1, \beta = 1, \gamma = 0$, which means, higher classes do not consider lower class traffic for the calculation of the class-dependent link costs, $t_l = 30s$, that is, each 30 seconds we calculate all routing graphs, and $t_s = 3s$ being the time-interval to reallocate the load on the existing successor set. In Test (5.7) we increase t_l by factor four. Since Algorithm-M^2DR performs route calculations for each class independently, this setup accounts for the increase of complexity due the class-based operation with four classes. We are particularly interested, if our algorithm has comparable performance with this reduction in routing updates compared with Algorithm-*MDR* in Test (5.6). In

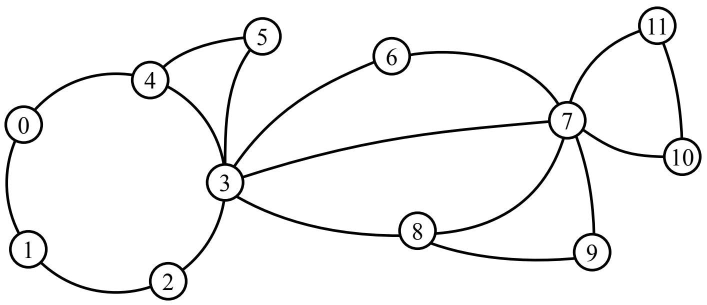


Figure 76: Cairn network topology. The topology provides for a low number of alternate paths. All links are of equal capacity and propagation delay. See Table 9 for the complete parameterization of the topology.

Table 10: Experimental Parameter Set for the Vutukury Topology Tests; Test-series (5.*).

Test	Test (5.6)	Test (5.7)	Test (5.8)
Topology	Vutukury (see Figure 77), Graph order $p = 19$, Graph size $q = 54$. Link bandwidth = 5,000 kbit/s, Propagation delay = 0.1 ms.		
Routing parameters	$\alpha = 1, \beta = 1, \gamma = 0$, $t_l = 30s$, $t_s = 3s$.	$\alpha = 1, \beta = 1, \gamma = 0$, $t_l = 120s$, $t_s = 3s$.	$\alpha = 1, \beta = 0.7, \gamma = 0$, $t_l = 30s$, $t_s = 3s$.

The common parameters for the Vutukury topology tests are as follows.
Simulation: 3 replications per experiment, 310 s simulation time.
Router: Processing delay = 0 s, Queue size = 80 packets, Algorithm-M^2DR uses 4 virtual queues each of size 20 packets instead of one queue with 80 packets.
Traffic: The source/destination pairs are chosen randomly. The traffic rates are chosen randomly and are equally distributed between 100 kbit/s and 1000 kbit/s. We use a fixed symmetric distribution of 25% for each traffic class $m \in \{1, 2, 3, 4\}$. The packet sizes are: 1000 byte + TCP-Header/UDP-Header + IP-Header.
Traffic elasticity: Each test has been performed using three different traffic mixes:
(1) 100% elastic traffic (TCP), 0% inelastic traffic (UDP),
(2) 50% elastic traffic (TCP), 50% inelastic traffic (UDP),
(3) 0% elastic traffic (TCP), 100% inelastic traffic (UDP).

Test (5.8) we study the influence of a reduced factor $\beta = 0.7$. Setting $\beta < 1$ should reduce the oscillations of Algorithm-*MDR*, because the links are loaded more conservatively. We perform the experiments for three different traffic mixes, each. Table 10 summarizes all tests performed using Vutukury's topology and gives the corresponding experimental parameter set.

The third topology investigated is the *RAN-II topology*. We again include three traffic mixes of elastic/inelastic traffic. The main purpose of the RAN-II topology is to study the algorithm performance in a setup with high number of equal cost paths. Additionally, we study the algorithms for different load situations. We start with a low load in

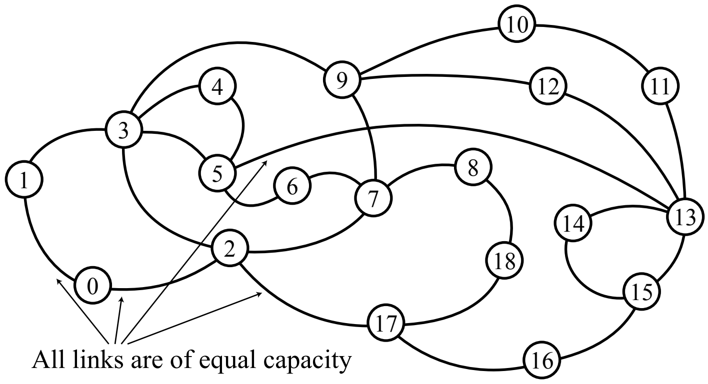


Figure 77: Vutukury's network topology. The topology provides for a moderate number of alternate paths. All links are of equal capacity and propagation delay. See Table 10 for the complete parameterization of the topology.

Table 11: Experimental Parameter Set for the RAN-II Topology Tests; Test-series (5.*).

Test	**Test (5.9)**	**Test (5.10)**	**Test (5.11)**	**Test (5.12)**	**Test (5.13)**
Topology	RAN-II (see Figure 78), Graph order $p = 27$, Graph size $q = 70$. Link bandwidth = 2,400 kbit/s, Propagation delay = 10 ms.				
Traffic load	7 flows	14 flows	28 flows	56 flows	112 flows

The common parameters for the RAN-II topology tests are as follows.
Simulation: 1 replication per experiment, 310 s simulation time.
Router: Processing delay = 0 s, Queue size = 80 packets, Algorithm-M^2DR uses 4 virtual queues each of size 20 packets instead of one queue with 80 packets. The routing parameters are set to: $t_l = 30s$, $t_s = 3s$, $\alpha = 1, \beta = 1, \gamma = 0$.
Traffic: The source/destination pairs are chosen randomly. The traffic rates are chosen randomly and are equally distributed between 100 kbit/s and 1000 kbit/s. We use two class distributions for each test: a fixed symmetric distribution of 25% for each traffic class $m \in \{1, 2, 3, 4\}$ and a fixed asymmetric distribution of 10% for class 1, 20% for class 2, 50% for class 3, 20% for class 4. The packet sizes are: 1000 byte + TCP-Header/UDP-Header + IP-Header.
Traffic elasticity: Each test has been performed using three different traffic mixes:
(1) 100% elastic traffic (TCP), 0% inelastic traffic (UDP),
(2) 50% elastic traffic (TCP), 50% inelastic traffic (UDP),
(3) 0% elastic traffic (TCP), 100% inelastic traffic (UDP).

Test (5.9) that is increased stepwise until we reach an overload situation in Test (5.13). We study an asymmetric class-distribution of traffic as well as a symmetric class-distribution for all tests. See Figure 78 for the RAN-II topology and Table 11 for the experimental parameter set. For all our experiments we use a random mix of constant bitrate, pareto, and exponential traffic.

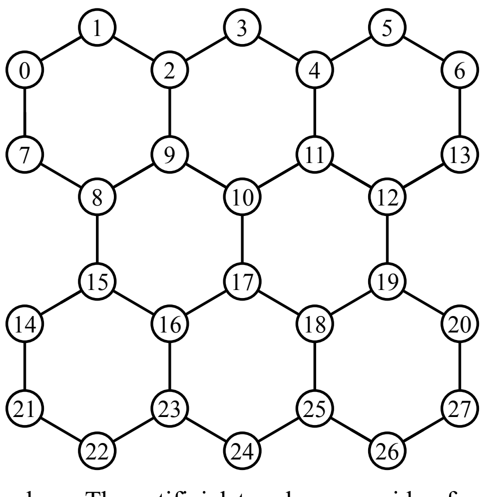


Figure 78: RAN-II network topology. The artificial topology provides for a high number of equal cost paths. See Table 11 for the complete parameterization of the topology.

5.6 Results

We have shown the feasibility of multipath QoS routing mechanisms for resource management in radio access networks in Chapter 4. The discussion of results for our novel algorithm covers the measurement of the network performance in more detail. The quantification of the influence of the predictor variables *multiclass behavior*, *network load*, *traffic elasticity*, and the influence of our *modeling parameters* is performed using the results obtained in the various sets of tests for the three topologies investigated. Moreover, a comparison of the influence of the *number of alternate paths* is performed by comparing the results over the individual topologies studied.

We structure the presentation of the results according to the three different topologies investigated. In Section 5.6.1 we present the results for the Cairn topology, where we compare the various *types of algorithms* studied. As response variables we present application characteristics including the *average end-to-end delay* as well as various *throughput related metrics*. For the Cairn topology we also study the impact of the *multiclass behavior* of our algorithm compared to the single-class routing behavior of the other algorithms tested. In Section 5.6.2 we investigate the sensitivity of our algorithms to different *modeling parameters* in Vutukury's topology. Finally, the results for the RAN-II topology are explained in Section 5.6.3. Here, we focus primarily on the behavior of the algorithms for *varying network load*. We, moreover, investigate different *distributions of the traffic classes*. We keep our response variables to be *throughput* and *delay*. For all experiments, we discuss the influence of *varying degrees of elastic/inelastic traffic*. A *summary* of the analysis and the obtained results is presented in Section 5.6.4.

5.6.1 Results for the Cairn Topology

Using the Cairn topology (see Figure 76 and Table 9) we perform a thorough study of the performance of the studied algorithms for different degrees of traffic elasticity. In Figure 79 we show the average delay with 95% confidence intervals in seconds over all traffic mixes for the tested algorithms. It is clearly visible that the end-to-end delay achieved using Algorithm-M^2DR is much better compared to the other algorithms. This decrease in delay is due to the prioritization of the higher class traffic, which is treated superior, while, at the same time, does only marginally suffer from packet losses. This comes at the expense of serving lower priority traffic worse. These results meet our expectations, because Algorithm-M^2DR burdens the low-priority classes with higher loss rates as well as higher round trip times by design. Also, the TCP-acknowledgements (ACKs) are sent in the same class as the original packets, which emphasizes this behavior. For the class unaware algorithms such a trade-off is not possible, hence all packets are treated equally (bad). It is worth to notice that the increase in delay for our algorithm with increasing amount of inelastic traffic is much lower than that of the other algorithms, too.

The shortest path algorithm, Algorithm-*DST*, and the equal cost multipath algorithm, Algorithm-*SMP*, produce nearly similar results. This is because the Cairn topology provides only for few alternate paths of equal cost (see Figure 76), which degrades Algorithm-*SMP* effectively to a single-path algorithm. Vutukury's algorithm, Algorithm-*MDR*, experiences the highest average delays of all algorithms. This seems counterintuitive for a minimum-delay routing algorithms on a first glance. However, in our simulations we experienced the problem of oscillations of Algorithm-*MDR*. These oscillations arise under very high load situations and lead to route flapping. Overloaded links to neighbors are pruned from the successor set because of their very high marginal delay even if the routes via these links are among the best routes towards the destination. The load is then shifted to newly established suboptimal paths. At the next re-routing interval, the pruned path is again included in the successor set. During our simulation, these oscillations have been clearly visible: periods of low delay and moderate loss interchange with periods of increased delay and high loss.

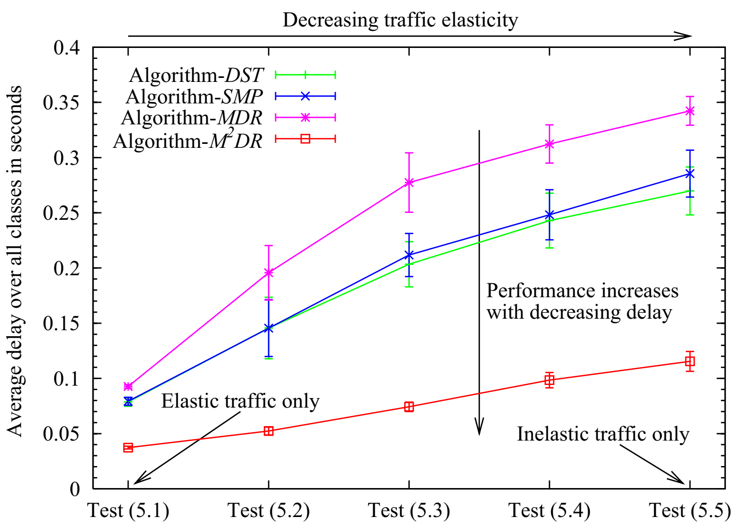


Figure 79: Average e2e delay and goodput for all algorithms in Test (5.1) to Test (5.5). The error bars give the 95% confidence intervals.

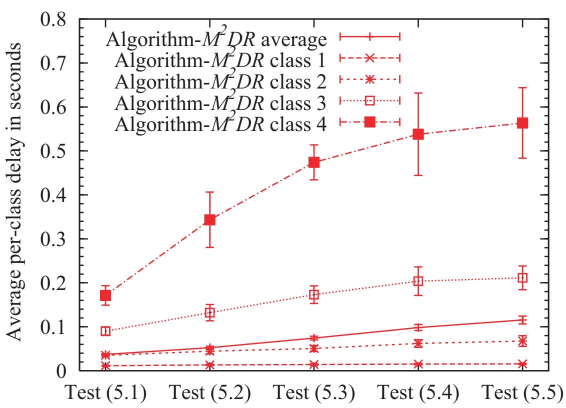


Figure 80: Per-class average e2e delay for Algorithm-M^2DR in Test (5.1) to Test (5.5).

For our highly loaded network we see a significant increase in delay if we move from elastic (TCP) to inelastic (UDP) traffic. This increase in delay is due to the non-existence of a feedback loop for inelastic traffic: since we use greedy sources, the network is being overloaded and operates at its capacity limits. Moreover, due to congestion, we see increasing packet losses with increasing level of inelastic traffic. For elastic traffic, the TCP feedback mechanism responds to this congestion situation and throttles the bandwidth of the flows. The *class-based* delay of Algorithm-M^2DR is shown in Figure 80. The figures clearly show the operation of our algorithm in combination with four priority classes. Higher class traffic experiences much better delay for all tests in combination with much lower loss rates. We can conclude that our algorithm is well suited for edge networks, such as provider radio access networks, where different priorities in traffic should be supported.

Figure 81 (a) depicts the average goodput, which is the fraction of packets delivered, of all algorithms for Test (5.1) to Test (5.5). The *class-based* goodput of Algorithm-M^2DR is shown in Figure 81 (b). For the case of elastic traffic only, the TCP mechanisms react to packet losses and, thus, all algorithms show nearly similar goodput in Test (5.1). It has to be noted that in this test Algorithm-*DST* and Algorithm-*SMP* deliver a higher absolute number of packets compared to the other algorithms as can be seen in Figure 82 (a) and (b). With increasing amount of inelastic traffic, the goodput decreases for all algorithms. Again the high load situation hinders Algorithm-*MDR* to perform as good as expected. The class-based comparison of the goodput for Algorithm-M^2DR shows that class 1 and class 2 packets are treated significantly better than the average of all packets. Class 3 and class 4 traffic suffers from the priority mechanism in place and show worse performance compared to Algorithm-*DST* and Algorithm-*SMP*. The increased amount of packets received for the higher amount of elastic traffic in Figure 82 (b) is due to the TCP acknowledgment packets (40 bytes acknowledgement packet vs. 1040 bytes for a data packet), which are also counted in our statistics. See Appendix B for additional results.

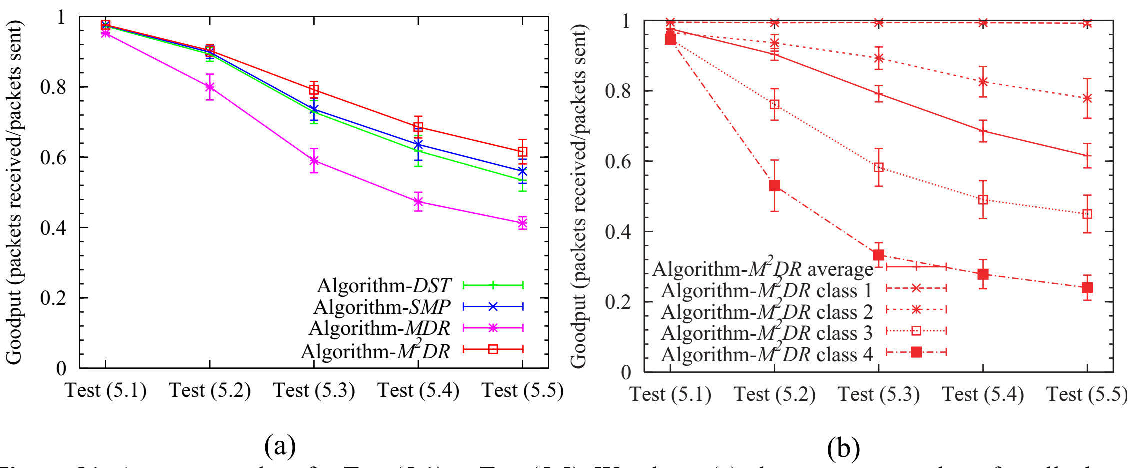
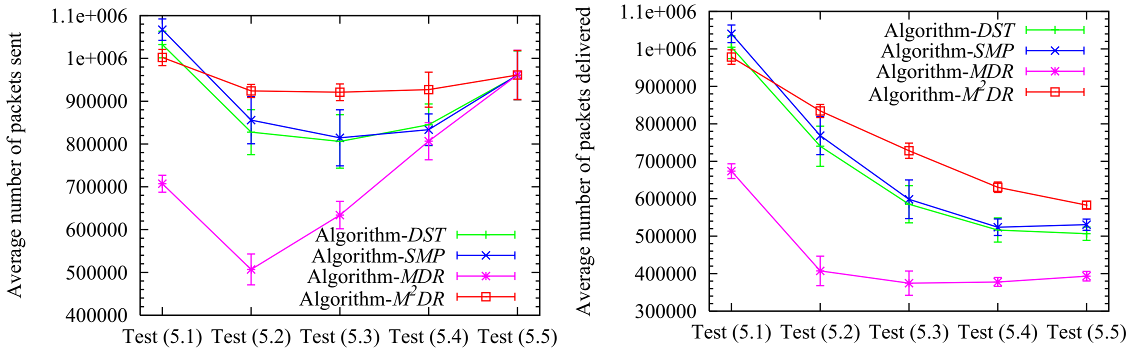


Figure 81: Average goodput for Test (5.1) to Test (5.5). We show (a) the average goodput for all algorithms and (b) the per-class average goodput for Algorithm-M^2DR.



(a) (b)

Figure 82: Average number of packets sent/delivered for Test (5.1) to Test (5.5). We show (a) the average number of sent packets and (b) the average number of delivered packets for all algorithms.

5.6.2 Results for Vutukury's Topology

Vutukury introduces a synthetic topology to analyze his algorithm in [Vut01]. Using this topology we concentrate on evaluating the influence of the *modeling parameters* of our algorithm. Since the topology has a higher number of alternate paths compared to the Cairn topology, the multipath algorithms are expected to outperform the single-class algorithms even more significantly. Moreover, we discuss the oscillations of Vutukury's algorithm. Again the experiments have been performed for different degrees of traffic elasticity; the network operates under very high load to stress the algorithms.

In Figure 83 (a), (b) and Figure 84 (a) we show the average delay of all investigated algorithms in Vutukury's topology. Additionally, for Test (5.6), we present the per-class delays of Algorithm-M^2DR in Figure 84 (b); it is clearly visible that the higher classes get a much better service compared to the lower classes. Since the individual replications performed for each test differ because of a random selection of traffic intensity for the fixed source/destination pairs we show each replication individually. Hence, only a direct comparison of the individual replications in the presented graphs is feasible.

In Test (5.7) and Test (5.8) we vary the update interval t_l and the modeling parameters β. These changes do not influence the algorithms operating on the metric hop count but only Algorithm-*MDR* and Algorithm-M^2DR shown in Figure 83 (a) and (b). The increase in t_l decreases the overhead of the algorithm, because the routing tables have to be calculated less often. From the perspective of computational complexity, the number of classes acts as a constant factor to the complexity of the underlying single-class algorithm. For example, for the four classes setup described, our algorithm needs to perform four times the number of calculations compared to Vutukury's single-class algorithm. To compare the algorithms performance, in Test (5.7), we increased the t_l duration by the factor four for our algorithm to obtain a similar complexity compared to Vutukury's algorithm for Test (5.6). For the described increase in t_l the performance of Algorithm-M^2DR remains nearly constant. In contrast, the trend for the average delay of Algorithm-

MDR is decreasing. This decrease is not expected but can be explained with the oscillating nature of the algorithm: the increasing t_l led to fewer mis-predictions in our particular setup and hence the delay decreased. The influence of the modelling parameter β of Test (5.8) is also shown in Figure 83 (a) and (b). Again the delay trend of Algorithm-*MDR* shows improvements in terms of average delay for a decrease in β because of less oscillations. See Appendix B for additional results.

In summary, Algorithm-M^2DR is able to perform very well while all other algorithms suffer from the high load condition in the network and produce much higher delays. Moreover, our algorithm is quite robust to larger update intervals because of our approach with possibly differing routing graphs for each class. These results stem from the inherent property of our algorithm to balance the load of different classes using dissimilar routing graphs per class and destination. Again, the low priority traffic for Algorithm-M^2DR suffers from higher loss and delay.

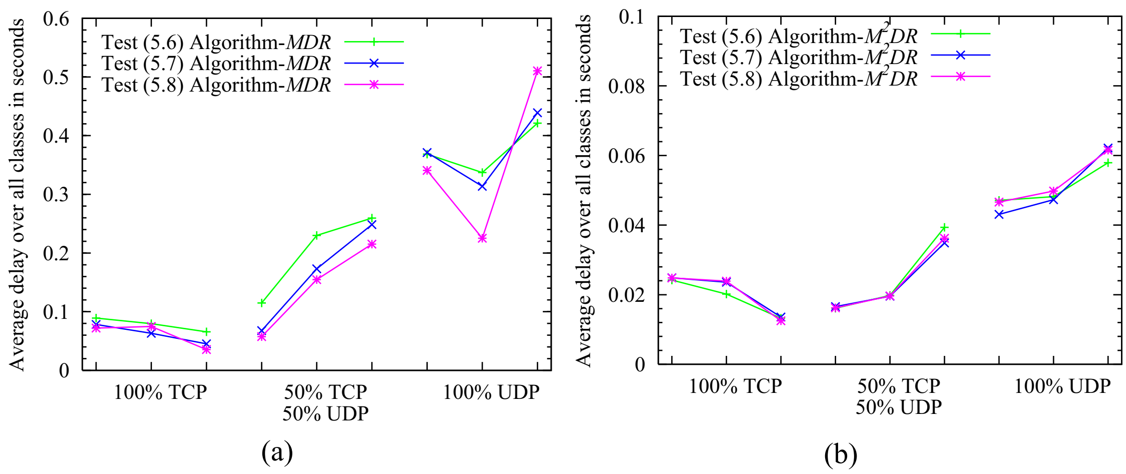


Figure 83: Average end-to-end delay for Tests (5.6) to (5.8). We show (a) Algorithm-*MDR* and (b) Algorithm-M^2DR. Please note the different y-axis.

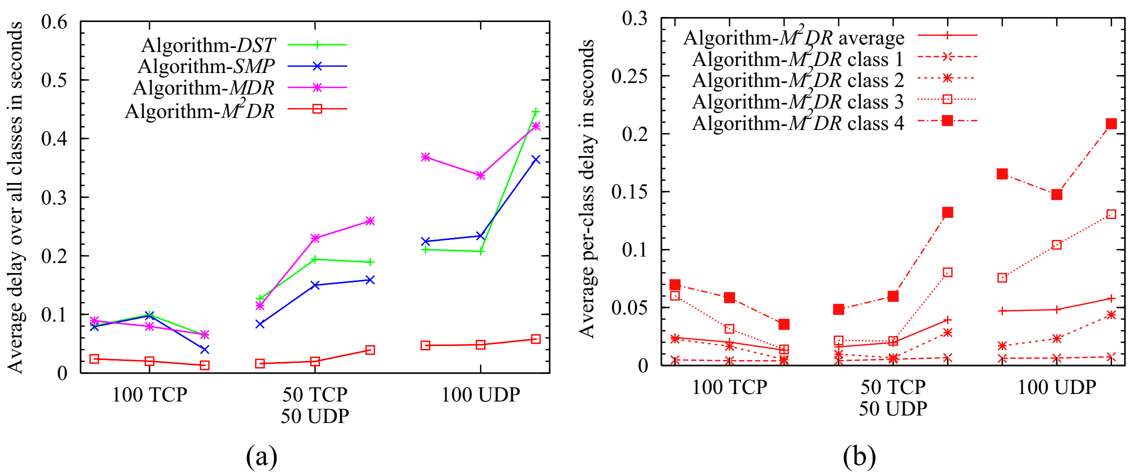


Figure 84: Average end-to-end delay for Test (5.6). We show (a) the average end-to-end delay and (b) the per-class average end-to-end delay. Please note the different y-axis.

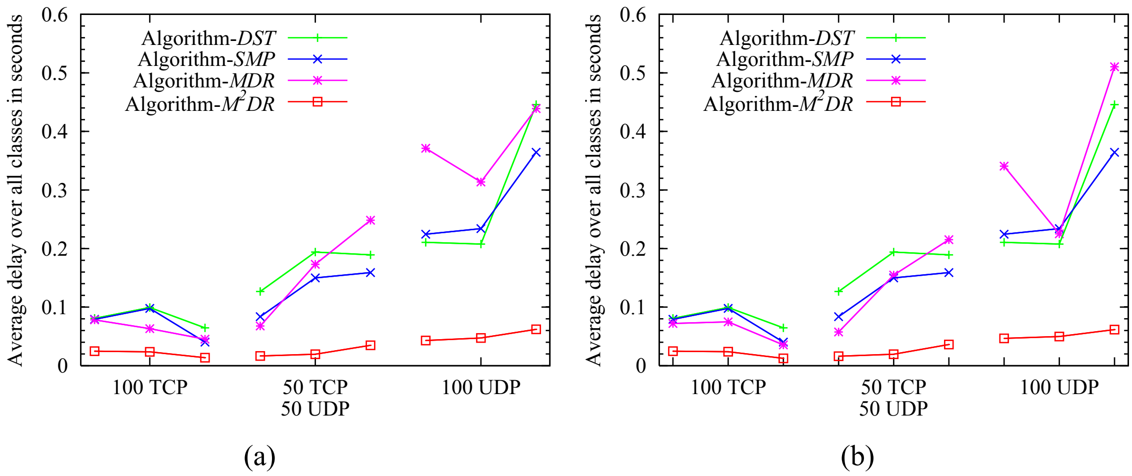


Figure 85: Average end-to-end delay for (a) Test (5.7) and (b) Test (5.8).

5.6.3 Results for the RAN-II Topology

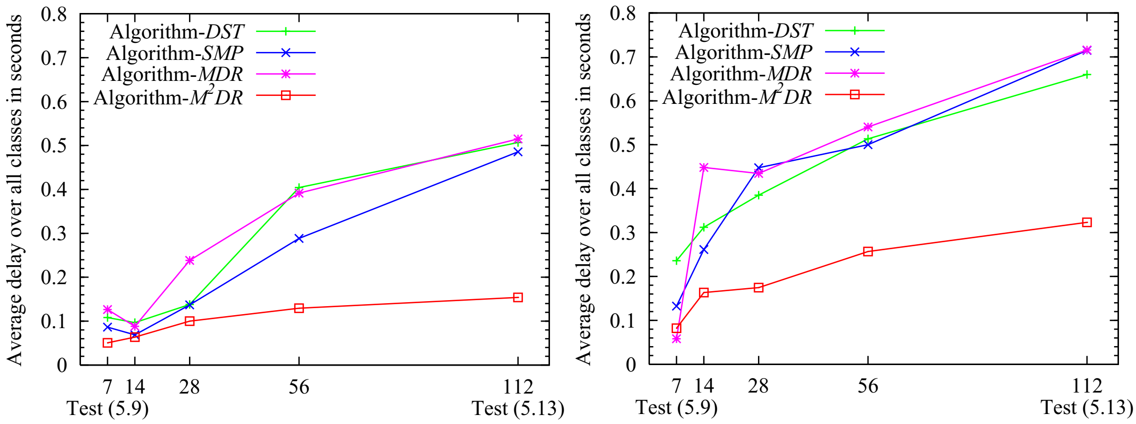
Using the RAN-II topology, we investigate the influence of the load on the network performance of the various algorithms and also study the influence of an asymmetric class distribution of traffic. We parameterized Test (5.9) to Test (5.13) to increase the number of sources in each test setup to be in the set {7, 14, 28, 56, 112}. Also, for all tests we use two class distributions of traffic: an asymmetric distribution (class 1 = 10%, class 2 = 20%, class 3 = 50%, class 4 = 20%) and a symmetric distribution (25% for each traffic class). Since the RAN-II topology provides for a high number of alternate paths, we expect all multipath capable algorithms to improve their performance compared to the single-path algorithms. Moreover, the ECMP algorithm is expected to perform well, because the regular topology has a high number of alternate equal cost paths.

In Figure 86 we show the average delay of the individual tests performed with asymmetric class distribution and (a) 50% elastic and 50% inelastic traffic and (b) inelastic traffic only. The increase of delay corresponds with an increase of load. Due to the random choice of source/destination pairs, some of the delay measurements do not follow this trend closely. Nevertheless, the general trend is clearly visible. Again, similar to the other topologies studied, the end-to-end delay achieved using Algorithm-M^2DR is superior compared to the delay of the other algorithms. We also notice that the increase in delay for our Algorithm-M^2DR is not as high as the increase for all other algorithms. Here, again, we can manifest that our algorithm is able to achieve load-balancing beyond the capabilities of traditional routing mechanisms because of the virtualization of the network into multiple classes.

The oscillatory behavior hinders Algorithm-*MDR* from reaching the superior steady state performance of a near-optimal algorithm. We obtain only minor changes in performance between the equal and unequal distribution of traffic classes. The visualization of the class-performance can be seen in Figure 87. Algorithm-*SMP* is class-unaware and, as a result, the measurements for the four classes show nearly similar behavior—the results only differ because of our random choice of source-destination pairs. In contrast, for Algorithm-M^2DR the class-based behavior is clearly visible.

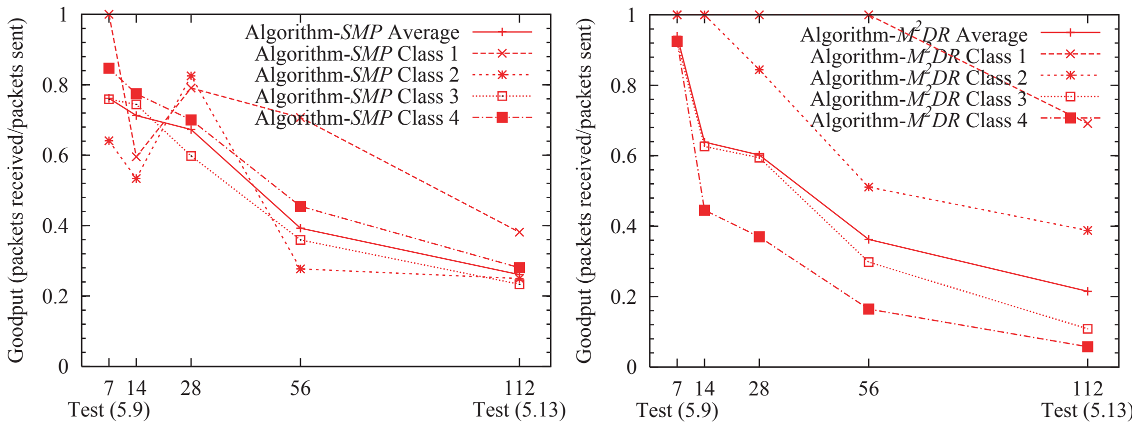
5.6.4 Summary of Results

Our analysis focused on the achievable performance region of our novel near-optimal multiclass minimum-delay algorithm, Algorithm-M^2DR. If we regard the important metrics average end-to-end delay and goodput, Algorithm-M^2DR clearly excels over the other algorithms. We can conclude that our routing algorithm in combination with the beneficial treatment of high-class traffic increases the perceived quality of service significantly. Moreover, we have been able to show that the oscillations of Algorithm-*MDR* can be significantly reduced using our approach. We observed that in low and moderate



(a) (b)

Figure 86: Average end-to-end delay for Test (5.9) to (5.13). We show (a) 50% elastic/50% inelastic traffic and (b) 100% inelastic traffic. The class distribution of traffic is class 1=10%, class 2=20%, class 3=50%, class 4=20%.



(a) (b)

Figure 87: Average per-class end-to-end delay for Tests (5.9) to (5.13). We show all traffic classes for (a) Algorithm-*SMP* and (b) Algorithm-M^2DR for asymmetric class distribution and 100% inelastic traffic.

load situations nearly all algorithms show a comparable performance. As the load increases, however, our algorithm gains performance advantages and is able to deliver a much better service to the higher classes. Moreover, the multipath algorithms are clearly favored compared with the shortest path algorithm as the number of alternate paths increases. For some additional results we refer to Appendix B and [Mog04]. In summary, our algorithm is well suited for edge networks, such as provider radio access networks, where different priorities in traffic should be supported. This improvement comes without the need for explicit signalling and without keeping per flow state in the network.

5.7 Related Work

In this chapter we touched two areas of related work, namely QoS routing and network architectures. The most interesting algorithms and concepts coupled with QoS routing have already been discussed in Section 4.6 together with some basic network architectures. We now shift our focus on advanced network architecture to support mobile communications. There exist various proposals for beyond 3G architectures. Here, we only regard approaches that follow multihop strategies or integrate variable topology concepts into current cellular architectures. Mainly focused on leveraging the capacity of the radio network, these architectures exhibit various problem areas, especially in the context of routing.

Cellular Architectures beyond 3G

The work of Lin and Hsu [LH00] describes "Multihop Cellular: A New Architecture for Wireless Communications" and is one of the first and most cited works that tries to integrate multihop paradigms with cellular network architectures. The main goal of the work is to minimize the costs for the wired infrastructure of a cellular network. The authors propose a *bases reduced* and a *power reduced* multihop cellular architecture, which builds on increasing the distance between two adjacent base stations or else reducing the transmission power of the cellular base stations. The end systems need to be able to act as multihop relays. Problems of the approach include that power or base reduction may result in only partial coverage of the network. Moreover, the authors provide no sufficient solution for routing problems, which may be coupled with the mobility of hosts and networks.

Wu et al. propose "iCAR, an architecture for Integrated Cellular and Ad Hoc Relaying Systems" in [WQDT01]. The main goal of the work is to reduce the call blocking rate and the call dropping ratio in cellular networks by using ad hoc relay stations. These stations are shifting traffic from congested cells to neighboring cells with spare capacity, thus, allowing for load-balancing within the network. iCAR assumes a relative dense mesh of relay stations, which can be seen as an overlay network of base-stations without fixed egress on top of the traditional cellular architecture. The overlay combines a structure of n cells to a virtual single super-cell so that the number of available data channels in each super-cell is n times that in a single cell. There are, however, multiple areas that limit the applicability of iCAR such as the use of multihop forwarding via the relays using special protocols, which provide only very restricted relaying functionality. Another limiting aspect of iCAR is the use of standard mechanisms within the "traditional" radio access network. In contrast to our work, the authors neither consider decentralized algorithms nor is their routing scheme able to balance the load within the scope of the radio access network.

The work of Yeh "ACENET: Architectures and protocols for High Throughput, Low Power and QoS Provisioning in Next-Generation Mobile Communications" [Yeh02], combines ideas from Multihop Cellular Networks [LH00] as well as those from iCAR [WQDT01]. In contrary to [WQDT01], the multihop functionality is not restricted to fixed relays at the boundary of the cell. In contrast to [LH00] the transmission range of the base stations and the number of base stations is not varied. ACENET provides full coverage like cellular networks and also includes a general routing and medium access control framework for the integration of various heterogeneous network technologies. The proposed routing algorithm is derived from both fixed networks as well as ad hoc networks. It may be summarized as a selective table-driven routing with different roles for the mobile devices in the system depending on the capacity and the ability of the node in question. Being derived from simple ad hoc algorithms, the ACENET routing is fairly restricted to finding a shortest path without neither giving guarantees nor considering resource management issues.

The work of Li and Lott et al. is about "Hierarchical Cellular Multihop Networks" [LLW+02] and "Multihop Communications in Future Mobile Radio Networks" [LWZ+03]. In [LLW+02] the general architecture is presented. In particular, the effects of relaying on the transmit power and capacity are derived. [LWZ+03] extends this work. Here, the authors assume a combination of fixed extension stations as well as multihop extensions of mobile nodes. A proprietary routing algorithm, which is based on dynamic source routing (DSR) [JMH04], is designed. This algorithm is executed only on base stations, which share full topology information. While the overall concept of this work is promising, the routing problem is only addressed in a very restricted manner.

Summarized, the related work in the area of future architectures for mobile communications introduces the important aspect of network variability. However, the routing problem and here especially the dependability aspect is neither studied nor solved adequately in the work surveyed.

Multiclass QoS Routing

We discussed single-class QoS routing algorithms in detail in the related work section of Chapter 4. In contrast, in the area of multiclass QoS routing algorithms there is only a small body of directly related work. This is especially true if we consider distributed algorithms only. To our knowledge the recent work of Wang and Nahrstedt is the first work that studies some problems of distributed routing algorithms in multiclass systems [NW02]. Starting from the observation that shortest path routing for the premium class in a DiffServ network [BBC+98] might lead to suboptimal routes for lower-class traffic, the goal of the work is to provide optimal paths for the premium class while the influences to other classes should be minimized. A major difference compared to our work is that [NW02] does follow a source-routing paradigm instead of a distributed one. The authors call their problem optimal premium-class routing (OPR) and derive a novel heuristic to solve the problem. In contrast, our work assumes a strict priority system to be in place, that is, each class is optimized without regarding lower class traffic.

5.8 Summary

In this chapter we addressed the problem of dependable routing for cellular networks from a perspective permitting variability of infrastructure components. We first addressed the design of a novel network architecture, and we second investigated feasible routing strategies for use within such networks.

Departing from our architecture developed in Chapter 4, we performed a thorough analysis of related work to capture the state-of-the-art in beyond third generation (B3G) architectures. Since the surveyed work was not flexible enough with respect to the synergy of the routing strategies with the underlying network architecture, we decided to introduce additional infrastructure components of different scope, namely wireless relays and wireless relay routers to form our variability-augmented cellular architecture. The merits of this architecture are clearly visible: the combination of decentralized routers in combination with the variable components allows for dynamic on-demand network reconfiguration if necessary. Our architecture is designed to allow for easy integration of heterogeneous access technologies and to allow for survivable, dependable, and self-organized operation of the network.

To leverage the potential power of our architecture we formulated a novel routing algorithm, which is tailored to fit our needs tightly. Our algorithm, Algorithm-M^2DR, capitalizes on the two orthogonal dimensions, path diversity of the network and class diversity of the traffic, respectively. Basic design criteria are self-adaptive and resilient operation to react to traffic and network dynamics. The possible scalability of the architecture by simply adding or removing components to the network is mirrored by the decentralized and distributed nature of the algorithm, which allows for scalability and extensibility. Our algorithm follows a multiclass multipath approach to minimize the overall network delay. To be quality of service aware, the algorithm allows to discriminate multiple classes of traffic, thus, providing proportional guarantees supporting real time and near real time applications. Each class is served with a logical separate routing graph, which is calculated for the virtual network reflecting the perceived capacity for the particular class. We have discussed the working principles of Algorithm-M^2DR in detail. By means of a comparative simulation analysis including various state-of-the-art routing algorithms, we have been able to show the superior performance of our algorithm. While providing for excellent routing performance, our algorithm maintains the simplicity of a decentralized and distributed routing algorithm, though.

"The price of reliability is the pursuit of the utmost simplicity.
It is a price, which the very rich find most hard to pay."
—Sir Antony Hoare

Part III
Dependable Routing for Ad hoc Networks

6 Analytical Model of the Route Acquisition Process

Starting with this chapter, we depart from the infrastructure based cellular networks described in Chapters 4/5 and look into the details of the self-organizing operation of mobile and wireless nodes in so-called ad hoc networks. While a number of experimental routing protocols has been designed, the quest for adaptability, scalability, quality of service, and security in routing for ad hoc networks has only recently started. Research in this area suffers especially from the lack of meaningful and realistic models, which may be used to approach the main goal of our studies: the analysis of routing dependability as described in Chapter 2. To overcome this problem, within this chapter, we propose an analytical model to describe the route acquisition process of realistic protocols in ad hoc networks. This model serves as a powerful, generically applicable tool for further analysis of ad hoc routing protocols. It is validated by means of simulation and instantiated for further usage. In Chapter 7 we extend the generic model presented in this chapter to cover the effects induced by node misbehavior and perform an extensive experimental analysis to study the effects induced on the routing system.

6.1 Motivation

The promise of self-organizing operation of mobile and wireless nodes gives rise to several interesting research challenges of which routing is a very prominent one. A number of experimental protocols have been designed; see, for example, the Ad Hoc On-demand Distance Vector Routing (AODV) by Perkins et. al [PBRD03], the Dynamic Source Routing (DSR) by Johnson et al. [JMH04], Jaquet's Optimized Link State Routing (OLSR) in [CJ03], and Blazevic's geographic Terminode Routing (TRR) in [Bla02]; Feeney [Fee99] as well as Royer and Toh [RT99] present fairly comprehensive taxonomies of ad hoc routing.

One key concept of ad hoc routing protocols is the adaptability to the decentralized and highly dynamic nature of these networks, induced by mobility (see Chapter 3) and the nature of the wireless medium. Despite the fact that extensive work is being performed in this area, large-scale ad hoc networks are not currently available for civilian applications. One of the key problems, as yet unsolved, is the dependable and scalable operation of ad hoc routing. From a research perspective, there is a lack of tools to sufficiently approach the adaptability and dependability problems tied to ad hoc networks. On one hand, there are many simulation-based studies that investigate special but restricted scenarios. These do not allow for easy generalization of the results. On the other hand, some fairly restricted analytical models exist. Most of these do not account for special routing protocols but assume ideal assumptions. To allow both, easy generalization of the results while accurately matching realistic protocol behavior, we need a combination of the models described above: Analytical models of realistic routing protocols. Only few such models exist, however.

Based on existing work we close this gap and formulate and validate an analytical model to describe the route acquisition process of realistic ad hoc routing protocols. Our model acts as basis for our investigation of routing dependability in ad hoc networks. We capture the characteristics of the AODV protocol described in [PBRD03], but our model may be extended to other routing protocols as well. Within this chapter, we provide:

- An analytical model of an *ideal route acquisition process*.
- An extension to cover *transmission errors* of the wireless channel.
- An extension to our model to describe a *realistic ad hoc routing protocol*. In particular, we have chosen the *basic AODV protocol*.
- Extensions to describe various protocol optimizations of AODV: *probabilistic flooding (gossiping)* [HHL02], *expanding ring search (ERS)*, and *reply by intermediate* [PBRD03].
- The *experimental validation* of these models.

Our results enable the prediction of the route-length distribution inside the network, which characterizes the overall network behavior. We regard this knowledge as crucial for the study of various issues that hinder the further evolution of ad hoc networks to reach a critical mass of wide deployment, namely the adaptability, scalability, and dependability of ad hoc routing protocols.

6.2 Outline

In order to give a precise description of the individual steps of modeling, we present our findings in two steps: first, the modeling of an ideal routing protocol; second, the extension of this basic model to cover realistic protocols in use today. The basic assumptions as well as the modeling of an ideal routing protocol are laid down in Section 6.3. In Section 6.4 we extend our model to reflect the behavior of wireless networks and actual routing protocols in greater detail, starting with the modeling of transmission errors in Section 6.4.1. Modeling of the basic features of the experimental AODV protocol [PBRD03] is described in Section 6.4.2. We further enhance the model to cover *probabilistic flooding* mechanisms for route discovery as proposed from Li et al. in [HHL02] and Sasson et al. in [SCS03]. We validate both models in Section 6.4.3. Moreover, in Section 6.4.4 and following, the protocol features *expanding ring search* and *reply by intermediate* [PBRD03] are added to the model. All models are validated by means of simulation and instantiated for our specific needs. The main results are summarized in Section 6.4.8 while a survey of related work, mainly covering analytical models, is performed in Section 6.5. We finish by outlining the merits and weaknesses of our analytical model in Section 6.6 and point to Chapter 7, which describes the dependability specific part of the model. The structure of this chapter is shown in Figure 88.

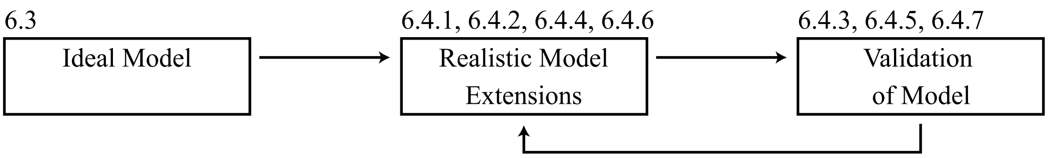


Figure 88: Structure of Chapter 6.

6.3 Modeling of Ideal Routing Protocols

There exists a large body of work to describe the basic properties of wireless multihop networks (see also Chapter 2 for important characteristics of ad hoc networks). An early and very prominent one was developed in the context of packet radio networks by Kleinrock and Sylvester [Kle78] [KS78]. We use this work as a foundation for our model. In contrast to the work of Kleinrock et al., which studies the performance and capacity of wireless multihop networks under ideal conditions, we are interested in meaningful metrics for the analysis of routing dependability. We want to obtain deeper insights on the overall network behavior for dedicated protocols—with a higher abstraction level than currently available through simulation.

Since the route acquisition process constitutes the essential behavior of ad hoc routing protocols we chose it to be the nucleus of our model. The primary metric of the model needs to sufficiently describe the overall network behavior while being highly responsive to abnormal network conditions including node misbehavior. In particular, we have chosen the distribution of route lengths as the metric for our study. This metric describes the statistical relation between the distance of two nodes inside the modeled area and the corresponding probability of being connected using a multihop route. In our case this distribution of route lengths is able to predict how many stable routes of certain lengths can be established under various conditions. Our metric operates on the routing layer and implicitly includes the effects induced by lower layers such as wireless losses, interference or shared medium access. Partly related work in the area of fixed networks that uses the same metric does exist, however. In [ZCD97] Zegura et al. introduce the metrics “length distribution” and “hop-length distribution” to compare graph-based models for Internet topology.

In the area of ad hoc networks there exists only few work directly related to ours. Recently, a model for ideal source routing was studied by Kail, Nemeth, et al. in [KNT01], [NTV01]. In parallel to our work, the work of Miller [Mil01] derived the distribution of link distances in a wireless network, which is part of our model as well. In [Mil01] and the follow-up work from Mullen [Mul03] more general results for the link-distance distribution than ours are provided. However, both works assume a single-hop connection between source and destination and omit the modeling and validation of realistic protocols. A more comprehensive survey and classification of related work is given in Section 6.5 below. In the following we describe our model.

6.3.1 Assumptions

Our model utilizes the geometrical properties of the ad hoc network. To describe networks of arbitrary connectivity, property, and size, we make the length parameters dimensionless. As shown in Chapter 3, synthetic models are not able to represent user mobility as observed in reality. Moreover, the distribution of the location of nodes is changing over time. However, to keep the closed forms of the route-length distribution at reasonable complexity, our model uses a uniform distribution. In [Mul03] Mullen also investigates non-uniform distributions as observed, for example, for the random-way-point model. Additionally, he gives simple approximations and describes the error of the model for these approximations and non-uniform node distributions.

Ad hoc networks are mostly considered to be mobile. In our analysis we do, however, regard quasi-static networks to simplify the analysis. In summary, we use the following set of assumptions:

- The investigated area A is a normalized square of side length 1.
- The x and y coordinates of the nodes are independently and identically uniformly distributed in the interval [0,1]. We denote the respective probability density functions (pdf) as $p_x(\alpha)$ and $p_y(\beta)$, α and β being random variables.

- The system consists of a set N of nodes, the number of nodes being $|N| = n$. The nodes share a uniform transmission range r, which is considerably smaller than the side length of the square.
- The nodes are not in motion.

Figure 89 shows a dimensionless network covered by our model and highlights the most important geometrical properties used within the following analysis. Hereby, the visualization is based on a network with 500 nodes, an area of 3,300 times 3,300 square meters, and a transmission range of 250m. As it may be noticed, below we use the same dimensions in some of the simulations performed.

6.3.2 Modeling of Ideal Source Routing

Performing an ex-post analysis of the geometrical node distribution, we find the area A_0 a node covers related to the entire area of investigation A to be

$$A_0 = \frac{A}{n}. \quad (39)$$

The radius r_0 of a circle and the side length b_0 of a square equivalent to A_0 are

$$r_0 = \sqrt{\frac{A}{n\pi}} \text{ and } b_0 = \sqrt{\frac{A}{n}} \text{ respectively (see Figure 89).} \quad (40)$$

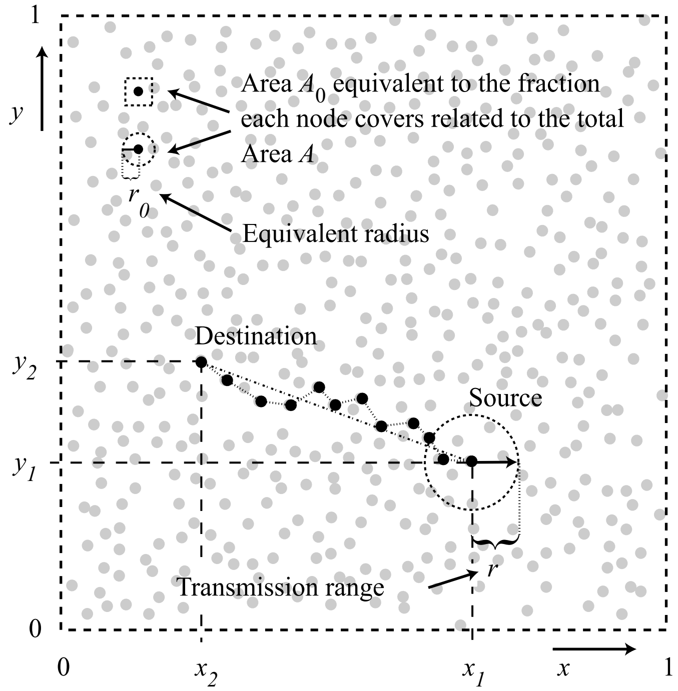


Figure 89: Normalized sample network of the modeling process. The highlighted nodes, areas, and distances represent the most important properties of the network.

The expected number of nodes in a transmission radius at any point, M (also coined the average degree of the network in [KS78]), can easily be obtained by

$$M = \frac{\pi r^2}{A_0} = \left(\frac{r}{r_0}\right)^2. \tag{41}$$

It is intuitive that the average number of neighbors, that is, the average degree of one node, equals $M-1$. Knowing of M allows one to predict how many nodes are influenced, on average, if one node transmits a signal. We are particularly interested in route lengths. Starting with idealized source routing, a first approximation of the shortest path between two nodes follows the direct line between them, assuming a very large or infinite number of nodes (see Figure 90 for the corresponding visualization). Thus, the estimated length of hops h between two nodes is a function of the geometric distance d.

The routing protocol uses neighboring nodes to transmit the packets from source to destination. The progress a packet makes in each step can be modeled as follows: Nodes are assumed to be connected directly if they are in range r of each other. The median distance, r_1, between two nodes that can reach each other is

$$r_1 = \frac{r}{\sqrt{2}} \text{ (see Appendix C for the derivation of } r_1\text{).} \tag{42}$$

If d is sufficiently large compared to r, we can approximate the average progress per routing step as r_1. As a result, the distance between source and destination is

$$d = h(d)r_1. \tag{43}$$

The geometric distance between two nodes on a plane is also given by the Euclidean distance between their positions. So for Node1 (x_1, y_1) and Node2 (x_2, y_2) the distance is

$$d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} \text{ or } d = \sqrt{\Delta x^2 + \Delta y^2} \tag{44}$$

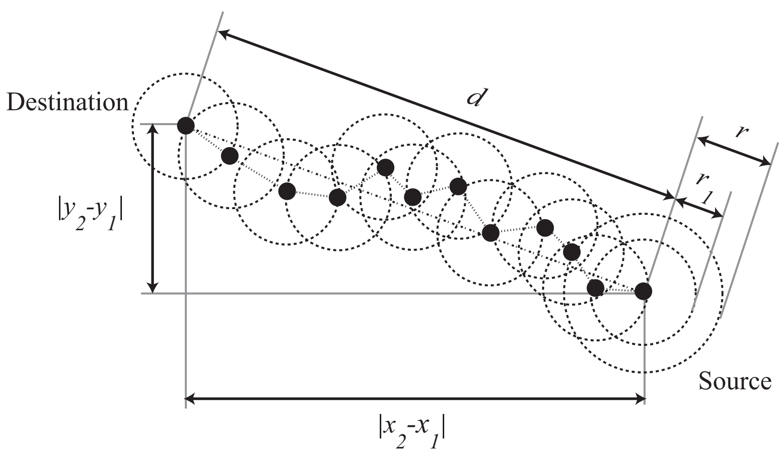
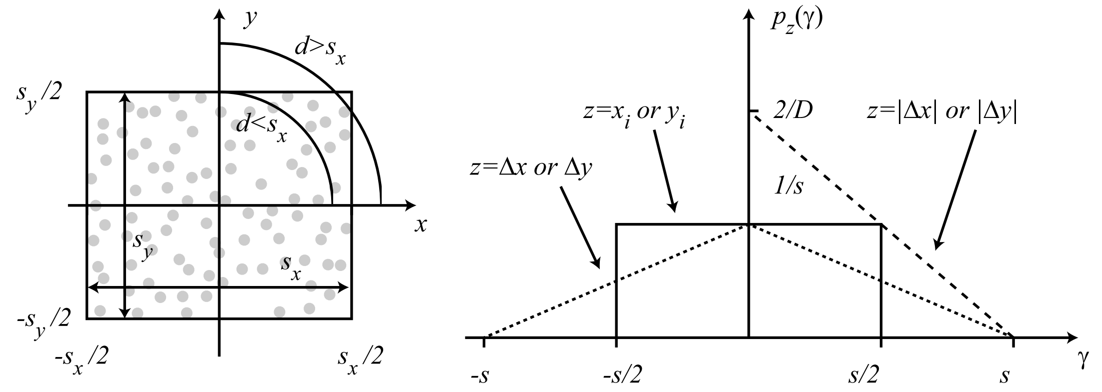


Figure 90: Geometrical measures within the sample network. The distance d between source and destination can be expressed using the hopcount $h = 11$ and the radius r_1.



(a) (b)

Figure 91: Domain of integration and probability densities in the modeling process. The figure shows (a) a rectangle with uniformly distributed nodes and the domain of integration, and (b) probability density functions for location, difference of location, and the absolute value of difference.

We replace the explicit positions in Equation (44) using their probability density. Communication can only take place within our area of investigation. For the general case of a rectangle with side lengths s_x and s_y as shown in Figure 91 we obtain the probability density functions $p_x(\alpha)$ and $p_y(\beta)$ for the uniform distribution of the x and y coordinates:

$$p_x(\alpha) = \frac{1}{s_x} \text{ for } |\alpha| \le \frac{s_x}{2} \text{ and } p_y(\beta) = \frac{1}{s_y} \text{ for } |\beta| \le \frac{s_y}{2}\text{, else } p_x(\alpha) = p_y(\beta) = 0 \tag{45}$$

Due to the independence of the x and y coordinates, the probability densities of the differences of locations are also independent and given by

$$p_{\Delta x}(\alpha) = \frac{(s_x - |\alpha|)}{(s_x)^2} \text{ for } |\alpha| < s_x \text{ and } p_{\Delta y}(\beta) = \frac{(s_y - |\beta|)}{(s_y)^2} \text{ for } |\beta| < s_y\text{, else}$$

$$p_{\Delta x}(\alpha) = p_{\Delta y}(\beta) = 0. \tag{46}$$

For our case, $s_x = s_y = s$. The integration of the probability densities of the geometrical distance, $p_d(\xi)$, $d = \sqrt{\Delta x^2 + \Delta y^2}$ equals the cumulative probability distribution function (cdf) $P_d(\xi)$. The cumulative probability distribution gives the probability P that two nodes are at most the distance d apart from each other. Hence we are able to obtain the cdf for the link distance between two nodes. Figure 91 (a) visualizes domain of integration while Figure 91 (b) shows the probability densities for our case, both of which are adopted from [Mil01].

$$P_d(\xi) = P\{d \le \xi\} = P\{\sqrt{\Delta x^2 + \Delta y^2} \le \xi\} = \iint\limits_{(\alpha,\beta)} p_{\Delta x, \Delta y}(\alpha, \beta)\,d\alpha\,d\beta \tag{47}$$

The lengthy evaluation of the integral, where various cases need to be considered, is documented in detail in [Mil01]. As a result we obtain the distribution $P_d(\xi)$ in closed form, which is used to calculate the model predictions.

$$P_d(\xi) = \begin{cases} \frac{1}{2}\xi^4 - \frac{8}{3}\xi^3 + \pi, & \xi \le 1 \\ 4\sqrt{\xi^2 - 1} + \frac{8}{3}\sqrt{(\xi^2-1)^3} - \frac{1}{2}\xi^4 - 2\xi^2 + \frac{1}{3} & \\ +2\xi^2\left(\arcsin\left(\frac{1}{\xi}\right) - \arccos\left(\frac{1}{\xi}\right)\right), & 1 < \xi \le \sqrt{2} \end{cases} \quad (48)$$

By differentiation, we finally obtain the probability density function $p_d(\xi)$ to be

$$p_d(\xi) = \begin{cases} 2\xi(\xi^2 - 4\xi + \pi), & \xi \le 1 \\ 8\xi\sqrt{\xi^2 - 1} - 2\xi^3 - 4\xi + 4\xi\left(\arcsin\left(\frac{1}{\xi}\right) - \arccos\left(\frac{1}{\xi}\right)\right), & 1 < \xi \le \sqrt{2} \end{cases}. \quad (49)$$

This serves as the central equation for our model. It describes the statistical relation between the distance of two nodes inside the unit square and the corresponding probability of being connected.

6.4 Modeling of Realistic Protocols

In the previous section we transferred existing results to our application domain. Starting from here, we formulate our novel model to describe realistic ad hoc protocols.

6.4.1 Modeling of Transmission Errors

The basic model assumes that there are no transmission errors on the lower layers (idealized physical and link layer without losses). A wireless medium is, in reality, often shared competitively. To describe this behavior analytically, we relax the error condition by introducing the success probability for each packet transmitted. Our model takes a global approach instead of a local one and sets all success probabilities to be equal (note that under some circumstances the probability may be near zero, as we show later).

For a single hop and a given loss probability q the success probability is $1 - q$. If we assume a multihop route consisting of h hops, the success probability is $(1 - q)^h$. Our base model assumes a success probability of 1. Consequently, errors are modeled using a correction term, which depends on the distance of the path. We introduce $(1 - q)^{h(d)}$, describing the success probability as a function of $h(d)$. We are now able to obtain the probability density functions for both; the successfully established routes and the routes hindered to be established, respectively. We are interested in a direct comparison of the number of routes between the error-free case and the case with errors, in order to predict the routing performance. Thus, besides the probability density as shown in

Equation (50), we give a probability measure, which accounts for the routes remaining unaffected by the error condition (see Equation (51)).

$$p_d(\xi) = \begin{cases} (1-q)^{h(\xi)} 2\xi(\xi^2 - 4\xi + \pi) \\ \quad + (1-(1-q)^{h(\xi)}) 2\xi(\xi^2 - 4\xi + \pi), & \xi \le 1 \\ (1-q)^{h(\xi)} (8\xi\sqrt{\xi^2 - 1} - 2\xi^3 - 4\xi) \\ \quad + (1-q)^{h(\xi)} 4\xi\left(\arcsin\left(\frac{1}{\xi}\right) - \arccos\left(\frac{1}{\xi}\right)\right) \\ \quad + (1-(1-q)^{h(\xi)})(8\xi\sqrt{\xi^2 - 1} - 2\xi^3 - 4\xi) \\ \quad + (1-(1-q)^{h(\xi)}) 4\xi\left(\arcsin\left(\frac{1}{\xi}\right) - \arccos\left(\frac{1}{\xi}\right)\right), & 1 < \xi \le \sqrt{2} \end{cases} \tag{50}$$

$$pm_d^{success}(\xi) = \begin{cases} (1-q)^{h(\xi)} 2\xi(\xi^2 - 4\xi + \pi), & \xi \le 1 \\ (1-q)^{h(\xi)} (8\xi\sqrt{\xi^2 - 1} - 2\xi^3 - 4\xi) \\ \quad + (1-q)^{h(\xi)} 4\xi\left(\arcsin\left(\frac{1}{\xi}\right) - \arccos\left(\frac{1}{\xi}\right)\right), & 1 < \xi \le \sqrt{2} \end{cases} \tag{51}$$

6.4.2 Modeling of AODV and Gossip-enhanced AODV

We now refine our model to include realistic routing protocols. As a consequence, the routes discovered differ from the ideal routes. AODV is a reactive routing protocol, that is, routes are discovered on-demand if necessary. All nodes in the network share uniform responsibilities and the routing process is based on the distance vector routing principle. Route acquisition follows a cycle: first, the source node broadcasts a route request (RREQ), which then is flooded through the network. A RREQ must never be broadcasted more than once by any node. At the same time, the intermediate nodes set up a reverse path pointing towards the source. Second, if the RREQ reaches the destination, a unicast route reply (RREP) is issued and sent along the reverse path of the RREQ. For route maintenance, each node uses timers to keep the route alive. The reverse path is purged after a time-out interval, which should be chosen sufficiently large to allow for the RREP to reach the source. Likewise, a routing table entry maintaining a forward path is purged if not used for a certain time-out interval. If no data is being sent using a particular routing table entry, this entry is deleted from the routing table even if the route may still be valid. Route error messages (RERR) are sent towards the source via unicast if a route break is discovered.

AODV imposes no overhead on data packets because the intermediate nodes maintain routing tables. Destination sequence numbers are employed to achieve freedom of loops as well as to solve various inherent problems of distance vector protocols. The characteristics of AODV make it most suitable for networks that can absorb a network wide broadcast rate and wherein the routing churn is high enough that proactive maintenance of routes is unproductive. See [PBRD03] for more details on the operation of the AODV protocol.

For our model, if we take the AODV protocol as an example, we obtain non-optimal routes, due to the loop-freedom criterion. Here, a node that receives a RREQ on a longer but faster route forwards this non-optimal request. Hence, the forward routing graph that is spanned according to the propagation of the routing messages is likely to be sub-optimal in the case of AODV. The route length increases due to this effect. For large networks this elongation of routes may be described by a factor θ. This factor acts multiplicatively on the geometrical distance d. All routes discovered can be described using $d' = \theta d$. Since routes may not be any shorter than the optimal distance, $\theta \geq 1$ for all protocols. θ depends on the routing protocol variant used. The curve resulting from Equation (51) appears contracted in the y-axis by the factor θ, and stretched in the x-axis by the factor θ. Equipped with this refined model, we are now able to directly compare the analytical results with simulation results, using θ as correction term for the protocol chosen. We obtain $h(d) = d'/r_1 = d\theta/r_1 = d(\theta/r_1) = d/r'_1$ where $r'_1 = r_1/\theta$.

Recently, various performance optimizations for the AODV protocol have been proposed. Within our analysis, we include an optimization of the flooding mechanisms used within the AODV route discovery cycle, that is, probabilistic flooding or gossiping (see [HHL02], [SCS03], and [Har03]). Instead of re-broadcasting each routing request received, the nodes only forward the requests with a certain probability. The choice of an appropriate probability is subject of heuristics and largely depends on the network properties. Targeted to lower the network load imposed during route discovery, the inherent property of the probabilistic mechanisms to skip some of the nodes increases the route lengths significantly.

We use the same approach as described above for AODV to model the gossip-enhanced AODV. The main difference between the two protocol variants is that non-optimal routes can rarely be observed within standard AODV while they can be observed more frequently using the probabilistic mechanisms. Our experiments show that θ for gossip-enhanced AODV [HHL02] is significantly larger than θ for pure AODV [PBRD03]. For a gossiping probability approaching 1, θ is the same for both protocol variants.

6.4.3 Experimental Validation of AODV and Gossip-enhanced AODV

To allow for experimental validation, we extended the Qualnet® network simulator to comply with the experimental AODV standard [PBRD03]. Table 12 provides an overview of the parameter set for the simulations used for experimental validation within this chapter. The most important parameters of the experimental design are also shown in Figure 92. We simulate stationary nodes to fulfil our model assumption of non-mobility. In our simulations we use IEEE 802.11b in distributed coordination function (DCF) mode as lower layer protocol [IEEE01].

Table 12: Experimental Parameter Set for Ad hoc Routing Model; Test-series (6.*).

Test	**Test (6.1), AODV**	**Test (6.2), Gossip**	**Test (6.3), AODV**	**Test (6.4), Gossip**	**Test (6.5), AODV)**	**Test (6.6), Gossip**
Duration	5050 s		5050 s		1500 s	
Gossip probability	1	0.7	1	0.7	1	0.7
Expanding ring search	no	no	yes	yes	yes	yes
Reply by intermediate	no	no	no	no	yes	yes
Traffic (# of generated routing requests)	every 10 s one flow (500)		every 10 s one flow (500)		every 500 ms one flow (3000)	

The common parameters for the test-series 6.* are as follows.
Simulation: 20 replications per experiment.
Mobility: no mobility.
Nodes density/placement: 500 randomly placed nodes. The size of the simulation area is $(3300{,}94\text{ m})^2$, the average density $M = 9$. The x and y coordinates are independently and identically uniformly distributed.
Traffic: The source/destination pairs are randomly selected. One packet is sent per flow using UDP as transport protocol. Packet size = 512 bytes.
Physical layer parameters: Propagation model = free space, transmission power = 7dBm, transmission range $r = 249.862$m, $r_0 = 83.287$m, $r_1 = 176.679$m.
Link layer parameters: MAC 802.11b with Distributed Coordination Function (DCF), max. transmission rate = 11 Mbit/s.
AODV parameters: Local repair = deactivated, hello messages = deactivated, net diameter = 35.

The model assumptions for the basic model have been validated using Test (6.1) for AODV and using Test (6.2) for gossip-enhanced AODV. Since we are mainly interested in investigating the distribution of route lengths, we generated a series of single packets. The rationale behind this configuration is to trigger route discoveries without loading the network unnecessarily. Using the protocol type set {AODV, gossip-enhanced AODV} as a predictor variable, the simulations validate our basic model. Moreover, we obtain a first estimate for θ. The possible *reply by intermediate* (see Section 6.4.6) was avoided by setting the pause time between the individual route requests to 10 seconds. AODV invalidates the cached reverse paths within this period. We measured the length and

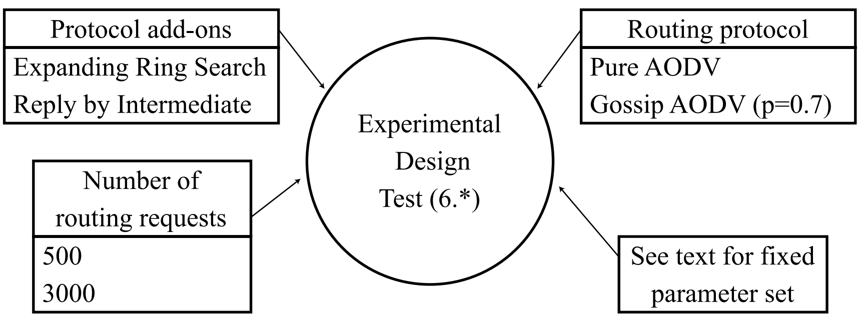


Figure 92: Experimental parameters for test-series (6.*).

number of valid routes as response variable as shown in Figure 93 for Test (6.1). The mean values were obtained in 20 simulation runs with different seeds. Unless stated otherwise, we also calculated and present the two-sided 95% confidence interval of all the data obtained experimentally.

We use the least-squares method to obtain the fitting parameter θ of the curve described by Equation (51) for the measured data. Routes longer than $h = 22$ hops for AODV are not used in the fitting. These routes can be considered mostly unstable and only account for less than 5% of all routes acquired. The application of Equation (51) produces Figure 94 (a) for Test (6.1) and Figure 94 (b) for Test (6.2). We obtain a good fit above $h = 5$ hops while for small h the measured values are too high. This can be explained by the average transmission range. Our initial assumption was that the number of hops h multiplied by the average progress per hop r_1 equals the estimated distance. For routes longer than 5 hops this holds. For neighboring nodes, however, the destination

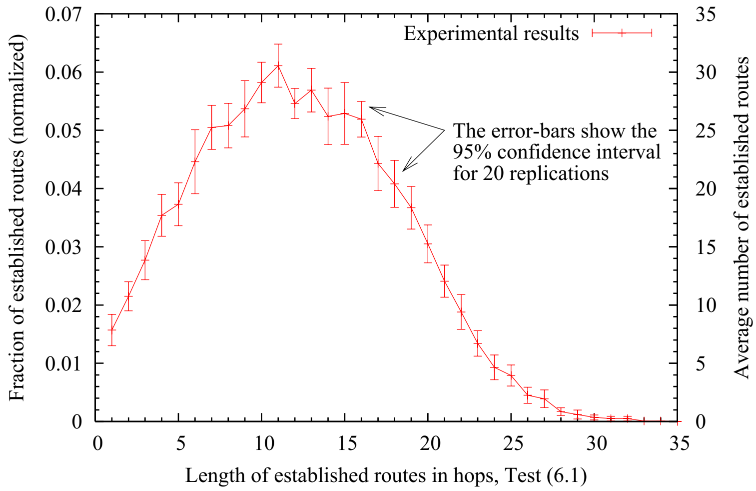


Figure 93: Route-length distribution for Test (6.1). We include the 95% confidence intervals.

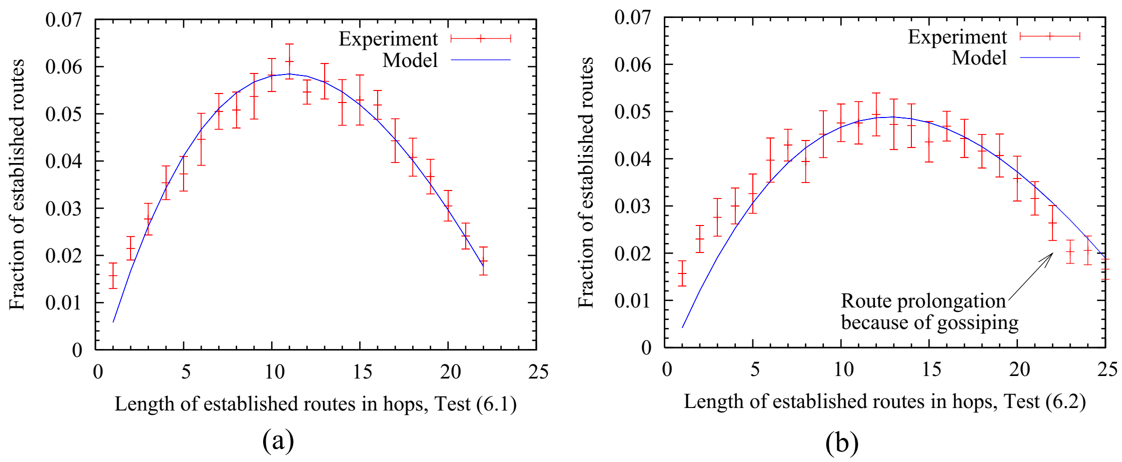


Figure 94: Route-length distribution for (a) Test (6.1) and (b) Test (6.2). We include the 95% confidence intervals. Our model predictions are fitted to the measurements using a least-squares fit.

answers directly even if outside the circle of radius r_1. This special behavior can be observed for routes up to approximately 5 hops. In theory one would need to model the average range r_1 as function of h. For the case $h > 5$, we expect $r_1(h) \approx r_1$; for the case $h = 1$, $r_1(h) = \sqrt{2}r_1 = r$. For the sake of simplicity, we omit the modeling of this special behavior. Figure 94 also gives the fit for the gossiping variant of AODV. We notice the same increase in short routes. Moreover, the elongation of routes is clearly visible. The fitting parameter θ as well as q, both obtained using the simulation, are as follows:

- For AODV (Test (6.1)): $\theta = 1.2$, $1-q = 0.99$.
- For gossip-enhanced AODV (Test (6.2)): $\theta = 1.43$, $1-q = 0.99$.

6.4.4 Modeling "Expanding Ring Search"

Expanding ring search is a protocol optimization which AODV uses to increase the protocol efficiency [PBRD03]. Given the assumption that the communicating nodes are located nearby, the pure flooding of route requests would generate an unnecessary amount of network traffic. *Expanding ring search* is a stepwise increase of the time-to-live of routing requests (RREQ). The RREQ is first propagated with hop count 1. If no route is found, the hop count is stepwise increased until it reaches an upper limit. As the propagation boundary increases, the network load increases as well. The limited propagation over the short distance produces nearly optimal graphs (spanning trees). As soon as the RREQ is flooded throughout the network, contention for the medium may introduce additional errors. This hinders an optimal propagation and, as a result, the graph degenerates. This needs to be considered within the model in two ways. First, the propagation needs to be divided into two distinct areas. For the area covered by the *expanding ring search,* h_{ers}, we obtain θ_{ers}; for the wider area we obtain θ_f as the correction factors. Second, the network load is influenced. If the *expanding ring search* is successful, the overall network load is reduced and the error probability decreases. For hop counts larger than h_{ers}, the errors follow the model introduced in Section 6.4.1.

As a result, we obtain a function defined in three sections. Consequently we only investigate the case where $h_{ers}r_1 < 1$, since the transition to flooding is usually smaller than the normalized distance $d = 1$. The corresponding model equations are Equation (52)-(55). The equations are only valid if appropriate fitting is performed.

$$d' = d\theta(d) \tag{52}$$

$$h = \frac{d\theta(d)}{r_1} \tag{53}$$

$$\theta(\xi) = \begin{cases} \theta_{ers}, & \xi \le r_1 \dfrac{h_{ers}}{\theta_{ers}} \\ \theta_f, & (h_{ers}+1)\dfrac{r_1}{\theta_f} \le \xi \le \sqrt{2} \end{cases} \quad (54)$$

$$pm_d^{ers}(\xi) = \begin{cases} (1-q_{ers})^{h(\xi)} 2\xi(\xi^2 - 4\xi + \pi), & \xi \le r_1 \dfrac{h_{ers}}{\theta_{ers}} \\ (1-q_f)^{h(\xi)} 2\xi(\xi^2 - 4\xi + \pi), & (h_{ers}+1)\dfrac{r_1}{\theta_f} \le \xi \le 1 \\ (1-q_f)^{h(\xi)}(8\xi\sqrt{\xi^2-1} - 2\xi^3 - 4\xi) & \\ +\,(1-q_f)^{h(\xi)} 4\xi\left(\arcsin\left(\dfrac{1}{\xi}\right) - \arccos\left(\dfrac{1}{\xi}\right)\right), & 1 < \xi \le \sqrt{2} \end{cases} \quad (55)$$

6.4.5 Experimental Validation of "Expanding Ring Search"

For experimental validation, we conducted two series of experiments: Test (6.3) to examine the behavior of AODV and Test (6.4) to examine the behavior of gossip-enhanced AODV, both with *expanding ring search* activated (see Table 12 for all simulation parameters). The remainder of the simulation parameters is set identical to the ones used to validate the base model (Test (6.1)/Test (6.2)). We carried out 20 independent replications for the experiment.

The results for Test (6.3) and Test (6.4) are shown in Figure 95. We see a significant increase in routes in the close vicinity of the source. Moreover, the steps of *expanding ring search* are clearly visible for both, AODV and gossip-enhanced AODV. The parameters for *expanding ring search* are set to *TTL_START*=1 and *TTL_INCREMENT*=2. The

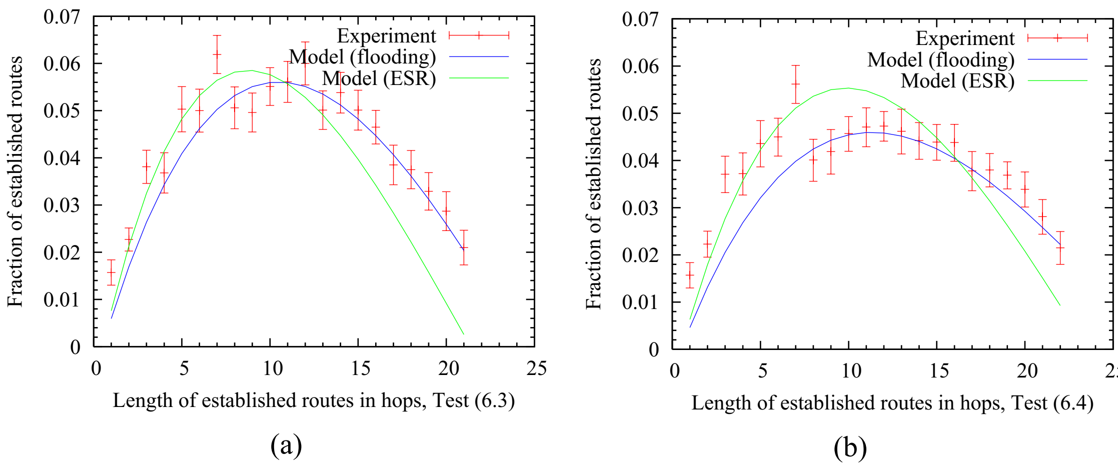


Figure 95: Route-length distribution for (a) Test (6.3) and (b) Test (6.4). We include the 95% confidence intervals. Our model predictions are fitted to the measurements using a least-squares fit.

upper bound for the search is *TTL_THRESHOLD*=7. The RREQ is flooded if no route is found using TTLs of 1, 3, 5, and 7. For our model, these increments induce different network loads and, thus, a stepwise function for the area $h = 1$, $1 < h \le 3$, $3 < h \le 5$, $5 < h \le 7$, and $h > 7$. The end of the curve is a result of RREQ flooding/gossiping throughout the entire network. For the sake of simplicity, we considered only the two main segments of the curve within Equation (55). The parameters obtained for these two segments are as follows (see also Table 13):

- For AODV (Test (6.3)):
  $\theta_{ers} = 1.04$, $\theta_f = 1.19$, $1 - q_{ers} = 0.958$, $1 - q_f = 0.98$.
- For gossip-enhanced AODV (Test (6.4)):
  $\theta_{ers} = 1.14$, $\theta_f = 1.35$, $1 - q_{ers} = 0.97$, $1 - q_f = 0.97$.

6.4.6 Modeling "Reply by Intermediate"

Another feature of AODV is the possibility that intermediate nodes with valid routes answer the RREQ on behalf of the destination node. The intermediate node generates a route reply for both, source and destination. Since both replies are unicasted, the network is relieved compared to flooding the routing request until it reaches the destination. First, the reply by intermediate feature may shorten the duration of the routing cycle. Second, the probability of acquiring long routes is heightened. Third, the resulting routes may be prolonged under special network conditions.

The correction term introduced in Section 6.4.1 needs to be adopted according to *reply by intermediate*. If we assume the length is reduced by half, we obtain $(1-q)^{h(d)/2}$ instead of $(1-q)^{h(d)}$. In general, route reductions from $h(d)$ to $\sigma h(d)$ (with $\sigma < 1$ as the correction term) can be modeled by $(1-q)^{\sigma h(d)} = ((1-q)^{\sigma})^{h(d)}$. The measurement of σ on the right side of the equation is not trivial, because we are only able to calculate the combination of σ and q from our experiments. The value of $(1-q)^{\sigma}$ may be obtained by knowing h. For an exact estimation of σ, we need to know the distance at which the replying node resides. Nevertheless, the results are suitable for our purpose. Quantitatively, we see a decreasing error probability and a subsequent increase in longer routes.

6.4.7 Experimental Validation of "Reply by Intermediate"

The experimental validation is performed with Test (6.5) to examine the behavior of AODV. Test (6.6) studies the same case for gossip-enhanced AODV. To allow for a greater number of active routes, we increased the number of *RREQs* while the rest of the simulation parameters was kept the same as previously used. The experiments were performed using 20 replications each.

The results for Test (6.5) and Test (6.6) are depicted in Figure 96. We notice a better reply behavior due to the increased activity and possible replies by intermediate nodes. Moreover, the *expanding ring search* characteristics produce two sections of the curve,

Table 13: Summary of Results for Test-series (6.*); AODV & Gossip-enhanced AODV.

	AODV			Gossip-enhanced AODV		
Test	**Test (6.1)**	**Test (6.3)**	**Test (6.5)**	**Test (6.2)**	**Test (6.4)**	**Test (6.6)**
θ_{ers}	n.a.	1.04	1.064	n.a.	1.14	1.11
θ_f	1.2	1.19	1.19	1.43	1.35	1.33
$1-q_{ers}$	n.a.	0.958	0.976	n.a.	0.97	0.96
$1-q_f$	0.99	0.98	0.98	0.99	0.97	0.97

which are clearly visible. The combination with *reply by intermediate* gives a smoother transition between the individual steps, thus, supporting our approach of modeling the whole equation in two steps. The fitting of the curves is also given in Figure 96 for both tests. The measured values and fitting parameters are as follows (see also Table 13):

- For AODV (Test (6.5)):
  $\theta_{ers} = 1.064$, $\theta_f = 1.19$, $1-q_{ers} = 0.976$, $1-q_f = 0.98$.
- For gossip-enhanced AODV (Test (6.6)):
  $\theta_{ers} = 1.11$, $\theta_f = 1.33$, $1-q_{ers} = 0.96$, $1-q_f = 0.97$.

6.4.8 Results

We present the summary of results of our simulations in Table 13. The success probability of finding a valid route is larger using *expanding ring search* and *reply by intermediate* than using pure AODV. The comparison of Test (6.2) and Test (6.3) reveals a remarkable result. One would generally expect that the much higher network load (by a factor of 20) should result in lower success probabilities. Since the increased traffic on the other hand allows for replies by intermediate nodes, the probability is nearly similar and in Test (6.3) even above the result with low network load.

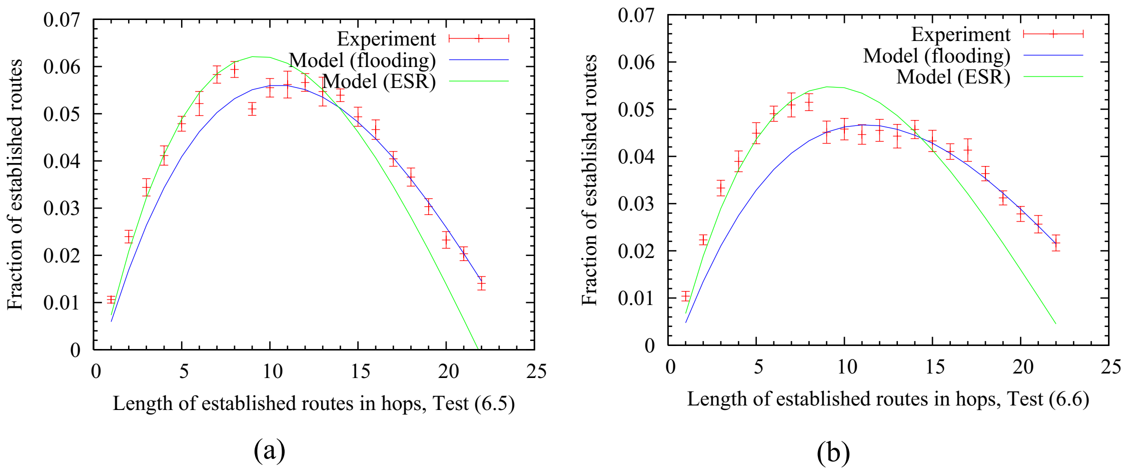


Figure 96: Route-length distribution for (a) Test (6.5) and (b) Test (6.6). We include the 95% confidence intervals. Our model predictions are fitted to the measurements using a least-squares fit.

6.5 Related Work

Our survey of related work starts with a brief analysis of the potential and the deficiencies of simulation studies. Following, we review various analytical models describing the basic properties of wireless multihop networks, which include models in the area of performance analysis, spacial capacity, network connectivity as well as the analysis of protocol complexity. The related work is concluded with some loosely related work in the area of infrastructure based networks.

Simulation Studies for Protocol Analysis

A large number of performance comparisons of various ad hoc routing protocols exist of which the work of Das et al. [DPR00] is an early and prominent example. These studies are of great importance for verifying exact protocol behavior in well-defined environments. However, they are likely to be imprecise due to the large set of predictor and response variables that need to be considered [Jai91] and cannot deliver qualitative metrics to describe overall network behavior and protocol scalability. These limitations are evident if we regard one of the largest and most comprehensive performance studies, which was performed by Lee et al. [LBRP03]. Here, the performance of the AODV protocol is studied for networks up to 10,000 nodes. Various flavors of the protocol are studied, however, to keep complexity reasonable, the surrounding parameter set was chosen to be quite restrictive. Despite the fact that a range from 50 to 10,000 nodes was modeled in the network, the study does only investigate 20 concurrent data streams for each setup. Moreover, the random waypoint mobility model—which is known to be problematic if not carefully parameterized [YLN03] [CBD02]—further limits the generality of the results.

In summary, [LBRP03] clearly points out the behavior of the different protocol optimizations for the given setup, although several open issues remain. Because of these and other deficiencies of simulation studies, the results obtained cannot be generalized. Nevertheless, the experimental analysis by means of simulation provides a powerful prediction tool for precise setups with restricted conditions.

Analytical Models for Network Analysis

The limitations of simulation studies clearly mandate the use of analytical models, which can provide more general results. Early work related to the capacity of multihop packet radio networks was performed by Kleinrock et al. (see, for example, [KS78], [TK84], and [NK84]). The results account for the link layer performance under various circumstances and are based on a sound analytical approach. They focus on the spatial capacity [NK84] as well as the optimal transmission ranges for randomly distributed packet radio terminals (see [KS78] for the initial work on optimum transmission radii, which was refined and corrected in [TK84]).

Recently, the work of Kleinrock et al. was enhanced by Gupta and Kumar in [GK00], a cornerstone of analytical capacity and performance estimation for large-scale ad hoc networks. Starting with assumptions motivated by current technology, the transport capacity of a multihop wireless network is derived. Besides this, [GK03] further enhances some assumptions to predict the achievable rates of large multihop networks. [GK00] has been extended at various points. As mentioned above, Grossglauser and Tse [GT01] introduce mobility of the individual nodes (with relaxed delay and memory conditions) and are able to show that mobility may increase the capacity of ad hoc networks. This seems counterintuitive first, but is achievable by exploiting the locality of nodes: mobility causes sources (or their neighbors, which act as relays) and destinations to move in proximity over time and allows for better usage of the wireless medium. There is other considerable work, which further explores the theoretical capacity of ad hoc networks. This includes, for example, [GV02] considering additional traffic patterns and [LBC$^+$01] estimating analytically the capacity of the 802.11 media access control (MAC) layer [IEEE01].

The capacity and the connectivity of ad hoc networks are interdependent and tightly coupled with the interference within the network—dense networks usually try to optimize capacity by minimizing interference as far as possible while sparse networks try to achieve connectivity even at the expense of strong interference. Thus, there exists work devoted to the connectivity in ad hoc networks (see, for example, [DTH02] for the connectivity of large-scale wireless ad hoc networks). Since connectivity directly depends on the distribution and mobility of nodes as well as the transmission range of the radios, these models are closely connected to mobility models. Significant work has been performed in this area. The properties of various mobility models have been studied in detail—most of these being random models (see [CBD02] for a summary of prominent random models). The main focus of the models is to analyze the stationary and dynamic properties of spatial node distribution and node connectivity (see, for example, [BHPC02], [CBD02], and [BRS02]). The non-random models that try to be as realistic as possible are, however, used for teletraffic modeling in infrastructure-based networks as already described in Chapter 3 in detail.

There is, however, little work that takes an analytical approach for describing the realistic characteristics of ad hoc routing protocols. This is especially true if we take protocol optimizations into consideration. The work of Kail, Nemeth, et al. [KNT01] estimates the possible capacity of ad hoc networks—using a model of the idealized source routing process. In contrast, Santivanez et al. [SMSR02] developed a general performance comparison metric on an abstract level. Their results account for various protocols mechanisms. In particular, they studied the complexity of the individual schemes with respect to the induced overhead. Although the work studies the protocol complexity, the characteristic network behavior is not explained further.

In contrast to the aforementioned work, our approach is to obtain deeper insights in the overall network behavior for dedicated protocols—with a higher abstraction level than currently available through simulation. We developed an own model, because the existing models are mostly too abstract to yield the necessary level of detail needed for our further analysis of routing dependability.

6.6 Summary

Within this chapter, we have discussed the realistic modeling of ad hoc routing protocols. As a first step, we analytically modeled the route acquisition process. Hereby, our model predicts the route-length distribution within the network. We extended the idealistic assumptions of existing models to cover transmission errors. We then modeled the behavior of the AODV protocol. To reflect more realistic protocol behavior, the AODV features *expanding ring search* and *reply by intermediate* were integrated as well. Additionally, we extended our base routing model to reflect recently proposed probabilistic optimizations of the route discovery process, namely gossiping.

We validated all models by means of experimental analysis. Our finding is that our model gives precise predictions of the route-length distribution within ad hoc networks operating with realistic protocols. Based on these promising results, we further enhance the model in Chapter 7 to cover the effects of various classes of node misbehavior. Finally, including these refinements, the model serves as a powerful tool to support our analysis of routing dependability within ad hoc networks.

Moreover, the broad scope of the model allows the use for applications besides routing dependability. Some promising application domains are the analysis of routing performance in ad hoc networks as well as the analysis of network scalability.

7 Node Misbehavior—Classification, Analytical Model, and Performance Trade-off

In Chapter 2, we highlighted the significant difference in characteristics between ad hoc networks and infrastructure-based networks. We found that the dependability of the routing system in ad hoc networks inherently relies on the behavior of nodes, which, in contrast to the infrastructure based case, cannot necessarily be controlled or predicted. In this chapter, building on top of our analytical model introduced in Chapter 6, we explore the impact of node misbehavior on the dependability of the routing system in ad hoc networks. We give a classification of node misbehavior and derive an analytical model of the route acquisition process for the special constraints imposed by misbehaving nodes. Here, we utilize and advance the results of Chapter 6. The enhanced model provides a tool to predict the network frailty in presence of misbehaving nodes. We show the need for more adaptability as well as dependability in the routing system—especially if nodes misbehave. Moreover, we explore the performance trade-offs of various protocol optimizations that may aid in mitigating misbehavior and enhancing the dependability of the routing system.

7.1 Motivation

The self-organizing and cooperative operation of mobile and wireless nodes within ad hoc networks bears several interesting research challenges, some of which have already been introduced in Section 6.1. There, we derived an analytical model that predicts the route-length distribution in the network. If combined with other metrics such as end-to-end delay and loss, this metric can be used to describes the overall performance of the ad hoc routing system.

Most protocols silently assume only well-behaving and cooperative nodes to allow for multihop operation of the network and until now we have also regarded only protocol compliant nodes. When operating outside of laboratory conditions the possibility of misbehaving nodes arises, however. In reality there may exist constrained, selfish or malicious nodes. Since the dependability of the routing system in ad hoc networks inherently relies on cooperation among nodes, the robustness of the routing system needs to be addressed under these unfriendly conditions. However, there exists only few work to analytically describe the effects of node misbehavior on the routing system performance.

In this chapter we discuss the influence of node misbehavior on the routing process. In particular, we derive a classification for misbehaving nodes and extend our analytical model of the route acquisition process executed by the Ad hoc On-demand Distance Vector (AODV) routing protocol [PBRD03] to cover different classes of misbehavior. Similar to the model presented in Chapter 6 our misbehavior model describes the behavior of quasi-static ad hoc networks. We perform an extensive simulation study to obtain further insights into the dynamics of misbehavior. The contributions of this chapter are:

- A *classification of node misbehavior*. This includes an intuitive model of node misbehavior as well as a classification of node misbehavior to suit analytical models.
- An *extension to our analytical model* presented in Chapter 6 to cover various classes of node misbehavior. We include the modeling of *inactive*, *selfish*, and *malicious* nodes.
- An extensive *experimental analysis studying the effects of node misbehavior vs. performance optimizations*. Our study covers the dynamic aspects of misbehavior and of mobility. We analyze the AODV protocol as well as various optimized variants of the protocol.

Our results enable the precise prediction of the effects node misbehavior induces on the overall network behavior within ad hoc networks. We believe that these results are of high importance for the further development of ad hoc routing protocols.

7.2 Outline

We approach the modeling of node misbehavior in two steps. In Section 7.3 we first give a classification of node misbehavior. The classification consists of a technical model covering node misbehavior and a classification scheme to suit the needs of analytical modeling. Second, in Section 7.4, we model various classes of misbehavior analytically. We build on the work presented in Chapter 6 and detail the modeling process of various classes of misbehavior. In Section 7.5 we describe the experimental design for our simulation study. In particular, we investigate the trade-off between performance optimization of the routing protocol and robustness of the routing system. The results are presented and discussed subsequently in Section 7.6. Section 7.7 presents related work in the area of ad hoc routing. We conclude with a brief summary of this chapter in Section 7.8. The coarse structure of this chapter is shown in Figure 97.

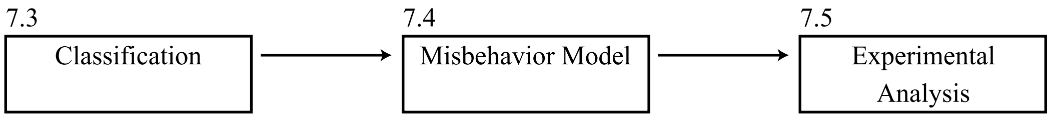


Figure 97: Structure of Chapter 7.

7.3 Classification of Node Misbehavior

There is no common classification of node misbehavior. The authors of [MM02] [Kar03] and the other related work given in Section 7.7 each introduce their own category of misbehavior using dissimilar nomenclatures. Since these categories and especially the accuracy of their definition do not suit analytical models like the one derived in Chapter 6, we need to classify node misbehavior differently.

Intuitive Model of Node Misbehavior

In a first step we intuitively model node misbehavior from a technical/implementation perspective. We limit our classification to misbehavior on network layer, though. The possibilities to misbehave can be manifold for ad hoc nodes. Some of these possibilities that affect the routing system are shown in Figure 98. We distinguish into control plane (routing) and data plane (forwarding). Since nodes operate autonomously they have the capability to alter, discard, or inject messages of either plane into the network. From a technical perspective we obtain the following dimensions of node misbehavior, which allow for implementation of a sophisticated misbehaving node:

- *Time*. We describe the on/off status of node misbehavior using t_{start} and t_{stop}.
- *Degree of behavior*. We specify the probability p with which the node follows its misbehavior. This allows for randomly misbehaving nodes.
- *Plane of behavior*. We specify the plane that is affected by misbehavior to be the *control plane*, the *data plane*, or *both planes*.
- *Type of behavior*. We determine the action the node performs with a specific packet: *forward packet, discard packet, inject packet*. Packets can be *modified* in the node.
- *Behavior against whom*. We specify, which nodes are affected from the behavior: *all nodes*, a *subset of nodes*, a *superset of nodes*, *no node* (a misbehaving node that affects no other node is in fact again a well-behaving node).

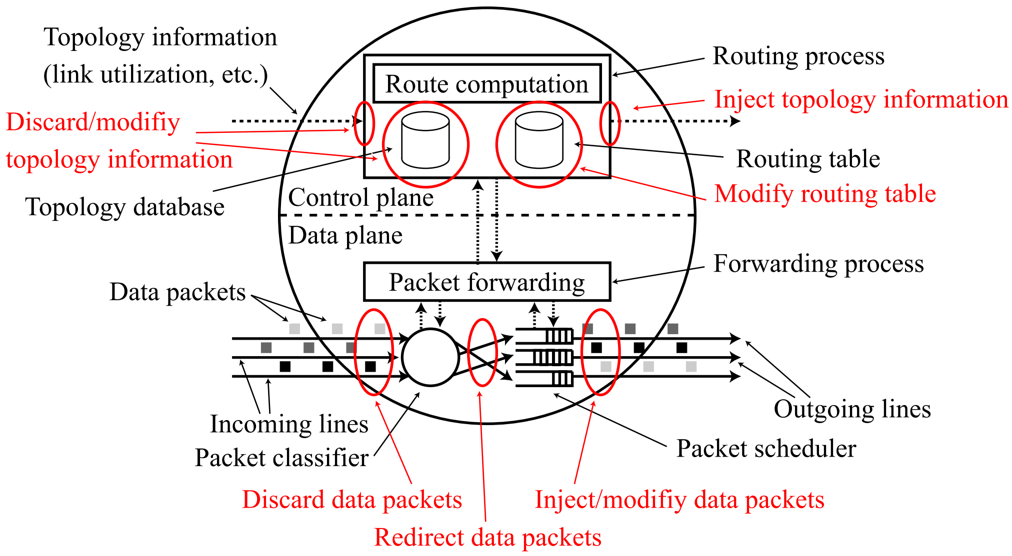


Figure 98: Conceptual misbehavior model for ad hoc nodes.

Please note that these dimensions are not necessarily orthogonal to each other and that arbitrary combinations may not make much sense. These technical dimensions are flexible enough to implement various types of misbehavior, though. For example, a node that forwards control packets unmodified but discards all data packets might cause severe packet loss in the network. We implemented the aforementioned flavors of behavior within the Qualnet® network simulator. For the reason of simplicity, we omitted "selective" malicious nodes, which only act maliciously against a subset of all nodes.

The sheer complexity of the intuitive approach towards node misbehavior prohibits its application in our analytical model. Therefore, we additionally introduce a characterization of node misbehavior using well-defined classes to allow for further analytical study.

Classification of Node Misbehavior for Analytical Modeling

Our class-based approach aggregates the types of node behavior. The classification allows, on the one hand, to be analytically tractable while, on the other hand, to model realistic behavior. We derived the following classes:

- *Cooperative nodes*, which comply to the standard, at all times.
- *Inactive nodes*, which include *lazy nodes* (unintentionally misconfigured) and *constraint nodes* (for example energy-constraint or field-strength-constraint).
- *Selfish nodes*, which optimize their own gain with neglect for the welfare of other nodes.
- *Malicious nodes*, which inject false information and/or remove packets from the network.

We note that, depending on the degree of non-cooperation the nodes exhibit, selfishness may partially overlap with inactivity or maliciousness. A more detailed description of the classes as well as further restrictions imposed to allow for modeling are given in the corresponding sections below.

7.4 Modeling of Node Misbehavior

Our model of node misbehavior is based on the model of the route acquisition process executed by the AODV protocol as described in Chapter 6. In this section, we extend the model to cover the effects of node misbehavior. We formulate the model for *inactive nodes*, *selfish nodes*, and *malicious nodes.* The deformation of the probability measure when misbehaving nodes are present allows to characterize the network behavior in comparison with the well-behaving case.

7.4.1 Modeling of Inactive Nodes

The behavior of *inactive nodes* can be easily described and traced analytically. In reality, these may be constrained nodes or lazy and misconfigured nodes, which are intentionally or unintentionally not participating in the route discovery and packet forwarding process.

Definition (14). *An* ***inactive node*** *is neither active on the control plane nor on the data plane. It does not cooperate during the routing process and does not forward any packets.*

Our model assumes that inactive nodes are neither the source nor the destination of a route. Since our definition of behavior concerns the network layer, these nodes may operate correctly on layer 2 and below. However, we assume that inactive nodes do not cause errors on the layers below the network layer. For our analysis we use the results and the notation of the modeling process in Chapter 6.

Within our model, inactive nodes are extracted from the network. The number of nodes is effectively decreased by the number of inactive nodes. Let the fraction of inactive nodes be f_{in} and the total number of nodes be n. The number of inactive nodes is then $f_{in}n$ and the number of active nodes $(1-f_{in})n$. Only the active nodes participate in the route discovery cycle. Obviously, the node density decreases as the number of inactive nodes increases. The average number of nodes within a transmission radius is given by the node degree M as introduced in Equation (41). We obtain the transformed node degree M' to be

$$M' = \left(\frac{r}{r'_0}\right)^2 = \left(\frac{r}{r_0}\right)^2(1-f_{in}) = M(1-f_{in}) . \tag{56}$$

The transformed equivalent radius r'_0 for the case of inactive node is

$$r'_0 = \sqrt{\frac{A}{\pi n(1-f_{in})}} = \frac{1}{\sqrt{(1-f_{in})}}\sqrt{\frac{A}{\pi n}} = \frac{r_0}{\sqrt{(1-f_{in})}} . \tag{57}$$

The change of the node density also influences the route-length distribution. As long as the network is sufficiently connected, normal operation is possible. With increasing number of inactive nodes and, thus, decreasing density, the network gets partitioned and the communication is restricted to subsets of nodes inside these partitions.

We experimentally tested the influence of node inactivity. We used a test setup similar to Test (6.1) from the previous chapter and deactivated 50, 100, 150, and 200 nodes. Figure 99 shows the results for these tests. The experimental validation within our 500 nodes setup with initially $M = 9$ showed 50 and 100 inactive nodes can easily be tolerated. From 150 inactive nodes onwards ($M' = 6.25$) the first drops are visible. For 200 ($M' = 5.38$) the routing is significantly burdened. Further increasing the number of inactive nodes renders the network nearly disconnected and unusable.

7.4.2 Modeling of Selfish Nodes

Selfish nodes maximize their own gain. Depending on the routing protocol, this selfishness may take various forms. It is optimally—from the selfish node's perspective—to only act cooperatively if packets from/to itself need to be processed. With this behavior the node is able to save considerable amounts of energy. However, depending on the routing protocol, information about the selfish node may be evanescent in routing tables of other nodes. As a consequence, we model a more perfidious strategy of selfishness: the cooperation for own data packets and all routing messages, and the non-cooperation for the data packets of all other nodes. In essence, these selfish nodes do not aid other nodes on the data-plane, but actively discard packets routed through them. To be able to send and receive packets from other nodes, they are cooperative on the control-plane.

Definition (15). *A* ***selfish node*** *does not forward any data packets for other nodes except for itself. The node cooperates during the route discovery cycle to maintain a correct routing table and to be present in other routing tables.*

Due to selfish nodes, routes that exist (cooperation for route discovery) may not be used to relay any packets to the destination (non-cooperation for data packets). Within standard AODV neither source nor destination are able to detect this misbehavior. From the perspective of the destination node there is an active route, however, no data arrives. From the perspective of the source node, the application level packets are sent, but there is no reply. Tracing the data packets inside the network is also then not possible.

We modify our model equations to describe the selfish behavior as follows. If we insert a fraction f_{sn} of nodes that do not forward data packets, we obtain on average an error probability of $q_{sn} = f_{sn}$ for neighboring nodes assuming a point-to-point connection. The errors on layer two and below may also add some additional loss. In the absence of selfish nodes $1-q$ is the success probability as described in Section 6.4.1. Since the errors induced by selfish nodes are independent of link layer errors, the combined success probability is $(1-q)(1-q_{sn})$. This holds for neighboring nodes except when the packets are sent to the selfish node itself. If we further exclude collusions

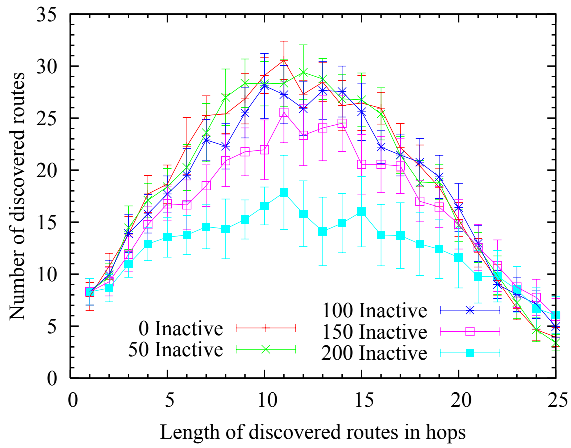


Figure 99: Route-length distribution of AODV with inactive nodes. We use the setup of Test (6.1) and increase the number of inactive nodes. We show the 95% confidence intervals for 20 replications.

among nodes, this probability holds for each node independent of predecessors. The resulting probability of a successful data packet transmission is

$$p_{selfish}^{success} = ((1-q)(1-q_{sn}))^{h(d)} \tag{58}$$

Combining of the resulting probability with the results of Chapter 6, we obtain a modified function for the probability measure, which now gives the estimated number and probability of routes, which carry data streams without errors. See Figure 100 for the resulting deformation of the probability measure. It is, however, very important to keep in mind that from an end system's perspective the deformation of the route-length distribution is not visible since our selfish nodes take part in the route discovery cycle. Hence the distribution of *discovered routes* does not change, reflecting the non transparency of selfish nodes to all other nodes. In contrast, the *discovered and operational routes* follow the described deformation. In summary, the behavior of this sort of nodes is more severe to the network than the behavior of inactive nodes.

7.4.3 Modeling of Malicious Nodes

Malicious nodes reduce the utility of the network without regard for their own gain. Maliciousness may naturally take on many forms.

Definition (16). *"A **malicious node** abuses the cooperation among nodes to hinder operation of the network."*

For our studies we choose *black holes*, which masquerade with a fake destination and try to attract routes and data packets, to represent malicious nodes.

Definition (17). *"A **black hole** answers each route request with a falsified route reply claiming to have a one hop route to the destination. If data packets arrive, the black hole discards these packets."*

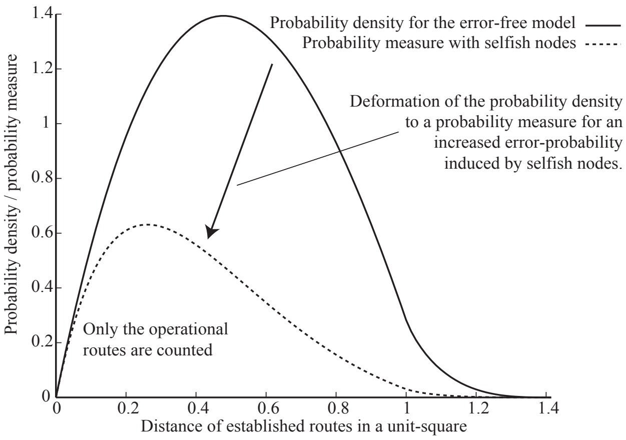


Figure 100: Deformation of the route-length distribution for selfish nodes. The shown curves visualize the expected trend and do not represent a specific setup.

Studying the behavior of standard AODV, the consequences of *black hole* behavior are obvious. RREQs are propagated until a node is or knows of the destination and answers with a RREP. The source of the request accepts the first incoming answer and then only the answers with the same or newer destination sequence number and lower hopcount. If the RREQ only reaches the intended destination, the issued RREP is correctly accepted by the source and the data transfer should also be successful. If the RREQ only reaches one or multiple black holes, the source sends data towards one of these. Normal protocol operation assumes only one destination node. Introducing black holes changes this behavior. The black hole acts as data sink, announcing itself as being one hop away from a fake destination. This may be described as multiple concurrent destinations. Due to the protocol operation of AODV, the node with the shortest route wins the "competition". We can model this behavior using the areas dominated by black holes vs. the area dominated by the original destination. A source only obtains a valid reply if it is located in the "sphere of influence" of the valid destination. The possible communication distance may serve as a metric to describe the influence of black holes.

Let us assume only one black hole within the network. All nodes (including the black holes) are randomly placed, and the number of nodes is very large. In this simple case we can model the sphere of influence of the black hole using basic geometrical relationships. The black hole effectively separates the network into two areas. All nodes that are located closer to the black hole as to the destination are trapped. The border between the areas is given by the perpendicular bisector between the black hole and the real destination. For networks with multiple black holes we can easily generalize this result. Let the fraction of black holes be f_{mn}. This leads to $f_{mn}n$ black holes in the network. The sphere of influence of each black hole is in average restricted to the $(f_{mn}n + 1)$ th part of the square simulation area A (see Appendix D for the mathematical proof of this relation). Figure 101 shows an example for a network with $n = 25$ nodes and (a) $f_{mn}n = 1$ and (b) $f_{mn}n = 3$ black holes.

Since our initial assumptions include the random placement of nodes, we can determine the number of nodes inside the individual spheres of influence of the black hole to be approximately $1/(f_{mn}n + 1)$ of all nodes. The consequences for the route discovery process are devastating because a successful transmission of data is only possible if the source node is located in the sphere of influence of the intended destination node. We are able to derive the expected size of the areas covered from black holes as well as from the real destination. We model these areas using a circle and a square and derive the estimated average distance to the next black hole (see Appendix D for the reasoning) in the following. We obtain the length in hops of the longest possible route if all nodes cooperate in the network to be

$$h_{max} = \frac{d_{max}}{r_1} = \frac{1}{r_1}\sqrt{2A}\,. \tag{59}$$

As described earlier, black holes cut down the size of the area in which a successful communication can be expected. We are able to obtain the expected average value of the maximum route length in hops within this restricted area to be

$$h^{\square}{}_{max} = \frac{1}{r_1}\sqrt{\frac{2A}{f_{mn}n + 1}} = h_{max}\sqrt{\frac{1}{f_{mn}n + 1}} \quad \text{for a quadratic area, and} \tag{60}$$

$$h^{\circ}{}_{max} = \frac{1}{r_1}\sqrt{\frac{4A}{\pi(f_{mn}n + 1)}} = h_{max}\sqrt{\frac{2}{\pi(f_{mn}n + 1)}} \quad \text{for a circular area.} \tag{61}$$

As mentioned earlier, a successful communication is only possible if the distance of the source node to the real destination is smaller than the distance between source node and black hole. As a consequence, all destinations farther than $h_{max}/2$ away from the source are probably black holes. The deformation of the resulting route-length distribution is very strong. For distances greater $h_{max}/2$, the number of valid routes heavily decreases since a black hole most probably answers the RREQ. Figure 102 shows a qualitative esti-

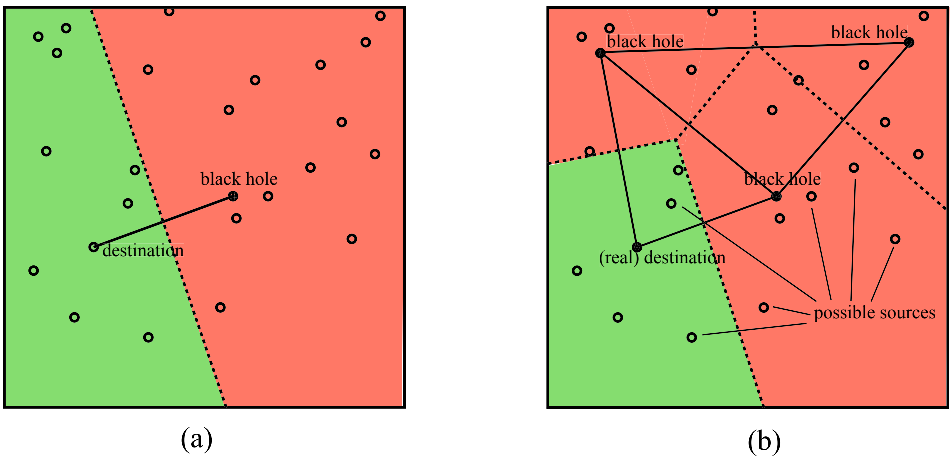


Figure 101: Sphere of influence of (a) one black hole and (b) three black holes.

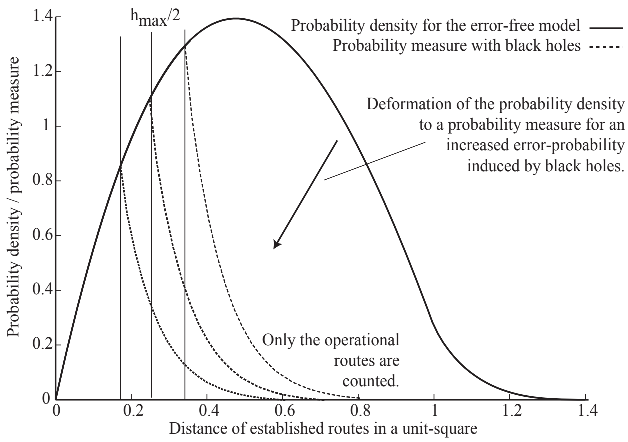


Figure 102: Deformation of the route-length distribution for malicious nodes. The shown curves visualize the expected trend and do not represent a specific setup.

mate of the resulting probability measure function. It is obvious that even a few black holes may hinder large areas of the network being connected. Black holes are able to inflict far more damage than the other types of misbehavior we discuss.

7.5 Experimental Analysis of Node Misbehavior

Correctly parameterized, our analytical model allows to predict the effect of node misbehavior in static ad hoc networks. While our metric, namely the route-length distribution, is very well suited to obtain insights into the overall network behavior, there exist other response variables of interest, which cannot be obtained analytically. Moreover, we aim to study the influence of node mobility, which is not captured in our model so far.

Until now, experimental studies of node misbehavior are fairly limited in size and evidence. To fill this gap, we performed an extensive experimental study to cover important aspects of the effects of node misbehavior not studied yet. We are specially interested in the trade-off that exists between protocol performance optimizations and network stability, if misbehaving nodes participate in the network.

7.5.1 Experimental Design

We analyze the influence of various forms and degrees of node misbehavior for the AODV protocol as well as various extensions. We again follow the methodological steps proposed in [Jai91]:

- *Definition* of the *system*, *goals*, and *services*: The analyzed system is a mobile ad hoc network. We study various routing protocols with respect to the effect of node misbehavior. The network's service is the best-effort delivery of datagrams.
- *Selection* of the *metrics*: Our metrics capture the performance and dependability of the routing system. In particular, we investigate the route-length distribution. Other metrics include application related metrics such as the end-to-end delay and the

goodput. We also observe routing related metrics such as the routing protocol overhead, the fractions of routing message types, and the average path length.

- *Definition* of the *parameters* to study: The most important parameters of our study are the influence of various types of misbehavior, the fraction of misbehaving nodes, as well as the variant of the routing protocol. Additionally, we study different sizes of networks and different mobility patterns.
- *Selection* of the *factors*. We linearly increase the fraction of selfish nodes ranging from 0% to 16% of all nodes and the fraction of malicious nodes ranging from 0% to 4% of all nodes. On top of AODV we parameterized various types of probabilistic flooding to optimize the route acquisition process: we include gossiping with fixed probability $p = 0.7$ (denoted as gossip-*p*), gossiping with probability $p = 1$ for the first hop $k = 1$ and $p = 0.7$ for subsequent hops $k > 1$ (denoted as gossip-*pk*), and gossiping with probability $p = 1$ for the first hop and if no other copy of the route request has been overheard, otherwise the probability is $p = 0.3$ for subsequent hops (denoted as gossip-*pkm*). Moreover, we vary the mobility of nodes between no mobility, low random waypoint mobility (1-2m/s), and high random waypoint mobility (5-7m/s). See Figure 103 for the factors studied and their respective levels. At the same time a number of factors has been kept constant during our experiments. This includes the degree of the network (average node density) and the workload as well as physical and link layer parameters. See Table 14 for the factors that have been kept constant.
- *Choice* of the *evaluation technique*. We simulate various setups using the Qualnet® network simulator. We enhanced the AODV [PBRD03] protocol within Qualnet®. In particular, we implemented probabilistic flooding (gossiping) [HHL02] in multiple variants. The changes in Qualnet® are documented in [Har03], [Sei04], [Spe03], and [Pet04].
- *Selection* of the *workload*. We chose a moderate workload for the network, that is, 8% of the sources act as sender for the AODV protocol. For example in the 250 node network we randomly pick 20 sources and 20 destinations for the tests. We use constant bitrate traffic over UDP.
- The *design* of the *individual experiments*, the *analysis* and *interpretation* of the obtained data, and the *presentation* of the result is described below.

See Figure 103 and Table 14 for an overview of the experimental design of the test-series performed for this chapter. Because of the large setup we differentiate the tests using the protocol variants. Prior to the experiments we ran a series of test to obtain appropriate gossiping parameters.

- Test (7.1): AODV [PBRD03], expanding ring search (ERS) is deactivated.
- Test (7.2): AODV, ERS is activated.
- Test (7.3): Gossip-*p* ($p = 0.7$) enhanced AODV [HHL02] with ERS.
- Test (7.4): Gossip-*pk* ($p = 0.7$, $k = 1$) enhanced AODV with ERS.
- Test (7.5): Gossip-*pkm* ($p = 0.3$, $k = 1$, $m = 1$) enhanced AODV with ERS.

Table 14: Experimental Parameter Set for Misbehavior Model; Test-series (7.*).

	AODV				
Test	**Test (7.1)**	**Test (7.2)**	**Test (7.3)**	**Test (7.4)**	**Test (7.5)**
Simulation area	Dependent on the number of nodes area={100: (1476.23 m)^2, 250: (2334.12 m)^2, 500: (3300.94 m)^2}				
Protocol	AODV	AODV-ERS	AODV-ERS Gossip-*p*	AODV-ERS Gossip-*pk*	AODV-ERS Gossip-*pkm*
Gossiping parameters	—	—	$p = 0.7$	$p = 0.7$, $k = 1$	$p = 0.3$, $k = 1$, $m = 1$
ERS parameters	—	hopcounts {1,3,5,7}	hopcounts {1,3,5,7}	hopcounts {1,3,5,7}	hopcounts {1,3,5,7}

The common parameters for the test-series 7.* are as follows.
Simulation: 20 replications per experiment, 900 s simulation time.
Nodes density/placement: The node density is kept constant to be $M = 9$. Nodes are randomly placed. The x and y coordinates are independently and identically uniformly distributed.
Traffic: The source/destination pairs are randomly selected and 16% of all nodes participate in the communication. We use constant bitrate traffic with UDP as transport protocol. The traffic rate is chosen to be 2 packets/s with 512 bytes each per flow.
Physical layer parameters: Propagation model = free space, transmission power = 7dBm, transmission range $r = 249.862$m, $r_0 = 83.287$m, $r_1 = 176.679$m.
Link layer parameters: MAC 802.11b with Distributed Coordination Function (DCF), max. transmission rate = 11 Mbit/s.
AODV parameters: Local repair = deactivated, hello messages = deactivated, net diameter = 35.

7.6 Results

We investigate the influence of node misbehavior on the performance of the routing system. In particular, we study probabilistic mechanisms to optimize the route acquisition process. We focus on the presentation of the results in the 250 node setup. This setup shows the various effects of misbehavior very clearly. The 100 node setup handled the amount of traffic very well, however, the results are biased towards very short route lengths because of network size. In a very low load configuration, the 500 node setup yielded results very similar to the 250 node network. However, selecting randomly 16% of all nodes as communication partners as specified in our experimental setup clearly showed the limitations of AODV [PBRD03] in combination with 802.11b [IEEE01] for this high number of nodes. The network was no longer able to tolerate the routing churn and congestion occurred.

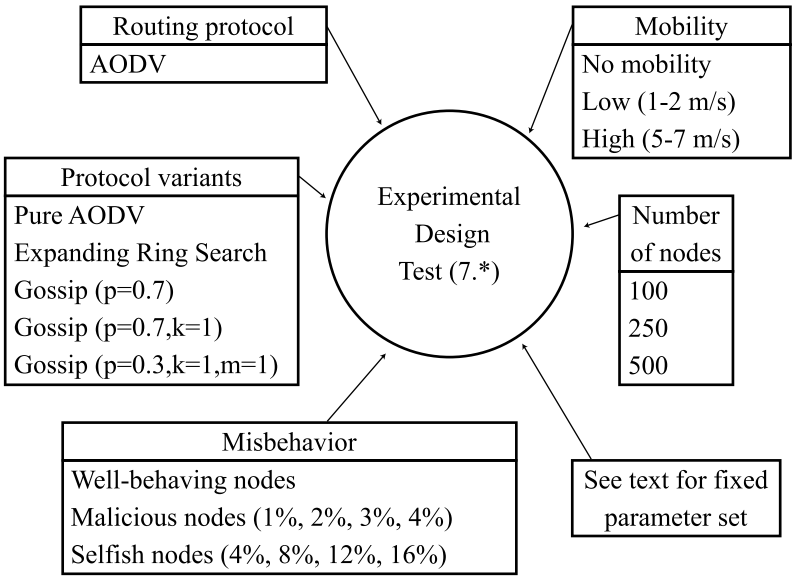


Figure 103: Experimental design for test-series (7.*).

7.6.1 Node Misbehavior vs. Performance Optimization

The performance of ad hoc routing protocols is very much related to the route discovery and maintenance operations. In proactive routing protocols the maintenance of up-to-date routing information requires the distribution of state information throughout the network, which might be very costly in highly dynamic environments. Reactive routing protocols discover routes on demand and, thus, the route acquisition process and the maintenance of active routes contribute significantly to the protocol overhead. Reactive protocols are usually preferred for highly dynamic environments, because the overhead is only generated if data is to be transmitted.

For the studied AODV protocol, the route acquisition process makes use of a flooding mechanism to discovery routes to unknown destinations. There exist various optimizations in the AODV protocol such as expanding ring search or reply by intermediate to limit the flooding area and reduce the control overhead as already discussed in Section 6.4.4 and Section 6.4.6. The process of flooding can also be optimized by adaptation to the characteristics of the network. In dense networks, for example, this adaptation should limit the re-broadcasting of the request to a subset of nodes, thus, limiting the network load. Recently proposed techniques such as probabilistic flooding or gossiping allow for static and dynamic adaptation to the network characteristics (see also Section 6.4.2). Additionally, they can be used to complement the integrated optimizations of AODV described earlier. At the time of writing, we are not aware of any existing work that investigates the effects of protocol performance optimization on the ability of the network to handle node misbehavior. In the following we tackle this challenge by means of an extensive experimental analysis. We first discuss the performance gains of the probabilistic optimization. Second, we explore the effects of various degrees of node selfishness. Third, the impact of malicious nodes is discussed.

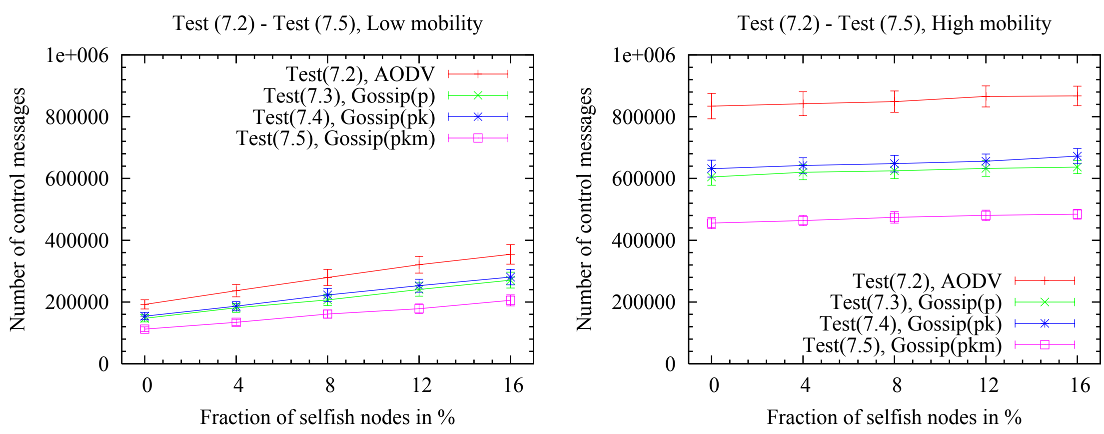
Performance Aspects of Probabilistic Flooding

Only recently and in parallel with our work, the effects and gains of probabilistic flooding have been studied performance-wise [HHL02] [SCS03]. While we expect the probabilistic optimization to yield a better routing performance, the resulting stability against misbehavior cannot be judged easily. Probabilistic flooding very much reduces the redundancy of the distribution of the route request. Moreover, it introduces randomness into the route acquisition process. On one hand the reduced redundancy during the route acquisition process might lead to a more fragile network, on the other hand the network resources that have been freed because of the more efficient route acquisition as well as the increased randomness of the discovery process may be beneficial for the re-acquisition of routes broken by misbehavior.

For our study we implemented the gossip variants proposed in [HHL02] for use in the Qualnet® simulation environment. The misbehaving nodes have been implemented as described in Section 7.3. We parameterized the optional parameters of node misbehavior as follows.

- Selfish and malicious nodes are active during the entire simulation time.
- Malicious (black hole) behavior cannot be detected from the source node.
- Selfish behavior is detected at the source node after a period of 15 seconds. After this period the source node re-issues the route request.

Figure 104 shows the total number of routing messages forwarded over the degree of selfishness for (a) low and (b) high node mobility. The effects of mobility and selfishness are clearly visible. Stepping up the node velocity from 1-2m/s to 5-7m/s increases the number of routing messages dramatically. The increase in selfish nodes also leads to a higher number of routing messages, because the source nodes issue requests as soon as



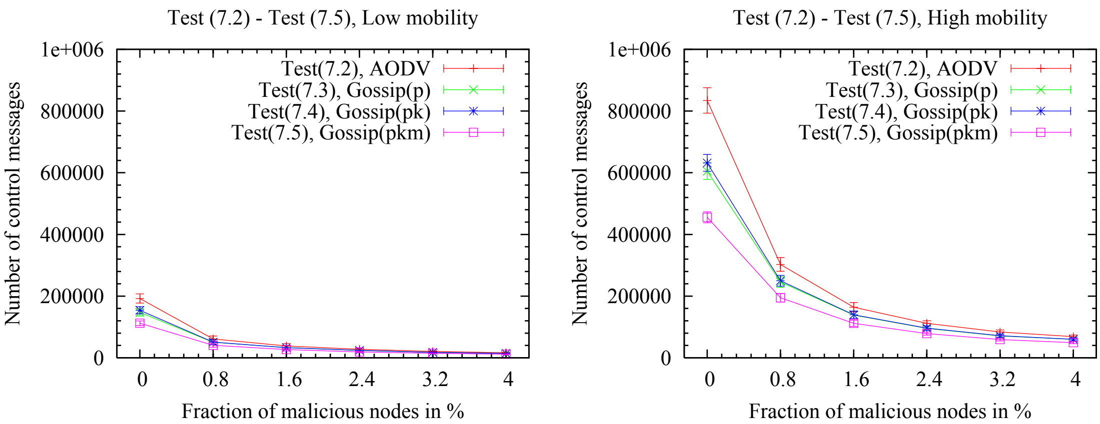
(a) (b)

Figure 104: Number of control messages with selfish nodes, Test (7.2) to Test (7.5). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals.

the selfish route is detected. Probabilistic flooding is able to reduce the number of routing messages significantly. The reduction in control message overhead is around 25% for gossip-*p*, ~21% for gossip-*pk* and as high as ~43% for gossip-*pkm* in the low mobility case. For high mobility the reduction is between ~15% and ~30%.

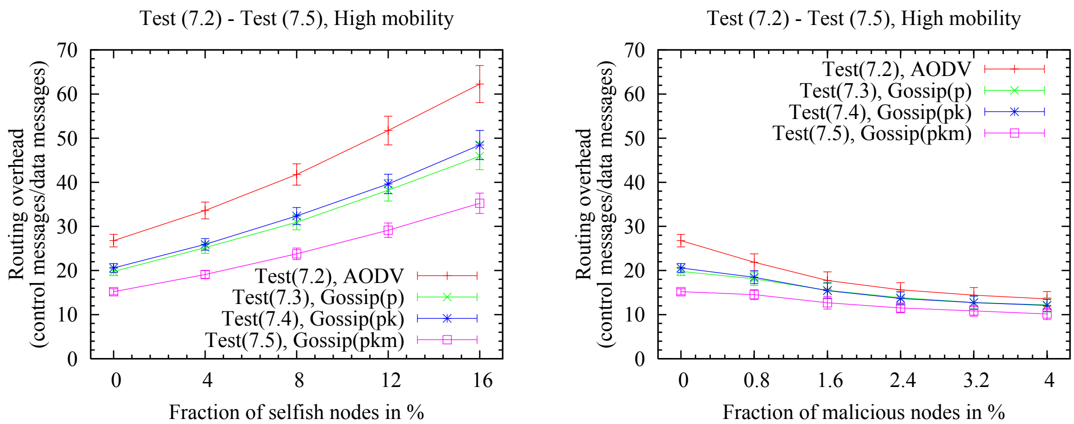
Figure 105 depicts the total number of routing messages over the degree of maliciousness. The increase in malicious nodes decreases the number of routing messages. This decrease in control messages is due to the fact that the expanding ring search feature of AODV limits the further dissemination of the routing request after the routing reply of the malicious node is received. Here, the decrease in control messages reflects the decreasing size of the area where communication is possible. Again, we observe a significant reduction in load for the gossiping variants of AODV.

Figure 106 shows the routing overhead, that is, the number of control packets per successfully received data packet in the high mobility scenario for (a) selfish nodes and (b) malicious nodes. This metric implicitly includes the number of packets dropped



(a) (b)

Figure 105: Number of control messages with malicious nodes, Test (7.2) to Test (7.5). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence



(a) (b)

Figure 106: Routing overhead for high node mobility, Test (7.2) to Test (7.5). Our measurements show (a) selfish node (b) malicious nodes. The error bars show the 95% confidence intervals.

because of selfish and malicious behavior. The observed trend in routing overhead for an increasing number of misbehaving nodes is similar to the trend in the number of control messages. The trends are contrasting for selfish and malicious nodes again. This is due to the fact that selfish nodes do only moderately influence the route length but incur significant packet loss. Malicious nodes, in contrast, heavily effect the route lengths, which reduces the overhead of flooding.

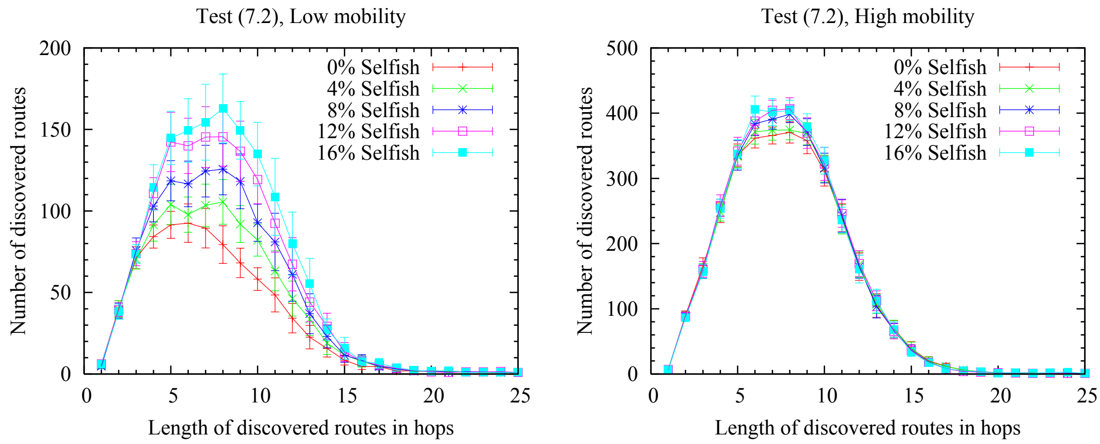
In summary, probabilistic mechanisms are able to reduce the network load significantly. This gain can be directly transferred to a reduction in routing overhead even if selfish or malicious nodes are in the network. Next, we study additional performance metrics related to selfish and malicious node behavior in ad hoc routing systems.

Selfish nodes

We tested the performance of all protocols for selfish nodes in the high and low mobility scenario. Figure 107 shows the route-length distribution of the AODV protocol with selfish nodes for (a) low node mobility and (b) high node mobility. We show *all discovered routes* instead of the *discovered and operational routes given by our model*, because this allows us to capture the dynamics of misbehavior as well. In particular, we cover the breakage of paths due to mobility and due to misbehavior. Figure 107 (a) shows an increase in discovered routes for an increase in selfish behavior. Moreover, the shape of the route-length distribution is slightly deformed. The absolute increase in discovered routes is attributed to the detection of misbehavior: as soon as the source node is informed of the invalid route a new discovery is initiated. Mobility and misbehavior account for the deformation of routes, because the probability of a route breakage due to mobility or selfishness is higher for longer routes.

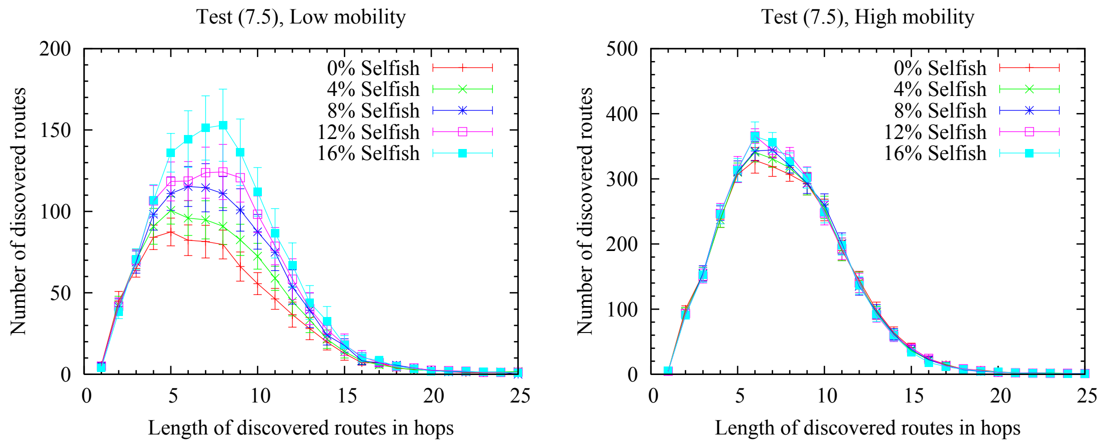
Our results of the high mobility scenario differ significantly compared with the results of the low mobility scenario. Under high mobility, Figure 107 (b) shows only marginal changes in the route-length distribution for different degrees of selfishness. However, the absolute number of discovered routes is much higher than for the low mobility scenario shown in Figure 107 (a). The topology dynamics induced by the increased mobility dominate the dynamics induced by misbehavior. This observation is in accordance with the results shown in Figure 104, where the increase in control messages because of selfishness is small compared with the absolute number for the high mobility scenario.

Figure 108 shows the route-length distributions for gossip-*pkm* and selfish nodes. Compared with pure AODV in Figure 107 small but not significant differences of the curves are visible. The total number of discovered routes using gossip-*pkm* is slightly reduced and the curves are slightly deformed compared with pure AODV. This small difference in absolute performance is remarkable, because the savings in control messages are between 30% and 43% using gossip-*pkm*. This gain comes at the expense of reduced redundancy for the broadcast route-acquisition of AODV. However, as shown, this loss in redundancy only marginally affects the number of discovered routes for selfish nodes.



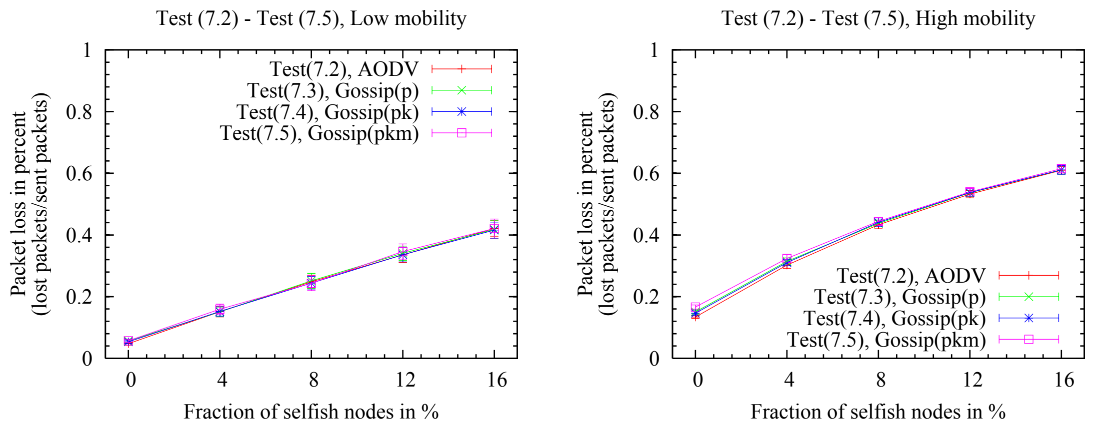
(a) (b)

Figure 107: Route-length distribution of AODV with selfish nodes, Test (7.2). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals.



(a) (b)

Figure 108: Route-length dist. of gossip-*pkm* enhanced AODV + selfish nodes, Test (7.5). Our measurements show (a) low node mobility and (b) high node mobility. Please note the different y-axis.



(a) (b)

Figure 109: Packet loss with selfish nodes, Test (7.2) to Test (7.5). Our measurements show the results for (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals.

Figure 109 shows the packet loss for all algorithms for selfish nodes. Again we visualize both mobility scenarios. The increase in packet loss is nearly linearly with the number of misbehaving nodes for the low mobility scenario shown in Figure 109 (a). For the high mobility scenario, we observe a higher initial loss. Moreover, the increase in loss is more abrupt in the beginning of the curve for high mobility. We provide additional results in Appendix D.

Malicious nodes

For the case of malicious nodes we present selected results for AODV, gossip-*pk* enhanced AODV, and gossip-*pkm* enhanced AODV. Figure 111 shows the route-length distribution of the AODV protocol with malicious nodes for (a) low node mobility and (b) high node mobility. Again our results include the dynamics induced by mobility. The *discovered routes* include all routes, that is, *valid routes* and *malicious routes*. The effect of node misbehavior is clearly visible and the curves follow our model predictions.

The bias in very short routes starting with route-length 2 hops is due to the malicious behavior: if a malicious node is a one hop neighbor of the source node, it answers the request using the reply by intermediate feature and announces a one hop route to the destination. The curves for high mobility show much higher numbers of discovered routes but are similar in shape compared with the low mobility curves. The influence of the gossiping variant on the route-length distribution for malicious nodes is shown in Figure 112 for gossip-*pk* and Figure 113 for gossip-*pkm*. Again, the reduced redundancy of the gossiping variants of AODV has only minor effects. See also Appendix D for additional results. Figure 110 shows the packet loss for all algorithms for malicious nodes. Again we visualize both mobility scenarios. The loss is no longer linear increasing with the number of misbehaving nodes. Moreover, the overall loss is very high. The individual algorithms show a comparable performance, though.

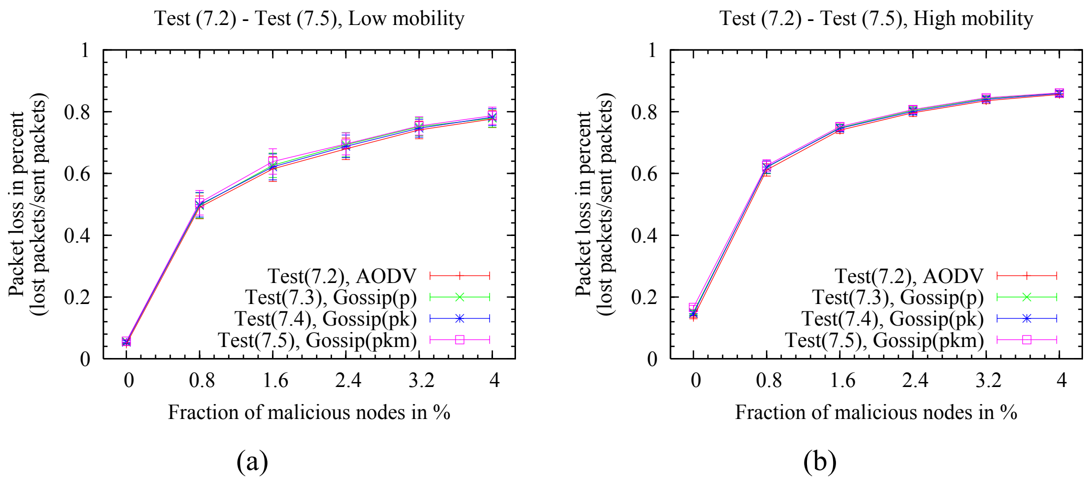
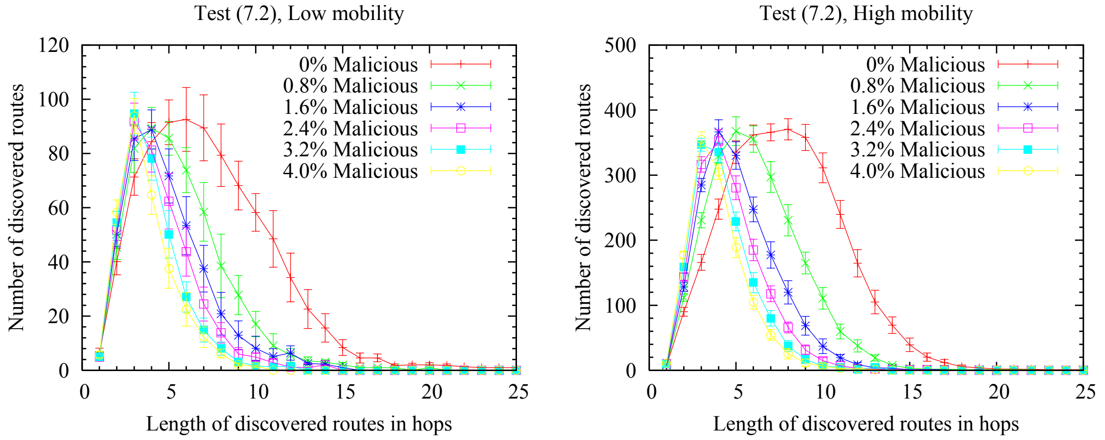
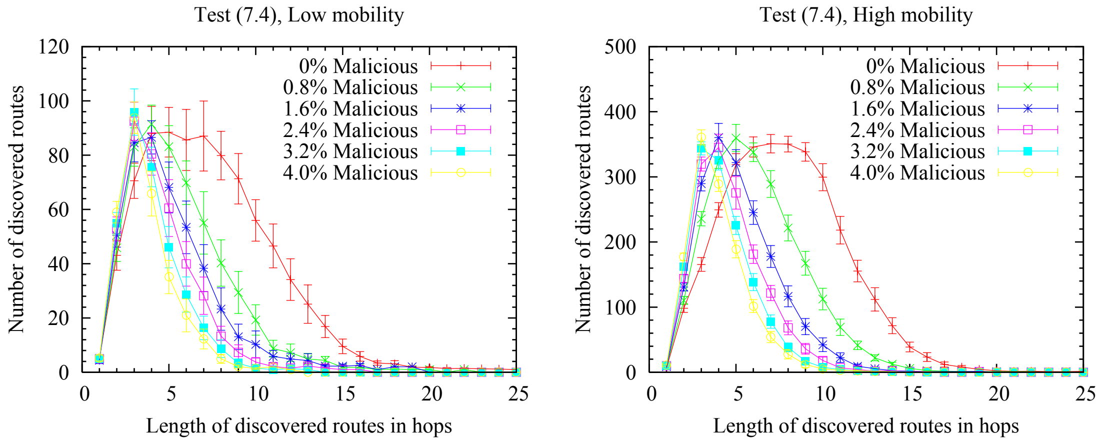


Figure 110: Packet loss with malicious nodes, Test (7.2) to Test (7.5). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals.



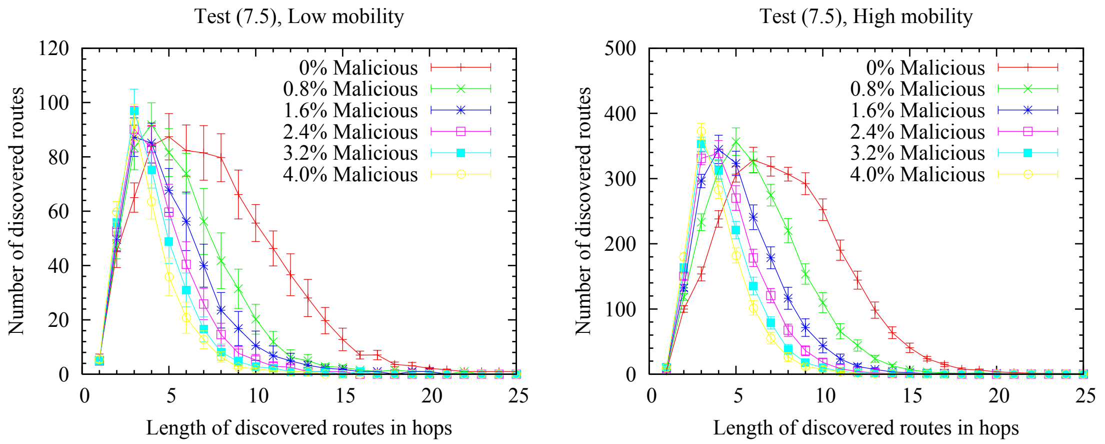
(a) (b)

Figure 111: Route-length distribution of AODV + malicious nodes, Test (7.2). Our measurements show (a) low mobility and (b) high mobility with 95% confidence intervals. Please note the different y-axis.



(a) (b)

Figure 112: Route-length dist. of gossip-*pk* enhanced AODV + malicious nodes, Test (7.4). Our measurements show (a) low node mobility and (b) high node mobility with 95% confidence intervals.



(a) (b)

Figure 113: Route-length dist. of gossip-*pkm* enhanced AODV + malicious nodes, Test (7.5). Our measurements show (a) low node mobility and (b) high node mobility with 95% confidence intervals.

Table 15: Summary of Results for Test (7.2); AODV with ERS, Low mobility, 250 nodes.

	Misbehavior										
Factor	**No**	**Selfish Nodes**				**Malicious Nodes**					
Fraction of misbehaving nodes	**0**	**4%**	**8%**	**12%**	**16%**	**0.8%**	**1.6%**	**2.4%**	**3.2%**	**4.0%**
Average end-to-end delay in ms	76,60	71,68	69,48	65,34	61,79	58,56	45,43	33,99	31,75	34,25
Average loss	4,8%	15,1%	24,8%	33,6%	42,1%	49,0%	61,5%	68,0%	74,1%	77,6%
Avg. path length	7,19	7,53	7,65	7,72	7,87	5,38	4,72	4,29	3,94	3,71
Control Msg. per Data Msg.	5,61	7,78	10,40	13,59	17,21	3,33	2,77	2,47	2,29	2,13
RREQ/Ctrl.Msg.	95,1%	94,8%	94,6%	94,5%	94,4%	90,2%	87,2%	84,9%	82,6%	80,0%
RREP/Ctrl.Msg.	3,3%	3,4%	3,4%	3,4%	3,4%	7,5%	10,2%	12,3%	14,3%	16,3%
RERR/Ctrl.Msg.	1,6%	1,8%	2,0%	2,1%	2,2%	2,3%	2,6%	2,8%	3,2%	3,6%

Summary

Having shown the most important results of our tests, we here present some additional response variables for one of our tests, Test (7.2), that is, default AODV with ERS, with 250 nodes in a low mobility setup. See Table 15 for the results. We show the averages over the 20 replications of the test and distinguish into application level and routing level results. Special emphasis lies on the end-to-end delay, the route length and the fraction of the individual control message types. With increasing misbehavior the average end-to-end delay decreases for selfish nodes. This corresponds with the reduced network load because of dropped packets. There is also a slight increase in average path length. This increase can be explained with the re-discovery of selfish paths. These paths are unstable as long as selfish nodes participate and get more stable if the route consists solely of cooperative nodes for our low mobility scenario.

The routing overhead, that is, the ratio of control messages to data messages, has already been discussed. Table 15 breaks down this overhead to account for the individual control message types. For selfish nodes, we see only minor changes in the fraction of reply (RREP) messages, but a slight decrease in the fraction of request (RREQ) messages to compensate the increase in RERRs. For malicious nodes, in contrast, we observe a decrease in fraction of RREQs going hand in hand with an increase in fraction of RREPs. This is due to the shortening of routes that limits the flooding of the request messages.

7.7 Related Work

The related work in the area of dependable ad hoc routing can be distinguished into two major groups. First, there is work studying impairments to routing dependability in general. Second, there exist routing protocols, which have special mechanisms to ensure the robust and dependable operation of the routing system.

Our work aims at investigating the trade-off between performance optimization and routing dependability, but does not consider solutions to the security problem in ad hoc networks, because there already exists a large body of work to address the latter issue. This work is related to our work in so far that various types of misbehavior are studied and strategies to mitigate the misbehavior are devised. Work in this area has been performed only recently. Zhou and Haas [ZH99] and Hubaux et al. [HBC01] elucidate common problems and threats related to ad hoc networks. In particular, [ZH99] investigates possible threats to ad hoc routing and discusses the necessary security goals for this class of networks. The authors introduce a distributed key management scheme to allow for secure ad hoc routing. In [HBC01], again, threats to ad hoc security are presented. Moreover, a self-organizing public key infrastructure is proposed as a solution to establish trust in ad hoc networks.

The effects of node misbehavior have been studied by Michiardi and Molva in [MM02]. They study the performance of the DSR protocol [JMH04] for three types of selfishness by means of a simulation study. However, the results of [MM02] cannot easily be generalized due to the setup of their simulation study, which does only consider small networks (20 and 60 nodes). Wang, Bhargava et al. in [WLB03] compare the performance of AODV [PBRD03] and DSR [JMH04] for node misbehavior. Again, the size of the network is very small, which limits the generality of the results. Marti et al. in [MGLB00] propose a scheme to detect selfish nodes in ad hoc networks. They describe a watchdog and a path-rater mechanism to detect the misbehavior of nodes. Unfortunately, the mechanisms can easily be circumvented by malicious nodes or even abused to introduce denial of service attacks. Also the performance penalties of node misbehavior are not discussed in detail. Kargl in [Kar03] discusses various building blocks for secure ad hoc routing. He includes a threat analysis and also studies misbehavior in a limited setup by means of simulation.

Most work related to ad hoc network security has been performed in secure routing protocols. Since the relevance of this work is only of minor interest for our work, we do only describe it briefly. Perrig et al. propose various protocols to allow for secure routing. In particular, ARIADNE [HPJ02] and SEAD [HJP02] are proposed to achieve various security goals. Similarly Papadimitratos and Haas propose a secure routing framework [PH02] [PH03]. Buchegger et al. study node misbehavior and propose the CONFIDANT scheme to increase collaboration among ad hoc nodes in [BB02a] and [BB02b].

7.8 Summary

We have discussed the effects of node misbehavior in ad hoc networks. Starting with a general and technical description of node misbehavior, we derived well-defined classes of misbehavior that are suitable for analytical study. An analytical model covering the different types of misbehavior was presented and adjoined to our analytical route acquisition process model introduced earlier in Chapter 6. To gather insights on the effects of misbehaving nodes, the estimated effect of these nodes on the overall routing performance was traced analytically as well as by means of simulation. Our results are that *inactive nodes* only moderately harm ad hoc networks, while *selfish nodes* and *black holes* may have devastating influence on the routing process.

The promise of ad hoc networks is built upon the premise of cooperation among nodes. We have shown the network frailty in the absence of such a cooperation. Our experimental study of node misbehavior covered the cross-influence of misbehavior vs. protocol optimizations as well. We have been able to show that the optimization of the route acquisition process using probabilistic flooding has only a minor influence on the robustness of the routing system. In contrast, we believe that the gain in network resources due to the optimization of flooding should be used to introduce mechanisms to increase the robustness of the routing. Our insights mandate for mechanisms to deal with mobility as well as misbehavior induced dynamics. A possible solution might be the use of multipath routes in combination with forward error correction and packet dispersion over these multipaths as proposed in [XN03]. The multipaths are likely to yield more stable routes and we are able to continue the transmission even if some paths out of the total set of paths fail.

“So eine Arbeit wird eigentlich nie fertig,
man muß sie für fertig erklären,
wenn man nach Zeit und Umständen
das mögliche getan hat.”
—Johann W. Goethe

Part IV
Summary, Conclusion, Outlook, and Appendices

8 Summary, Conclusion, and Outlook

In the preface we projected the vision of the rise of virtual cities, which are empowered by the seamless and pervasive nature of communication networks. Using the analogy to real world cities and the physical world's transportation networks, we motivated our quest for dependable routing in cellular and ad hoc networks to fulfil our vision. In Chapter 2 and Chapter 3 we then introduced the concept of routing dependability and found today's routing architectures to be prepared only insufficiently to handle the influence of real world phenomena such as mobility. Our main contributions in the area of cellular networks are related to the performance aspects of routing dependability. In particular, we investigated novel static and variable network architectures for cellular networks and formulated a powerful new routing paradigm in Chapter 4 and Chapter 5, respectively. For the field of ad hoc networks we focused on the effects of node-misbehavior in Chapter 6 and Chapter 7. We derived and validated an analytical model of the route acquisition process covering various patterns of non-cooperation among nodes. Moreover, the dynamic aspects of misbehavior have been studied. In summary, we have investigated some intriguing aspects of the fundamental concepts of mobility-aware routing systems. We contributed systematic advances in research methodologies, and we obtained a variety of application oriented results.

8.1 Summary

The theme of this dissertation has been to advance the routing dependability of cellular and ad hoc networks to support mobile communications. In particular, the relationship between the network performance and the dependability of the underlying routing system has been addressed.

The first part of the dissertation has served as introduction and motivation. Moreover, we have provided background information on routing dependability. In Chapter 1 "Introduction" we have discussed why the dependability of the Internet's routing system is at stake if we require seamless and pervasive mobililty support using Internet technology. Cellular and ad hoc networks have been named as possible passage-ways from today's Internet towards future networks that support mobility. The identification of research challenges within these networks has been followed by a precise definition of the goals of our investigation. We have highlighted the contributions of our work and briefly surveyed important related work.

In Chapter 2 "Routing Dependability" we have given accurate definitions and presented a conceptual model of the subject of this dissertation, namely routing dependability for ad hoc and cellular networks. A literature survey and a thorough discussion of the reviewed work has been performed. Moreover, we have precisely described the dimensions of dependability for the respective network classes. For both classes of networks we have derived the most important characteristics influencing dependability. Based on these characteristics, we have deduced the dimensions of dependability for further study. In particular, we have identified the dimension of network control to be of special interest for infrastructure-based networks such as cellular networks. The corresponding work items have been performed in Part II of this dissertation. For ad hoc networks we have shifted our focus to the effects induced by node misbehavior. We have carefully analyzed these effects and provided for means to adapt the network to mitigate the loss of dependability in Part III of our work.

In Chapter 3 "Mobility in Cellular and Ad hoc Networks" we have presented a novel approach towards a mobility/workload model describing the macroscopic effects observed within urban areas. Moving from traditional cellular networks to architectures that include hot spots and wireless local area networks results in smaller cells and an increase in capacity. Our model addresses the lack of adequate workload models to capture these requirements. We have formulated an analytical model, which is a hybrid of an empirical mobility model and a synthetic traffic model. For further usage, we have instantiated the model for the case of Darmstadt, a large city in Germany. The model predictions have been studied and we found that mobility introduces major fluctuations of user densities and user traffic. This concluded the first part of this dissertation, which has mainly served as motivation and framework for the rest of our work where concrete problems of routing systems for cellular and ad hoc networks have been discussed.

The challenge of routing dependability for cellular network has been investigated in Part II of the dissertation. In Chapter 4 "Dependable Routing for Cellular Networks with

Static Infrastructure” we have discussed a cellular network architecture to improve the performance of radio access networks for metropolitan scenarios. In particular, an architecture to extend the scope of the routing domain towards the edge of the network and to allow for more flexible topologies than strict tree hierarchies has been designed. We have evaluated the performance of various routing strategies within this architecture by means of a simulation study. Using the workload modeled in Chapter 3 we have been able to show multipath routing is able to enhance the performance of the network and allows for efficient resource management/load-balancing within our novel network architecture.

In Chapter 5 “Dependable Routing for Cellular Networks with Variable Infrastructure” we have approached the challenge of routing dependability and performance from a perspective permitting variability of infrastructure components. We have designed a novel network architecture and feasible routing strategies for use within this architecture. The architecture bases on the assumption that the network topology can be reconfigured on demand and allows for easy integration of hot spots. We have formulated a novel routing algorithm, Algorithm-M^2DR, which optimizes the network performance on the two orthogonal dimensions path diversity of the network and class diversity of the traffic, respectively. As a proof of concept the algorithm has been implemented in the ns-2 simulation environment. A simulation study has shown the superior performance of our algorithms compared with selected state-of-the-art routing algorithms. While providing for excellent QoS routing performance, our algorithm maintains the simplicity of a decentralized and distributed routing algorithm, though.

Part III of this dissertation has dealt with dependable routing for ad hoc networks. In Chapter 6 “Analytical Model of the Route Acquisition Process” we have formulated a realistic model of ad hoc routing protocols to serve as a versatile tool for our study of dependability in ad hoc networks. The model predicts the route-length distribution, which we have put into novel use to describe the overall performance of ad hoc routing systems. In particular, we have formulated an analytically model of the route acquisition process of the AODV protocol [PBRD03]. The model includes protocol optimizations such as expanding ring search, reply by intermediate, and probabilistic flooding.

The results of Chapter 6 form the basis for the analytical modeling of non-cooperating nodes in Chapter 7 “Node Misbehavior—Classification, Analytical Model, and Performance Trade-off”. Here, we have discussed the effects of node misbehavior in ad hoc networks in detail. We have derived well-defined classes of misbehavior that are suitable for analytical study and have adjoined these types of misbehavior to our analytical model. Moreover, the dynamic effects of node misbehavior have been studied by means of an extensive simulation study. Similar to Part II of our work, we have implemented the respective algorithms and the misbehavior of nodes in a network simulator. Our results show that inactive nodes only moderately harm ad hoc networks, while selfish nodes and black holes may have devastating influence on the routing process. Moreover, the probabilistic flooding is able to achieve significant performance optimization without sacrificing the robustness of the system.

8.2 Conclusion

The major conclusion that can be drawn from this dissertation is that routing dependability acts as one of the key factors to determine if a routing system is able to maintain acceptable performance even under critical conditions. We have shown this for the area of cellular and ad hoc networks as described in the preceding chapters and summarized in Section 8.1. Dependability comes in many forms, which are connected to the type and characteristics of the network, tough. We argue to question the “cure-all” approach of Internet technology. In fact the Internet has never been designed to support the heterogeneity, dynamics, and mobility it faces in our particular scenarios. In cellular networks, for example, the combination of these factors with the currently deployed services and applications break many of the architectural ties of the original Internet model. Our approach to introduce routing technology into the radio access part of the network has been shown to be very promising, though. The basics of our architecture and algorithms have been borrowed from existing work in cellular networks and the Internet. We have also been inspired by paradigms from the area of ad hoc networks to address the mobility related aspects. Enhancements of the dependability and performance by means of our algorithm aided in approaching the achievable performance region of routed radio access networks beyond the capabilities of Internet technology. Of course, there are further operational issues that need to be taken into account in real world networks. We also argue for a more engineered and planned approach than available today, to reach an appropriate level of dependability in routing networks. For example, the class of ad hoc networks has only recently been recognized as a very versatile technology to enhance existing communication systems. However, many of the protocol designs focus on specific and high level protocol optimizations while sacrificing a robust and dependable routing foundation. We have been able to show that this leads to inferior routing systems, which are not well equipped to withstand random failures or willful threats of non-cooperating nodes. We believe that the quest for dependable routing in wireless networks will not come to an end soon. Upcoming scenarios such as disruption tolerant networking again introduce novel demands for the routing system while featuring most challenging characteristics. We conclude that there is no panacea for the routing dependability problem. However, there exist solutions to particular problems hindering the seamless and dependable operation of wireless networks as has been shown in this dissertation.

8.3 Outlook

Our contributions advance the state-of-the-art in dependable routing in various areas. We have presented solutions to the challenges outlined in the introduction to our work. These solutions do not only answer the research problems we identified but also open up novel avenues of in research. Next, we present some appealing challenges, which can be taken up on top of our results.

In routing networks, the basic concepts of survivability, trustworthiness, and dependability have only recently drawn attention. In particular, our work provided the basic foundations of routing dependability for cellular and ad hoc networks. These foundations may further be improved and extended. While our categorization of dimensions mainly focuses on the aspects of mobility, possible future directions include, for example, the study of dependability in heterogeneous systems and the investigation of characteristics other than mobility (scalability, etc.).

We presented a workload model for use in the special scenario of metropolitan scale radio access networks. To be applicable for other scenarios as well, the mobility part of our workload model can be extended with a (synthetic) microscopic mobility model. The major advantage of such a model would be the realism for microscopic and macroscopic analysis. The possible applications of this extension include the modeling of large scale metropolitan ad hoc networks—a research field that has not drawn attention, yet.

Our architecture and algorithm proposed for routing in cellular networks are capable of increasing the performance of the routing system. We believe that our results can be transferred to other networking domains as well. Novel networking paradigms such as, for example, mesh networks combine infrastructure-based and ad hoc mechanisms. Our algorithm presents one possibility to deal with the resulting topology dynamics. We believe that further gains can be obtained if cross-layer interactions with the link layer and the physical layer are explored, though.

Our contributions in the area of ad hoc networks can be regarded as initial step towards a novel research field. The extension of our analytical model to cover the dynamics of mobility and misbehavior is certainly a very challenging problem for future research in this area. Moreover, the design of dependable and well performing ad hoc routing frameworks can make beneficial use of our results.

Last but not least, we regard the self-organized establishment of routing dependability to be of high interest. In the area of ad hoc and sensor networks this may translate into mechanisms to establish trust using totally decentralized, distributed and self-organizing mechanisms.

Bibliography

All webpages cited in this work have been checked in October 2004. However, due to the dynamic nature of the World Wide Web, webpages can change.

[3GPP03] 3GPP. QoS Concept and Architecture. TS 23.107, Release 5, Version 5.9.0, June 2003.

[3GPP04] 3GPP. 3rd Generation Partnership Project. Available at http://www.3gpp.org/, October 2004.

[ACE+02] D. O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao. Overview and Principles of Internet Traffic Engineering. Informational Request for Comments, RFC 3272, May 2002.

[Ack02] R. Ackermann. *Gateways and Components for Supplementary IP Telephony Services in Heterogeneous Environments*. PhD thesis, Department of Computer Science, Darmstadt University of Technology, Germany, July 2002.

[ALR00] A. Avizienis, J.-C. Laprie, and B. Randell. Fundamental Concepts of Dependability. In *Proceedings of the 3rd IEEE Information Survivability Workshop (ISW 2000), Boston, MA, USA*, pages 7–12. IEEE, October 2000.

[AWK+99] G. Apostolopoulos, D. Williams, S. Kamat, R. Guerin, A. Orda, and T. Przygienda. QoS Routing Mechanisms and OSPF Extensions. Experimental Request for Comments, RFC 2676, September 1999.

[BB02a] S. Buchegger and J.-Y. Le Boudec. Nodes Bearing Grudges: Towards Routing Security, Fairness, and Robustness in Mobile Ad Hoc Networks. In *Proceedings of the 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing (EUROMICRO 2002), Dortmund, Germany*, pages 403–410. IEEE, January 2002.

[BB02b] S. Buchegger and J.-Y. Le Boudec. Performance Analysis of the CONFIDANT Protocol. In *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc 2002), Lausanne, Switzerland*, pages 226–236. ACM Press New York, June 2002.

[BBC+98] D. Black, S. Blake, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Services. Informational Request for Comments, RFC 2475, December 1998.

[BCS94] R. Braden, D. Clark, and S. Shenker. Integrated Services in the Internet Architecture: an Overview. Informational Request for Comments, RFC 1633, June 1994.

[Bet01a] C. Bettstetter. Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(3):55–66, July 2001.

[Bet01b] C. Bettstetter. Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks. In *Proceedings of the 4th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2001), Rome, Italy*, pages 19–27. ACM Press New York, July 2001.

[BHPC02] C. Bettstetter, H. Hartenstein, and X. Perez-Costa. Stochastic Properties of the Random Waypoint Mobility Model: Epoch Length, Direction Distribution, and Cell Change Rate. In *Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2002), Atlanta, GE, USA*, pages 7–14. ACM Press New York, 2002.

[BKOKR04] T. Breyer, M. Klein, P. Obreiter, and B. König-Ries. Activity-Based User Modeling in Service-Oriented Ad-Hoc-Networks. In *Proceedings of First Working Conference on Wireless On-demand Network Systems, WONS 2004, Madonna Di Campiglio, Italy*, pages 43–56. Springer-Verlag Berlin/Heidelberg/New York, January 2004.

[Bla02] L. Blazevic. *Scalable Routing Protocols with Applications to Mobility*. PhD thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, February 2002.

[BMY04] A. Barbir, S. Murphy, and Y. Yang. Generic Threats to Routing Protocols. Internet Draft, draft-ietf-rpsec-routing-threats-06, April 2004. Work in progress.

[BRS02] D. M. Blough, G. Resta, and P. Santi. A Statistical Analysis of the Long-Run Node Spatial Distribution in Mobile Ad Hoc Networks. In *Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2002), Atlanta, GE, USA*, pages 30–37. ACM Press New York, 2002.

[BVBR02] A. Balachandran, G. Voelker, P. Bahl, and P. V. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN. *ACM SIGMETRICS Performance Evaluation Review*, 30(1):195–205, June 2002.

[BW02] C. Bettstetter and C. Wagner. The spatial node distribution of the random waypoint mobility model. In *Proceedings of the First German Workshop on Mobile Ad-Hoc Networks (WMAN 2002), Ulm, Germany*, pages 41 –58. GI, March 2002.

[CAI96] Collaborative Advanced Internet Research Network (CAIRN). Available at http://www.isi.edu/CAIRN/, September 1996.

[Car00] B. Carpenter. Internet Transparency. Informational Request for Comments, RFC 2775, February 2000.

[CAT88] C. G. Cassandras, M. V. Abidi, and D. Towsley. Distributed Routing with on-line Marginal Delay Estimation. In *Proceedings of the 7th Annual Joint Conference of the IEEE Computer and Communication Societies (INFOCOM 1988), New Orleans, LA, USA*, pages 603–612. IEEE, March 1988.

[CBD02] T. Camp, J. Boleng, and V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communications & Mobile Computing (WCMC): Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.

[CBFP03] D. Clark, R. Braden, A. Falk, and V. Pingali. FARA: Reorganizing the Addressing Architecture. *ACM SIGCOMM Computer Communication Review*, 33(4):313–321, October 2003.

[CCS96] I. Castineyra, N. Chiappa, and M. Steenstrup. The Nimrod Routing Architecture. Informational request for Comments, RFC 1992, August 1996.

[CG74] D. G. Cantor and M. Gerla. Optimal Routing in a Packet-switched Computer Network. *IEEE Transactions on Computers*, C-23(10):1062–1069, October 1974.

[CHM+03] J. Crowcroft, S. Hand, R. Mortier, T. Roscoe, and A. Warfield. Plutarch: An Argument for Network Pluralism. *ACM SIGCOMM Computer Communication Review*, 33(4):258–266, October 2003.

[CJ03] T. Clausen and P. Jacquet. Optimized Link State Routing Protocol (OLSR). Experimental Request for Comments, RFC 3626, October 2003.

[CK74] V. G. Cerf and R. E. Kahn. A Protocol for Packet Network Intercommunication. *IEEE Transactions on Communications*, 22(5):637–648, May 1974.

[Cla88] D. Clark. The Design Philosophy of the DARPA Internet Protocols. *ACM SIGCOMM Computer Communication Review*, 18(4):106–114, August 1988.

[CN98] S. Chen and K. Nahrstedt. An Overview of Quality-of-Service Routing for the Next Generation High-speed Networks: Problems and Solutions. *IEEE Network Magazine, Special Issue on Transmission and Distribution of Digital Video*, 12(6):64–79, November-December 1998.

[CN99] S. Chen and K. Nahrstedt. Distributed Quality-of-Service Routing in Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications*, 17:1488–1505, August 1999.

[CST96] P. Camarda, G. Schiraldi, and F. Talucci. Mobility Modeling in Cellular Communication Networks. In *Proceedings of 21st IEEE Conference on Local Computer Networks (LCN 1996), 1996, Minneapolis, MN, USA*, pages 518–525. IEEE, October 1996.

[Daf72] S. Dafermos. The Traffic Assignment Problem for Multiclass-User Transportation Networks. *Transportation Science*, 6:73–88, February 1972.

[DDD+00] D. Daly, D. D. Deavours, J. M. Doyle, P. G. Webster, and W. H. Sanders. Möbius: An Extensible Tool for Performance and Dependability Modeling. In *Proceedings of the 11th International Conference on Computer Performance Evaluation: Modelling Techniques and Tools (TOOLS 2000), Schaumburg, IL, USA*, pages 332–336. Springer-Verlag Berlin/Heidelberg/New York, March 2000.

[Dij59] E. W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1(1):269–271, January 1959.

[DPR00] S. R. Das, C. E. Perkins, and E. M. Royer. Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks. In *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000), Tel Aviv, Israel*, volume 1, pages 3–12. IEEE, March 2000.

[DTH02] O. Dousse, P. Thiran, and M. Hasler. Connectivity in Ad-hoc and Hybrid Networks. In *Proceedings of the 21th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, NY, USA*, volume 2, pages 1079–1088. IEEE, June 2002.

[Eck04] C. Eckert. *IT-Sicherheit: Konzepte, Verfahren, Protokolle, 3. berarbeitete und erweiterte Auflage*. R. Oldenbourg Verlag, 2004. ISBN 3-486-20000-3.

[EFL+99] R. J. Ellison, D. A. Fischer, R. C. Linger, H. F. Lipson, T. Longstaff, and N. R. Mead. Survivable Network Systems: an Emerging Discipline. Technical Report CMU/SEI-97-TR-013, Carnegie Mellon University, Pittsburg, PA, USA, November 1999.

[Ell01] C. Elliott. Building the Wireless Internet. *IEEE Spectrum*, pages 14–16, January 2001.

[EVB01] J. Eberspächer, H.-J. Vögel, and C. Bettstetter. *GSM Switching, Services and Protocols: Second Edition*. John Wiley & Sons, March 2001. ISBN 0-471-49903-X.

[Fal03] K. Fall. A Delay-tolerant Network Architecture for Challenged Internets. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2003), Karlsruhe, Germany*, pages 27–34. ACM Press New York, August 2003.

[Fee99] L. M. Feeney. A Taxonomy for Routing Protocols in Mobile Ad Hoc Networks. Technical Report T99/07, Swedish Institute of Computer Science, October 1999. Available at ftp://ftp.sics.se/pub/SICS-reports/Reports/SICS-T--99-07--SE.ps.Z.

[FV04] K. Fall and K. Varadhan. The Network Simulator - ns-2. Software and manual available at http://www.isi.edu/nsnam/ns/, October 2004.

[Gal77] R. G. Gallager. A Minimum Delay Routing Algorithm Using Distributed Computation. *IEEE Transactions on Communications*, 25:73–84, January 1977.

[GK00] P. Gupta and P. R. Kumar. The Capacity of Wireless Networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[GK03] P. Gupta and P. R. Kumar. Towards an Information Theory of Large Networks: An Achievable Rate Region. *IEEE Transactions on Information Theory*, 49(8):1877–1894, August 2003.

[GSM04] GSM Association. GSM Subscriber Statistics. Available at http://www.gsmworld.com/news/statistics/pdf/mar04.pdf, March 2004.

[GT01] M. Grossglauser and D. N. Tse. Mobility Increases the Capacity of Ad-hoc Wireless Networks. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), Anchorage, AK, USA*, pages 1360–1369. IEEE, April 2001.

[GV02] M. Gastpar and M. Vetterli. On the Capacity of Wireless Networks: the Relay Case. In *Proceedings of the 21th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, NY, USA*, volume 3, pages 1577–1586. IEEE, June 2002.

[Har03] R. Hartmann. Experimental Analysis of Gossip-based Strategies to Optimize the Reliability of Routing Mechanisms within Mobile Ad Hoc Networks. Diploma thesis KOM-D-181, Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany, 2003.

[HBC01] J.-P. Hubaux, L. Buttyan, and S. Capkun. The Quest for Security in Mobile Ad hoc Networks. In *Proceedings of the 2nd ACM International Symposium on Mobile Ad hoc Networking & Computing (MobiHoc 2001), Long Beach, CA, USA*, pages 146–155. ACM Press, October 2001.

[HBE+01] H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radimirsch, and D. Vollmer. Position-aware Ad hoc Wireless Networks for Inter-vehicle Communications: The Fleetnet Project. In *Proceedings of the 2nd ACM International Symposium on Mobile Ad hoc Networking & Computing (MobiHoc 2001), Long Beach, CA, USA*, pages 259–262. ACM Press New York, October 2001.

[Hel99] B. E. Helvik. Dependability Issues in Smart Networks. In *Proceedings of IFIP, Fifth International Conference on Intelligence (SMARTNET 1999), Bangkok, Thailand*, pages 53–76. Kluwer Academic Publishers, November 1999.

[Hec04] O. Heckmann. *A System-oriented Approach to Efficiency and Quality of Service for Internet Service Providers*. PhD thesis, Department of Computer Science, Darmstadt University of Technology, Germany, December 2004.

[HGBV01] J. P. Hubaux, Th. Gross, J. Y. Le Boudec, and M. Vetterli. Towards Self-organized Mobile Ad hoc Networks: the Terminodes Project. *IEEE Communications Magazine*, 39:118–124, January 2001.

[HGPC99] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. A Group Mobility Model for Ad Hoc Wireless Networks. In *Proceedings of the 2nd ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 1999), Seattle, WA, USA*, pages 53–60. ACM Press New York, August 1999.

[HHL02] Z. Haas, J. Y. Halpern, and E. L. Li. Gossip-based Ad Hoc Routing. In *Proceedings of the 21th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, NY, USA*, volume 3, pages 1707–1716. IEEE, June 2002.

[HJP02] Y.-C. Hu, D. B. Johnson, and A. Perrig. SEAD: Secure Efficient Distance Vector Routing in Mobile Wireless Ad Hoc Networks. In *Proceedings of Fourth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2002)*, pages 3–13. IEEE, June 2002.

[HKS+03] M. Hollick, T. Krop, J. Schmitt, H.-P. Huth, and R. Steinmetz. Comparative Analysis of Quality of Service Routing in Wireless Metropolitan Area Networks. In *Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks (LCN 2003), Bonn/Königswinter, Germany*, pages 470–479. IEEE Computer Society Press, Los Alamitos, October 2003. ISBN 0-7695-2037-5.

[HMKR04] M. Hollick, I. Martinovic, T. Krop, and I. Rimac. A Survey on Dependable Routing in Sensor Networks, Ad hoc Networks, and Cellular Networks. In *Proceedings of the 30th EUROMICRO Conference (EUROMICRO 2004), Rennes, France*, pages 495–502. IEEE Computer Society Press, Los Alamitos, September 2004.

[Ho87] Y.-C. Ho. Performance Evaluation and Perturbation Analysis of Discrete Event Dynamic Systems. *IEEE Transactions on Automatic Control*, 32:563–572, July 1987.

[Hol00] M. Hollick. The Evolution of Mobile IP Towards Security. In *Sicherheit in Netzen und Medienströmen, Tagungsband des GI-Workshops "Sicherheit in Mediendaten"*, pages 38–49. Springer-Verlag Berlin/Heidelberg/New York, September 2000. ISBN 3-540-67926-X.

[Hol01] M. Hollick. Secure Service Centered Networking for Nomadic Usage. In *Proceedings of Communications and Multimedia Security Issues of the New Century, IFIP TC6 / TC11 Joint Working Conference on Communications and Multimedia Security (CMS 2001), Darmstadt, Germany*, pages 375–380. Kluwer Academic Publishers Boston/Dordecht/London, May 2001.

[HPJ02] Y.-C. Hu, A. Perrig, and D. B. Johnson. Ariadne: A Secure On-demand Routing Protocol for Ad hoc Networks. In *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking (MobiCom 2002), Atlanta, GA, USA*, pages 12–23. ACM Press New York, September 2002.

[HPSS03] O. Heckmann, K. Pandit, J. Schmitt, and R. Steinmetz. KOM ScenGen - The Swiss Army Knife for Simulation and Emulation Experiments. In *Proceedings of the International Workshop on Multimedia Interactive Protocols and Systems, MIPS 2003, Naples, Italy*, pages 91–106, November 2003.

[HR86] D. Hong and S. S. Rappaport. Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Handoff Procedures. *IEEE Transactions on Vehicular Technology*, 35(3):77–92, August 1986. An updated version of the paper as of June 1999 is available as Technical Report No. 773, College of Engineering and Applied Sciences, Department of Electrical and Computer Engineering, State University of New York, Stony Brook, New York 11794-2350.

[HS01] M. Hollick and R. Steinmetz. Security Goals for Service Discovery in Ad-hoc Environments. In *Proceedings of International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2001), Split, Croatia*, pages 1025–1032. FESB-Split, October 2001.

[HS02] H.-Y. Hsieh and R. Sivakumar. Towards a Hybrid Network Model for Wireless Packet Data Networks. In *Proceedings of IEEE International Symposium on Computers and Communications (ISCC 2002), Taormina, Italy*, pages 264–274. IEEE, July 2002.

[HSHS04] M. Hollick, Jens Schmitt, Hans-Peter Huth, and Joachim Sokol. MobQoS - Mobility Aware QoS Routing. Available at http://www.kom.tu-darmstadt.de/Research/MobQoS/MobQoS.html, October 2004.

[Hus00] G. Huston. Next Steps for the IP QoS Architecture. Informational Request for Comments, RFC 2290, November 2000.

[IEEE99a] IEEE, Computer Society LAN MAN Standards Committee. IEEE Std. 802.11-1999, IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Network, Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1999.

[IEEE99b] IEEE, Computer Society LAN MAN Standards Committee. IEEE Std. 802.11a-1999, IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Networks, Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 1: High-speed Physical Layer in the 5 GHz Band, 1999.

[IEEE01] IEEE, Computer Society LAN MAN Standards Committee. IEEE Std. 802.11b-1999/Cor1-2001, IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Networks, Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 2: Higher-speed Physical Layer (PHY) Extension in the 2.4 GHz Band, Corrigendum1, 1999.

[IEEE02] IEEE, Computer Society LAN MAN Standards Committee. IEEE Std. 802.3-2002, IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Network, Specific Requirements, Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications, 2002.

[IEEE03] IEEE, Computer Society LAN MAN Standards Committee. IEEE Std. 802.11g-2003, IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Networks, Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 4: Further Higher-speed Physical Layer (PHY) Extension in the 2.4 GHz Band, 1999.

[IEEE04] IEEE, Institute of Electrical and Electronics Engineers. IEEE CS Technical Committee on Fault Tolerant Computing. Available at http://www.dependability.org/, October 2004.

[IFIP04] IFIP, International Federation for Information Processing. IFIP WG10.4 on Dependable Computing and Fault Tolerance. Available at http://www.dependability.org/, October 2004.

[ITU94] ITU, International Telecommunication Union, Telecommunication Standardization Sector. ITU-T Recommendation E.800: Terms and Definitions related to Quality of Service and Network Performance including Dependability, August 1994.

[Jai91] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience, New York, NY, USA, May 1991. ISBN 0-471-50336-3.

[JB99] E. Jugl and H. Boche. A New Mobility Model for Performance Evaluation of Future Mobile Communication Systems. In *Proceedings of IEEE International Conference on Communications (ICC 1999), Vancouver, BC, Canada*, pages 1751–1755. IEEE, June 1999.

[JFA03] A. Jonsson, M. Folke, and B. Ahlgren. The Split Naming/Forwarding Network Architecture. In *Proceedings of the First Swedish National Computer Networking Workshop, Arlandastad, Sweden*, September 2003.

[JM96] D. B. Johnson and D. A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In *Mobile Computing, 3rd Edition, Tomasz Imielinski and Hank Korth (Eds.)*, pages 153–181. Kluwer Academic Publishers, 1996.

[JMH04] D. B. Johnson, D. A. Maltz, and Y.-C. Hu. The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR). Internet Draft, draft-ietf-manet-dsr-09.txt, July 2004.

[JP04] A. Josang and S. Lo Presti. Analysing the Relationship Between Risk and Trust. In *Proceedings of the Second International Conference on Trust Management, iTrust 2004, Oxford, UK*, pages 135–145. Springer-Verlag Berlin/Heidelberg/New York, April 2004.

[JPA04] D. B. Johnson, C. E. Perkins, and J. Arkko. IP Mobility Support in IPv6. Proposed Internet Standard, RFC 3775, June 2004.

[KA98] S. Kent and R. Atkinson. Security Architecture for the Internet Protocol. Proposed Internet Standard, RFC 2401, November 1998.

[Kar00] M. Karsten. *QoS Signalling and Charging in a Multi-service Internet using RSVP*. PhD thesis, Department of Computer Science. Darmstadt University of Technology, Juli 2000.

[Kar03] F. Kargl. *Sicherheit in Mobilen Ad hoc Netzwerken*. PhD thesis, Faculty of Computer Science, University of Ulm, Germany, October 2003.

[KE02] D. Kotz and K. Essien. Analysis of a Campus-Wide Wireless Network. In *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking (MobiCom 2002), Atlanta, GA, USA*, pages 107–118. ACM Press New York, September 2002.

[Kel91] J. J. Kelleher. Tactical Communication Network Modelling and Reliability Analysis: Overview. Technical Report AD-A245339, J.S. Lee Associates, Inc., Rockville, ML, USA, November 1991. Available at http://w3.antd.nist.gov/wctg/netanal/gphthy_kelleher.pdf.

[Ken02] T. Kenyon. *Data Networks: Routing, Security and Performance Optimization*. Harcourt, July 2002. ISBN 1-555-58271-0.

[KG83] H. Kobayashi and M. Gerla. Optimal Routing in Closed Queuing Networks. *ACM Transactions on Computer Systems*, 1(4):294–310, 1983.

[Kit88] R. Kitamura. An Evaluation of Activity-based Travel Analysis. *Transportation*, 15(1):9–34, 1988.

[KKKM02] F. A. Kuipers, T. Korkmaz, M. Krunz, and P. Van Mieghem. A Review of Constraint-Based Routing Algorithms. *IEEE Communications Magazine*, 40(12):50–55, December 2002.

[KKW+03] H. T. Kaur, S. Kalyanaraman, A. Weiss, S. Kanwar, and A. Gandhi. BANANAS: An Evolutionary Framework for Explicit and Multipath Routing in the Internet. *ACM SIGCOMM Computer Communication Review*, 33(4):277–288, October 2003.

[Kle76] L. Kleinrock. *Queueing Systems, Volume 2: Computer Applications*. Wiley-Interscience, January 1976. ISBN 0-471-49111-X.

[Kle78] L. Kleinrock. Principles and Lessons in Packet Communications. *Proceedings of the IEEE, Special Issue on Packet Communications*, 66(11):1320–1329, November 1978.

[Kle96] L. Kleinrock. Nomadic Computing. In *Proceedings of the IFIP/ICCC International Conference on Information Network and Data Communication, Trondheim, Norway*, pages 223–233, June 1996.

[Kle01] L. Kleinrock. Breaking Loose. *Communications of the ACM*, 44(9):41–45, September 2001.

[KNT01] E. Kail, G. Nemeth, and Z. R. Turanyi. The Effect of the Transmission Range on the Capacity of Ideal Ad Hoc Networks. In *Proceedings of the 4th International Symposum on Wireless Personal Multimedia Communications (WPMC 2001), Aalborg, Denmark*, September 2001.

[Kro02] T. Krop. Experimental Analysis of QoS-Routing Mechanisms for Load-Balancing in Metropolitan Radio Access Networks. Diploma thesis KOM-D-182, Multimedia Communications Lab (KOM), Department of Computer Science, Darmstadt University of Technology, Germany, December 2002.

[KS78] L. Kleinrock and J. Silvester. Optimum Transmission Radii in Packet Radio Networks or Why Six is a Magic Number. In *Proceedings of National Telecommunications Conference, Birmingham, AL, USA*, pages 4.3.1–1.3.5, December 1978.

[LAN04] LANL, Los Alamos National Laboratory. TRANSIMS - TRansportation ANalysis SIMulation System. Available at http://transims.tsasa.lanl.gov/, January 2004.

[Lap85] J.-C. Laprie. Dependable Computing and Fault Tolerance: Concepts and Terminology. In *Proceedings of the 15th IEEE International Symposium on Fault-Tolerant Computing (FTCS 1985), Ann Arbor, MI, USA*, pages 2–11, June 1985.

[LBC+01] J. Li, C. Blake, D. S. J. De Couto, H. Imm Lee, and R. Morris. Capacity of Ad Hoc wireless networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom 2001), Rome, Italy*, pages 61–69. ACM Press New York, 2001. ISBN1-58113-422-3.

[LBRP03] S.-J. Lee, E. M. Belding-Royer, and C. E. Perkins. Scalability Study of the Ad hoc On-demand Distance Vector Routing Protocol. *International Journal on Network Management*, 13:97–114, 2003.

[LC03] Z. Liao and M. Tow Cheung. Challenges to Internet e-Banking. *Communications of the ACM*, 46(12):248–250, December 2003.

[LCW97] D. Lam, D. C. Cox, and J. Wilson. Teletraffic Modeling for Personal Communication Services. *IEEE Communications Magazine*, 35(2):79–87, February 1997.

[LH00] Y.-D. Lin and Y.-C. Hsu. Multihop Cellular:A New Architecture for Wireless Communications. In *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000), Tel Aviv, Israel*, pages 1273–1282. IEEE, March 2000.

[LLW+02] H. Li, M. Lott, M. Weckerle, W. Zirwas, and E. Schulz. Multihop Communications in Future Mobile Radio Networks. In *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002), Lisboa, Portugal*, volume 1, pages 54–58, September 2002.

[LWZ+03] M. Lott, M. Weckerle, W. Zirwas, H. Li, and E. Schulz. Hierarchical Cellular Multihop Networks. In *Proceedings of the Fifth European Personal Mobile Communications Conference (EPMCC 2003), Glasgow, UK*, March 2003.

[Max75] N. F. Maxemchuk. Dispersity Routing. In *Proceedings of International Conference on Communications (ICC 1975), San Francisco, CA, USA*, pages 41.10–41.13, June 1975.

[MBH+02] J. Manner, L. Burness, E. Hepworth, A. Lopez, and E. Mitjana. Provision of QoS in Heterogeneous Wireless IP Access Networks. In *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002), Lisbon, Portugal*, volume 2, pages 530–534, September 2002.

[Men04] M. Menth. *Efficient Admission Control and Routing for Resilient Communication Networks*. PhD thesis, University of Würzburg, July 2004.

[MGLB00] S. Marti, T. J. Giuli, K. Lai, and M. Baker. Mitigating Routing Misbehavior in Mobile Ad Hoc Networks. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom 2000), Boston, MA, USA*, pages 255–265. ACM Press New York, August 2000.

[Mil01] L. E. Miller. Distribution of Link Distances in a Wireless Network. *Journal of Research of the National Institute of Standards and Technology*, 106(2):401–412, March-April 2001.

[MLA98] J. G. Markoulidakis, G. L. Lyberopoulos, and M. E. Anagnostou. Traffic Model for Third Generation Cellular Mobile Telecommunication Systems. *Wireless Networks*, 4:389–400, July 1998.

[MLM+02] J. Manner, A. Lopez, A. Mihailovic, H. Velayos, E. Hepworth, and Y. Khouaja. Evaluation of Mobility and QoS Interaction. *Computer Networks*, 38(2):137–163, 2002.

[MLTS97] J. G. Markoulidakis, G. L.Lyberopoulos, D. F. Tsirkas, and E. D. Sykas. Mobility Modeling in Third-generation Mobile Telecommunications Systems. *IEEE Personal Communications*, 4(4):41–56, August 1997.

[MM02] P. Michiardi and R. Molva. Simulation-based analysis of security exposures in mobile ad hoc networks. In *Proceedings of the Conference on Next Generation Wireless Networks: Technologies, Protocols, Services and Applications (European Wireless 2002)*, February 2002.

[Mog04] P. S. Mogre. Near-Optimal Multiclass Minimum-Delay Routing for Cellular Networks with Variable Topology. Master thesis KOM-D-209, Department of Computer Science and Engineering, Indian Institute of Technology Guwahati and Multimedia Communications Lab (KOM), Department of Computer Science, Darmstadt University of Technology, Germany, May 2004.

[Moy98] J. Moy. OSPF Version 2. Proposed Internet Standard, RFC 2328, April 1998.

[Mul03] J. P. Mullen. Robust Approximations to the Distribution of Link Distances in a Wireless Network Occupying a Rectangular Region. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(2):80–91, April 2003.

[MWE02] P. B. Gove (Editor), Merriam-Webster (Editor). *Webster's Third New International Dictionary, Unabridged, 3rd edition*. Merriam-Webster, Inc., January 2002. ISBN 0-877-79201-1.

[Nan93] S. Nanda. Teletraffic Models for Urban and Suburban Microcells: Cell Sizes and Handoff Rates. *IEEE Transactions on Vehicular Technology*, 42(4):673–682, November 1993.

[Net72] M. Netter. Equilibrium and Marginal Cost Pricing on a Road Network with Several Traffic Flow Types. In *Proceedings of the 5th International Symposium on the Theory of Traffic Flow and Transportation*, pages 155–163. Elsevier, 1972.

[NK84] R. Nelson and L. Kleinrock. The Spatial Capacity of a Slotted ALOHA Multihop Packet Radio Network with Capture. *IEEE Transactions on Communications*, 32(6):684–694, June 1984.

[NSA+04] N. Niebert, A. Schieder, H. Abramowicz, G. Malmgren, J. Sachs, U. Horn, C. Prehofer, and H. Karl. Ambient Networks – An Architecture for Communication Networks Beyond 3G. *IEEE Wireless Communications, Special Issue on 4G Mobile Communications – Towards Open Wireless Architecture*, 11(2):14–22, April 2004.

[NTV01] G. Nemeth, Z. R. Turanyi, and A. Valko. Throughput of Ideally Routed Wireless Ad-hoc Networks. *ACM Mobile Computer and Communication Review (MC2R)*, 5(4):40–46, October 2001.

[NW02] K. Nahrstedt and J. Wang. Hop-by-hop Routing Algorithms for Premium-class Traffic in DiffServ Networks. In *Proceedings of the 21th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, NY, USA*, volume 2, pages 705–714. IEEE, June 2002.

[Opp95] N. Oppenheim. *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*. John Wiley & Sons, January 1995. ISBN 0-471-55723-4.

[PBRD03] C. E. Perkins, E. M. Belding-Royer, and S. R. Das. Ad hoc On-demand Distance Vector (AODV) Routing. Experimental Request for Comments, RFC 3561, July 2003.

[Per99] R. Perlman. *Interconnections: Bridges, Routers, Switches, and Internetworking Protocols (2nd Edition)*. Addison-Wesley, September 1999. ISBN 0-201-63448-1.

[Per02] C. Perkins. IP Mobility Support for IPv4. Proposed Internet Standard, RFC 3344, August 2002.

[Pet04] M. Peter. Mehrwegerouting in Ad Hoc Kommunikationsnetzen: Experimentelle Analyse. Diploma thesis KOM-D-189, Multimedia Communications Lab (KOM), Department of Computer Science, Darmstadt University of Technology, Germany, April 2004.

[PH02] P. Papadimitratos and Z. Haas. Secure Routing for Mobile Ad hoc Networks. In *Proceedings of the Conference on Communication Networks and Distributed Systems Modeling and Simulation (SCS 2002), San Antonio, TX, USA*, pages 27–31, January 2002.

[PH03] P. Papadimitratos and Z. Haas. Secure Data Transmission in Mobile Ad hoc Networks. In *Proceedings of the 2003 ACM Workshop on Wireless Security (WISE 2003), San Diego, CA, USA*, pages 41–50. ACM Press New York, September 2003.

[PKH+02] C. Prehofer, W. Kellerer, R. Hirschfeld, H. Berndt, and K. Kawamura. An Architecture Supporting Adaption and Evolution in 4G Mobile Communications Systems. *Journal of Communications and Networks*, 4(4):336–343, December 2002.

[PR02] P. Paul and S.-V. Raghavan. Survey of QoS Routing. In *Proceedings of the Fifteenth International Conference on Computer Communication, ICCC'02, Mumbai, India*, August 2002.

[PSL01] S. Park, Y. Serbest, and S. Li. Minimum Congestion Traffic Engineering with Multi-resource Constraint. In *Proceedings of IEEE International Conference on Computer Communications and Networks, ICCCN 2001, Scottsdale, AZ, USA*, pages 130–137, October 2001.

[QYZS03] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker. On selfish routing in internet-like environments. In *Proceedings of the Annual Conference of the Special Interest Group on Data Communication (SIGCOMM 2003), Karlsruhe, Germany*, pages 151–162, August 2003.

[Rap01] T. Rappaport. *Wireless Communications: Principles and Practice, 2nd Edition*. Prentice Hall PTR, Boston, MA, USA, December 2001. ISBN 0-130-422-320.

[Rev90] P. Revesz. *Random Walk in Random and Non-random Environments*. World Scientific, Singapore, September 1990. ISBN 981-02-0237-7.

[RMdS00] M. N. Rocha, G. R. Mateus, and S. L. da Silva. QoS and Simulation Models in Mobile Communication Networks. In *Proceedings of the 3rd ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2000), Boston, MA, USA*, pages 119–122, August 2000.

[Roe02] U. Roedig. *Firewall Architectures for Multimedia Applications*. PhD thesis, Department of Computer Science, Darmstadt University of Technology, Germany, December 2002.

[RT99] E. M. Royer and C.-K. Toh. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks. *IEEE Personal Communications*, 6(2):46–55, April 1999.

[SAZ+04] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana. Internet Indirection Infrastructure. *IEEE/ACM Transactions on Networking (TON)*, 12(2):205–218, April 2004.

[Sch99] F. B. Schneider. *Trust in Cyberspace*. National Academies Press, Washington, 1999. ISBN 0-309-06558-5.

[Sch00] B. Schneier. *Secrets and Lies: Digital Security in a Networked World*. John Wiley & Sons, August 2000. ISBN 0-47125-311-1.

[Sch01] J. Schmitt. *Heterogeneous Network Quality of Service Systems*. Kluwer Academic Publishers, June 2001. ISBN 0-792-37410-X.

[Sch03] J. Schiller. *Mobile Communications, 2nd Edition*. Addison-Wesley, August 2003. ISBN 0-321-12381-6.

[SCS03] Y. Sasson, D. Cavin, and A. Schiper. Probabilistic Broadcast for Flooding in Wireless Mobile Ad hoc Network. In *Proceedings of IEEE Wireless Communications and Networking Conference, WCNC 2003, New Orleans, LA, USA*, pages 1124–1130, March 2003.

[SE01] P. Srisuresh and K. Egevang. Traditional IP Network Address Translator (Traditional NAT). Informational Request for Comments, RFC 3022, January 2001.

[Seg77] A. Segall. The Modeling of Adaptive Routing in Data-Communication Networks. *IEEE Transactions On Communications*, COM-25(1):85–95, January 1977.

[Sei04] C. Seipl. Der Einfluss von Knotenfehlverhalten in Ad Hoc Netzen - Modellierung und Experimentelle Analyse. Student thesis KOM-S-158, Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany, 2004.

[SHHS03] J. Schmitt, P. Hurley, M. Hollick, and R. Steinmetz. Per-flow Guarantees under Class-Based Priority Queueing. In *Proceedings of IEEE Global Telecommunications Conference 2003 (GLOBECOM'03), San Francisco, CA, USA*, volume 7, pages 4169–4174. IEEE, December 2003.

[Shi00] R. Shirey. Internet Security Glossary. Informational Request for Comments, RFC 2828, May 2000.

[SK99] J. Scourias and T. Kunz. Activity-based Mobility Modeling: Realistic Evaluation of Location Management Schemes for Cellular Networks. In *Proceedings of Wireless Communications and Networking Conference, WCNC 1999, New Orleans, LA , USA*, pages 296–300. IEEE, September 1999.

[SM91] W. H. Sanders and J. F. Meyer. A Unified Approach for Specifying Measures of Performance, Dependability, and Performability. In *Dependable Computing for Critical Applications*, pages 215–237. Springer-Verlag Berlin/Heidelberg/New York, March 1991.

[SMSR02] C. A. Santivanez, B. McDonald, I. Stavrakakis, and R. Ramanathan. On the Scalability of Ad Hoc Routing Protocols. In *Proceedings of the 21th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, NY, USA*, volume 3, pages 1688–1697. IEEE, June 2002.

[SN02] S. H. Shah and K. Nahrstedt. Predictive Location-Based QoS Routing in Mobile Ad Hoc Networks. In *Proceedings of IEEE International Conference on Communications (ICC 2000), New York, NY, USA*, pages 1022–1027. IEEE, April 2002.

[SN04a] R. Steinmetz and K. Nahrstedt. *Multimedia Applications*. Springer-Verlag Berlin/Heidelberg/New York, January 2004. ISBN3-540-40849-5.

[SN04b] R. Steinmetz and K. Nahrstedt. *Multimedia Systems*. Springer-Verlag Berlin/Heidelberg/New York, January 2004. ISBN 3-540-40867-3.

[Sob01] J. L. Sobrinho. Algebra and Algorithms for QoS Path Computation and Hop-by-Hop Routing in the Internet. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), Anchorage, AK, USA*, volume 2, pages 727–735. IEEE, April 2001.

[Sol98] J. D. Solomon. *Mobile IP: The Internet Unplugged*. Prentice Hall, Upper Saddle River, NJ, USA, January 1998. ISBN 0-138-56246-6.

[Spe03] T. Spengler. Mehrwegerouting in Ad Hoc Kommunikationsnetzen: Implementierung und Experimentelle Analyse. Diploma thesis KOM-D-189, Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany, 2003.

[SW04] R. Steinmetz and K. Wehrle. Peer-to-Peer Networking and Computing. *Informatik Spektrum*, 27(1):51–54, February 2004.

[SWH95] N. Suri, C. J. Walter, and M. M. Hugue. *Advances in Ultra-Dependable Distributed Systems*. IEEE Computer Society Press, january 1995. ISBN 0-818-662875-].

[Tan03] A. S. Tanenbaum. *Computer Networks, 4th edition*. Prentice Hall PTR, September 2003. ISBN 0-13-066102-3.

[TB02] D. Tang and M. Baker. Analysis of a Metropolitan-Area Wireless Network. *Wireless Networks*, 8(2-3):107–120, March-May 2002.

[TK84] H. Takagi and L. Kleinrock. Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals. *IEEE Transactions on Communications*, 32(3):246–257, March 1984.

[Usk03] S. Uskela. Key concepts for evolution toward beyond 3G networks. *IEEE Wireless Communications Magazine*, 10(1):43–48, February 2003.

[VGLA99] S. Vutukury and J. J. Garcia-Luna-Aceves. A Simple Approximation to Minimum-Delay Routing. In *Proceedings of the Annual Conference of the Special Interest Group on Data Communication (SIGCOMM 1999), Cambridge, MA, USA*, pages 227–238, August 1999.

[VGLA01] S. Vutukury and J. J. Garcia-Luna-Aceves. MDVA: A Distance-Vector Multipath Routing Protocol. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), Anchorage, AK, USA*, pages 557–564. IEEE, April 2001.

[Vil02] C. Villamizar. OSPF Optimized Multipath (OSPF-OMP). Internet Draft, draft-ietf-ospf-omp-03, January 2002. Work in progress.

[Vut01] S. Vutukury. *Multipath Routing Mechanisms for Traffic Engineering and Quality of Service in the Internet*. PhD thesis, University of California Santa Cruz, March 2001.

[WC96] Z. Wang and J. Crowcroft. Quality-of-Service Routing for Supporting Multimedia Applications. *IEEE Journal on Selected Areas in Communications*, 14(7):1228–1234, September 1996.

[Wei91] M. Weiser. The Computer for the Twenty-First Century. *Scientific American*, pages 94–104, September 1991.

[Wei93] M. Weiser. Some Computer Science Issues in Ubiquitous Computing. *Communications of the ACM*, 36(7):75–84, July 1993.

[WLB03] W. Wang, Y. Lu, and B. Bhargava. On Security: Study of Two Distance Vector Routing Protocols for Mobile Ad Hoc Networks. In *Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom'03), Fort Worth, TX, USA*, pages 179–186, March 2003.

[WN04] Princeton University, Cognitive Science Laboratory. WordNet a Lexical Database of the English Language. Available at http://www.cogsci.princeton.edu/<HardSpacewn, October 2004.

[WQDT01] H. Wu, C. Qiao, S. De, and O. Tonguz. Integrated Cellular and Ad Hoc Relaying Systems: iCAR. *IEEE Journal on Selected Areas in Communications*, 19(10):2105–2115, October 2001.

[XN03] Y. Xue and K. Nahrstedt. Fault Tolerant Routing in Mobile Ad Hoc Networks. In *Proceedings of of IEEE Wireless Communications and Networking Conference (WCNC 2003), New Orleans, LA, USA*, volume 4, pages 1174–1179. IEEE, March 2003.

[Yan03] X. Yang. NIRA: A New Internet Routing Architecture. *ACM SIGCOMM Computer Communication Review*, 33(4):301–312, October 2003.

[Yeh02] C.-H. Yeh. ACENET: Architectures and protocols for High Throughput, Low Power and QoS Provisioning in Next-Generation Mobile Communications. In *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002), Lisboa, Portugal*, pages 779–786, September 2002.

[YIY01] H. Yumiba, K. Imai, and M. Yabusaki. IP-Based IMT Network Platform. *IEEE Personal Communications Magazine*, 8(5):18–23, October 2001.

[YLN03] J. Yoon, M. Liu, and B. D. Noble. Random Waypoint Considered Harmful. In *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003), San Francisco, CA, USA*, volume 2, pages 1312–1321. IEEE, April 2003.

[ZCD97] E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A Quantitative Comparison of Graph-based Models for Internet Topology. *IEEE/ACM Transactions on Networking (TON)*, 5(6):770–783, December 1997.

[ZD97] M. M. Zonoozi and P. Dassanayake. User Mobility Modeling and Characterization of Mobility Patterns. *IEEE Journal on Selected Areas in Communications*, 15(7):1239–1252, September 1997.

[ZH99] L. Zhou and Z. J. Haas. Securing Ad hoc Networks. *IEEE Network Magazine*, 13(6):24–30, 1999.

[ZS04] F. A. Zdarsky and J. B. Schmitt. Enhancing Mobile QoS Based on Movement Contracts. In *Proceedings of Twelfth IEEE International Workshop on Quality of Service (IWQoS 2004), Montreal, Canada*, pages 3–9. IEEE, June 2004.

Author's Publications

Journal Articles

M. Hollick, T. Krop, J. Schmitt, H.-P. Huth, and R. Steinmetz. Modeling Mobility and Workload for Wireless Metropolitan Area Networks. *Computer Communications*, 27(8):751–761, Mai 2004.

X. Masip-Bruin, M. Yannuzzi, J. Domingo-Pascual, A. Fonte, M. Curado, E. Monteiro, F. Kuipers, P. Van Mieghem, S. Avallone, G. Ventre, P. Aranda-Gutiérrez, M. Hollick, R. Steinmetz, L. Iannone, K. Salamatian. QoS Routing Open Agenda. *Computer Communications*, to appear.

Patents/Patent Applications

M. Hollick, P. Mogre, T. Krop, J. Schmitt, and H.-P. Huth. Routingverfahren für Nachrichten verschiedener Klassen. German patent application.

Conference Contributions

M. Hollick, I. Martinovic, T. Krop, and I. Rimac. A Survey on Dependable Routing in Sensor Networks, Ad hoc Networks, and Cellular Networks. In *Proceedings of the 30th EUROMICRO Conference (EUROMICRO 2004), Rennes, France*, pages 495–502. IEEE, September 2004.

M. Hollick, J. Schmitt, C. Seipl, and R. Steinmetz. On the Effect of Node Misbehavior in Ad Hoc Networks. In *Proceedings of IEEE International Conference on Communications, ICC'04, Paris, France*, volume 6, pages 3759–3763. IEEE, June 2004.

M. Hollick, J. Schmitt, C. Seipl, and R. Steinmetz. The Ad Hoc On-Demand Distance Vector Protocol: An Analytical Model of the Route Acquisition Process. In *Proceedings of Second International Conference on Wired/Wireless Internet Communications, WWIC 2004, Frankfurt (Oder), Germany*, pages 201–212. Springer-Verlag Berlin/Heidelberg/New York, February 2004.

J. Schmitt, P. Hurley, M. Hollick, and R. Steinmetz. Per-flow Guarantees under Class-Based Priority Queueing. In *Proceedings of IEEE Global Telecommunications Conference 2003 (GLOBECOM'03), San Francisco, CA, USA*, volume 7, pages 4169–4174. IEEE, December 2003.

M. Hollick, T. Krop, J. Schmitt, H.-P. Huth, and R. Steinmetz. Comparative Analysis of Quality of Service Routing in Wireless Metropolitan Area Networks. In *Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks, LCN'03, Bonn/Königswinter, Germany*, pages 470–479. IEEE Computer Society Press, Los Alamitos, October 2003.

M. Hollick, T. Krop, J. Schmitt, H.-P. Huth, and R. Steinmetz. A Hybrid Workload Model for Wireless Metropolitan Area Networks. In *Proceedings of 49th Vehicular Technology Conference, VTC-Fall 2003, Orlando, FL, USA*, volume 3, pages 1989–1993. IEEE, October 2003.

M. Hollick and R. Steinmetz. Security for Ad-Hoc Service Information - Threat Analysis of the Service Location Protocol. In *Enterprise Security*, pages 163–176. IT Verlag für Informationstechnik GmbH, Höhenkirchen, March 2002.

M. Hollick, A. Meissner, L. Wolf, and R. Steinmetz. A Security Model for Multicast Group Integrity Management, Poster Abstract, November 2001. Available at http://www-mice.cs.ucl.ac.uk/ngc2001/poster-abstracts.pdf.

M. Hollick. Security Awareness in Service Discovery for Multimedia Collaboration. In *Proceedings of ACM Multimedia 2001 Workshops, Multimedia and Security: New Challenges*, pages 60–63. ACM, New York, October 2001.

A. Meissner, L. Wolf, M. Hollick, and R. Steinmetz. Security Issues in Group Integrity Management for Multimedia Multicasting. In *Proceedings of Third International Conference on Information, Communications & Signal Processing (ICICS 2001), Singapore*, page n.a. Nanyang Technological University, Singapore, October 2001.

M. Hollick and R. Steinmetz. Security Goals for Service Discovery in Adhoc Environments. In *Proceedings of International Conference on Software, Telecommunications and Computer Networks, IEEE SoftCOM 2001*, pages 1025–1032. FESB-Split, October 2001.

M. Hollick, W. Rosenow, T. Bormuth, and R. Steinmetz. Sicherheitsaspekte bei der Einsatzplanung von Microsoft Windows 2000. In *Innovative Anwendungen in Kommunikationsnetzen, Proceedings der 15. DFN-Arbeitstagung über Kommunikationsnetze, Düsseldorf*, pages 101–111. Gesellschaft für Informatik, Bonn, June 2001.

M. Hollick. Secure Service Centered Networking for Nomadic Usage. In *Proceedings of Communications and Multimedia Security Issues of the New Century, IFIP TC6 / TC11 Joint Working Conference on Communications and Multimedia Security (CMS'01), Darmstadt, Germany*, pages 375–380. Kluwer Academic Publishers, Boston/Dordecht/London, May 2001.

M. Hollick. The Evolution of Mobile IP Towards Security. In *Sicherheit in Netzen und Medienströmen, Tagungsband des GI-Workshops "Sicherheit in Mediendaten"*, pages 38–49. Springer-Verlag, Berlin/Heidelberg/New York, September 2000.

Technical Reports/Miscellaneous

M. Hollick and J. Schmitt. The Ad hoc On-demand Distance Vector Protocol: An Analytical Model of the Route Acquisition Process. Technical Report TR-KOM-2002-11, Darmstadt University of Technology, Multimedia Communications Lab, 2002. Available at ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2002-11.pdf.

M. Hollick, T. Krop, and J. Schmitt. Modeling Mobility for Cellular Networks Based on Statistical Data Collection. Technical Report TR-KOM-2003-02, Darmstadt University of Technology, Multimedia Communications Lab, 2003. Available at ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2003-02.pdf.

M. Hollick, T. Krop, and J. Schmitt. A Hybrid Workload Model for Wireless Metropolitan Area Networks. Technical Report TR-KOM-2003-03, Darmstadt University of Technology, Multimedia Communications Lab, 2003. Available at ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2003-03.pdf.

M. Hollick, T. Krop, and J. Schmitt. Comparative Analysis of Quality of Service Routing in Wireless Metropolitan Area Networks. Technical Report TR-KOM-2003-04, Darmstadt University of Technology, Multimedia Communications Lab, 2003. Available at ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2003-04.pdf.

M. Hollick and J. Schmitt. On the Effect of Node Misbehavior in Ad Hoc Networks. Technical Report TR-KOM-2003-07, Darmstadt University of Technology, Multimedia Communications Lab, 2003. Available at ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2003-07.pdf.

M. Hollick. Sicherheit und Mobilität in IP-basierten Netzwerken, Exposè zur Dissertation. Internal Report, February 2000.

R. Steinmetz and M. Hollick. Schwerpunktentwicklungen der Computertechnik und Telekommunikation, Benutzungsmodelle und Werkzeuge: Zukünftige Entwicklung von Kommunikationstechnologie und Multimedia. Referatemappe des Print und Media Kongresses im Rahmen der Digimedia 1999, August 1999.

Curriculum Vitae

Personal Details

Name	Matthias Hollick.
Date/Place of birth	March 2, 1973/Frankfurt (Main), Germany.
Nationality	German.

Education

Since 04/1999	To be obtained: Doktor-Ingenieur (Dr.-Ing.) in Electrical Engineering and Information Technology.
10/1993 - 12/1998	Studies of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany. *Degree*: Diplom-Ingenieur (Dipl.-Ing.) in Electrical Engineering and Information Technology.
08/1979 - 05/1992	Elementary School, Secondary School and High School. *Degree*: Allgemeine Hochschulreife.

Professional Experience

Since 01/2002	Researcher at the Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany.
04/1999 - 07/2002	Doctoral candidate (part-time) at the Mobile Group, Fraunhofer Institute for Integrated Publication and Information Systems (IPSI), Darmstadt, Germany.
12/1998 - 03/2002	Co-founder and partner of the Nformation GmbH, Mühlheim, Germany. IT-Consultant (part-time).
Since 06/1996	Freelancer as IT-Professional and IT-Trainer. Since 01/1999 Microsoft Certified Systems Engineer (MCSE). Since 02/1997 Microsoft Certified Professional (MCP). From 12/1999 - 12/2003 Microsoft Certified Trainer (MCT).

Supervision of Diploma-, Master-, and Student-Theses

Mr. P. S. Mogre, *Design of a Novel Radio Access Network Architecture and Analysis of QoS Routing Mechanisms for Resource Management*, Master-thesis in collaboration with IIT-Guwahati, India, KOM-D-209, 05/2004.

Mr. M. Peter, *Mehrwegerouting in Ad Hoc Kommunikationsnetzen: Experimentelle Analyse*, Diploma-thesis KOM-D-189, 04/2004.

Mr. T. Spengler, *Mehrwegerouting in Ad Hoc Kommunikationsnetzen: Implementierung und Experimentelle Analyse*, Diploma-thesis KOM-D-200, 07/2003.

Mr. C. Seipl, *Der Einfluss von Knotenfehlverhalten in Ad Hoc Netzen: Modellierung und Experimentelle Analyse*, Student-thesis KOM-S-158, 03/2003.

Mr. T. Krop, *Experimental Analysis of QoS-Routing Mechanisms for Load Balancing in Radio Access Networks*, Diploma-thesis KOM-D-182, 12/2002.

Mr. D. Theiler, *Klassifizierung und Vergleich von Routingverfahren für Mobile Adhoc Netzwerke hinsichtlich Sicherheit und Dienstgüte*, Diploma-thesis KOM-D-180, 10/2002.

Mr. T. Schmidt, *Last-Balancierung in mobilen Zugangsnetzen*, Diploma-thesis KOM-D-172, 10/2002.

Mr. R. Hartmann, *Experimental Analysis of Gossip-based Strategies to Optimize the Reliability of Routing Mechanisms within Mobile Ad Hoc Networks*, Diploma-thesis KOM-D-181, 09/2002.

Ms. S. Wüst, *System and Protocol Analysis of Future Authentication, Authorization and Accounting Infrastructures with Respect to Mobile Users and End Systems*, Student-thesis KOM-S-093, 11/2000.

Mr. A. Lang, *Netzwerksicherheit von Mobile IP: Analyse von Verfügbarkeitsaspekten*, Diploma-thesis in collaboration with Hochschule Anhalt (FH), IN-96, 11/2000.

Mr. H. Kirchner, *Systeme zur Diensterkennung und zur Dienstverwaltung im mobilen Umfeld*, Thesis to fulfil the prerequisites for entering a PHD program, 09/2000.

Appendix A: Supplementary Results to Chapter 4

In the following we present additional plots for the simulation study described in Chapter 4. There, a detailed interpretation of results has already been given. Hence, we here only present the figures without additional explanation. The factors that distinguish the shown test from the tests presented in Chapter 4 are printed in italics.

Supplementary Results for the RAN-I Topology

- Figure 114 shows the *average delay* for Test (4.1) to Test (4.9).

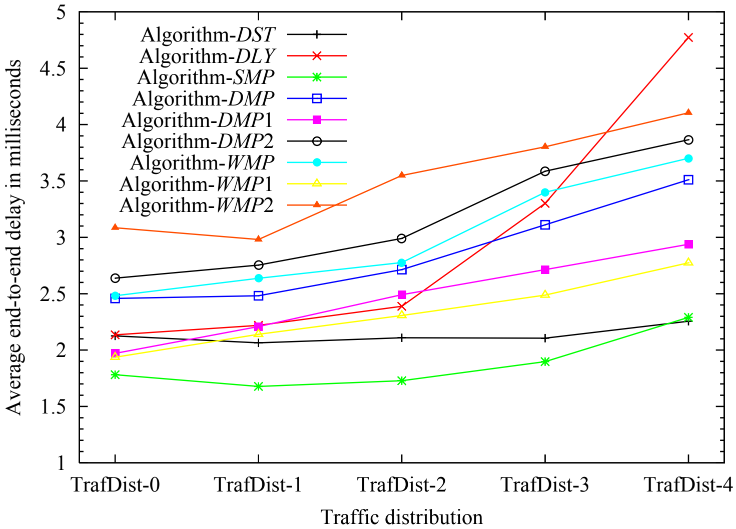


Figure 114: Average delay for all algorithms over 24 h. All traffic distributions are shown.

Supplementary Results for the RAN-I Topology

- Figure 115 shows the *average loss* for Test (4.1) to Test (4.9).

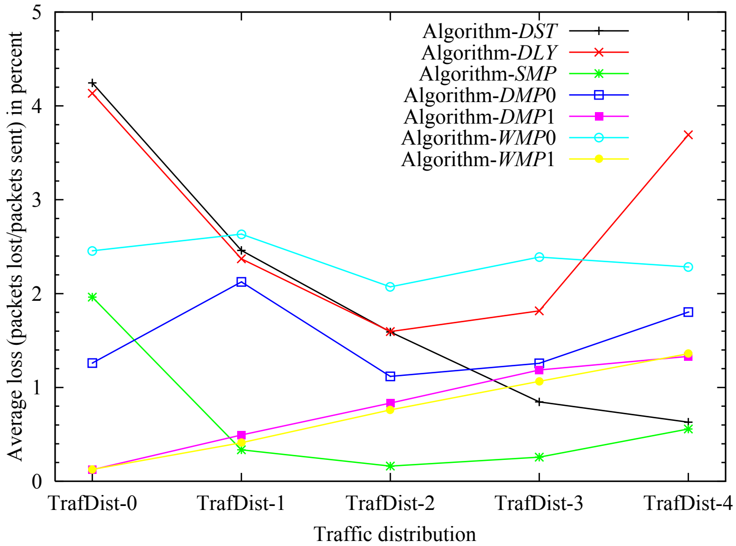


Figure 115: Average loss for all algorithms over 24 h. All traffic distributions are shown.

Appendix B: Supplementary Results to Chapter 5

In the following, we present additional plots for the simulation study described in Chapter 5. There, a detailed interpretation of results has already been given. Hence, we here only present the figures without additional explanation. The factors that distinguish the shown test from the tests presented in Chapter 5 are printed in italics.

Supplementary Results for the Cairn Topology

- Figure 116 shows the average number of *sent packets* for Test (5.1) to Test (5.5).

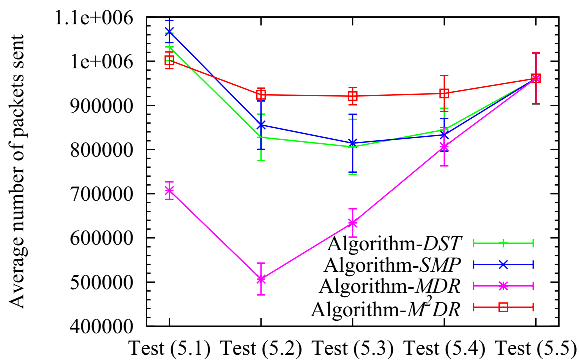


Figure 116: Average number of sent packets for Test (5.1) to Test (5.5). We show all algorithms.

Supplementary Results for the Vutukury Topology

- Figure 117 shows the per-class average end-to-end delay for *Test* (5.7) and *Test* (5.8).

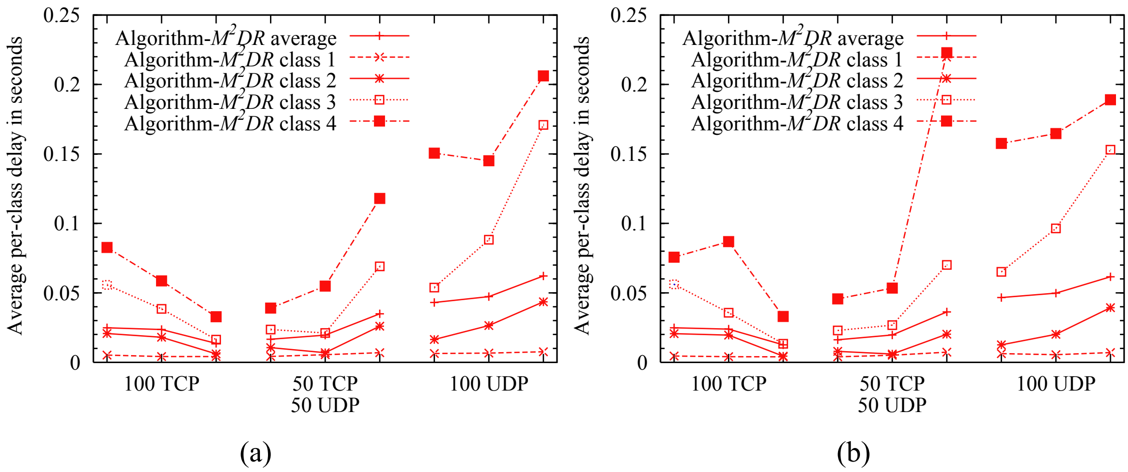


Figure 117: Average per-class end-to-end delay for (a) Test (5.7) and (b) Test (5.8).

Supplementary Results for the RAN-II Topology

- Figure 118 shows the per-class average end-to-end delay for Test (5.9) to Test (5.13) for 100% inelastic traffic and *symmetric class distribution*.

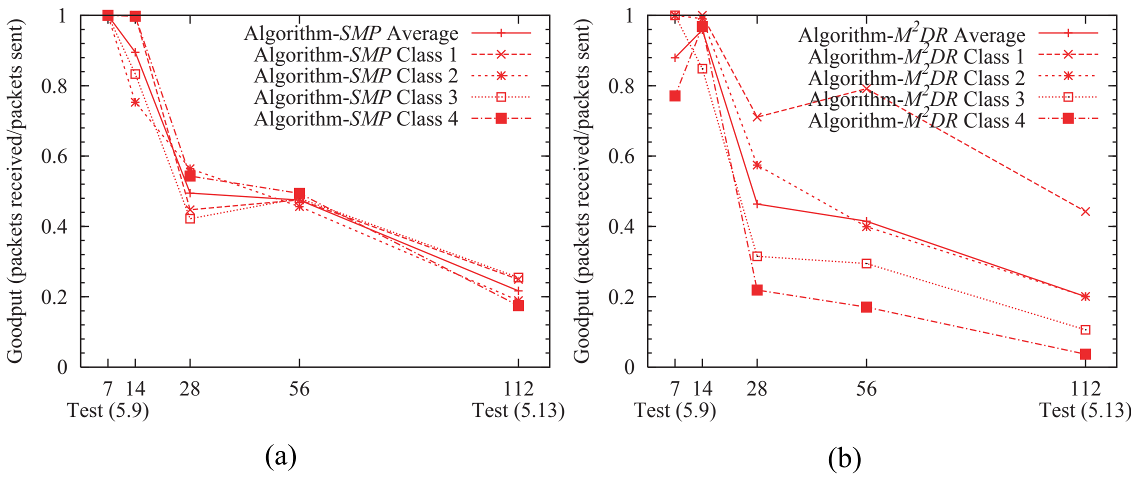


Figure 118: Average per-class end-to-end delay for Tests (5.9) to (5.13). We show all traffic classes for (a) Algorithm-*SMP* and (b) Algorithm-M^2DR for symmetric class distribution and 100% inelastic traffic.

Appendix C: Supplementary Results to Chapter 6

Derivation of r_1

Let A_n denote the area a node covers with its radio:

$$A_n = \pi r^2 \tag{62}$$

Given a random distribution of all nodes located within this area, we calculate the area which hosts 50% of the nodes:

$$\frac{A_n}{2} = \frac{\pi r^2}{2} \tag{63}$$

The radius covering this area is:

$$r_1 = \sqrt{\frac{r^2}{2}} = \frac{r}{\sqrt{2}} \tag{64}$$

This radius r_1 describes the median distance between two neighboring nodes. Since a random node is expected to propagate a broadcast such as a routing request we expect the routing progress per hop to be in average r_1.

Please note that in reality the average progress per hop is coupled to the node density of the target network as well as to the routing ability of the routing protocol to find the shortest path.

Appendix D: Supplementary Results to Chapter 7

Mean Sphere of Influence of a Black Hole

Assume a network which consists of exactly one destination node and exactly one *black hole* serving as a fake destination node, the positions of these nodes being identically and independently distributed. Each possible constellation is complemented by exactly one opposite constellation of *black hole* and destination node. Likewise, the sphere of influence from the real destination node and the fake destination node are interchangeable. We denote these areas A_1 and A_2. We calculate the mean area which is dominated from one particular node as $\bar{A}$:

$$\bar{A} = \frac{A_1 + A_2}{2} = \frac{A}{2} \text{ with } A_1 + A_2 = A. \tag{65}$$

The generalization to cover one destination node and m *malicious nodes* gives $(m+1)!$ constellations. Each node can appear at $(m+1)$ positions, which it occupies $m!$ times. The mean area $\bar{A}$ is given by:

$$\bar{A} = \sum_{j=1}^{m} m! A_j \cdot \frac{1}{(m+1)!} = m! \sum_{j=1}^{m} A_j \cdot \frac{1}{(m+1)!} = \sum_{j=1}^{m} A_j \cdot \frac{1}{m+1} = \frac{A}{m+1}. \tag{66}$$

Especially the concentration of *black holes* in certain areas may lead to other results. The areas are moreover not necessarily of the same shape. Considering these boundary conditions, the formula generally holds.

h_{max} in the Sphere of Influence of a Black Hole

As shown above, the generalized sphere of influence of a *black hole* is $A/(m+1)$. An area-equivalent square has a side length of $\sqrt{A/(m+1)}$. We obtain an estimate for the expected maximum distance in hops between two nodes h_{max} for a network with black holes. The maximum length inside the square is the diagonal line with length $\sqrt{(2A)/(m+1)}$. The expected maximum hopcount for the square area is then:

$$h^{\square}_{max} = \frac{1}{r_1}\sqrt{\frac{2A}{m+1}} = h_{max}\sqrt{\frac{1}{m+1}}. \quad (67)$$

Under the assumption that the covered area is represented better by a circular area than a rectangle, we can transform the result into an area-equivalent circle. The radius being $\sqrt{A/(\pi(m+1))}$. The maximum distance equals the diameter and the hopcount is:

$$h^{\circ}_{max} = \frac{1}{r_1}\sqrt{\frac{4A}{\pi(m+1)}} = h_{max}\sqrt{\frac{2}{\pi(m+1)}}. \quad (68)$$

These results now can easily be compared to the error free case of h_{max}:

$$h_{max} = \frac{d_{max}}{r_1} = \frac{1}{r_1}\sqrt{2A}. \quad (69)$$

Experimental Validation of the results obtained for h_{max}

We have performed tests to validate the *malicious node* model for pure AODV and gossip-*p* enhanced AODV. The experimental setup is shown in Table 16. We simulate 500 nodes. In order to quantify the impact of black holes on the data plane, we randomly selected 25 source-destination pairs and injected 25 continuous CBR flows with a rate of 4 packets/sec of size 512 bytes, each. We also randomly selected 2% of the nodes to be *black holes* (10 from 500 nodes) for AODV and gossip-*p* enhanced AODV respectively. The observed degradation of the routing performance is noteworthy and can be seen in Figure 119. Comparing the numerical result with our prognosis we obtain

$$h_{max} = \frac{\theta_{AODV}}{r_1}\sqrt{2A} = \frac{1.064}{176.679}\sqrt{2(3742.92m)^2} = 31.8 \approx 32,$$

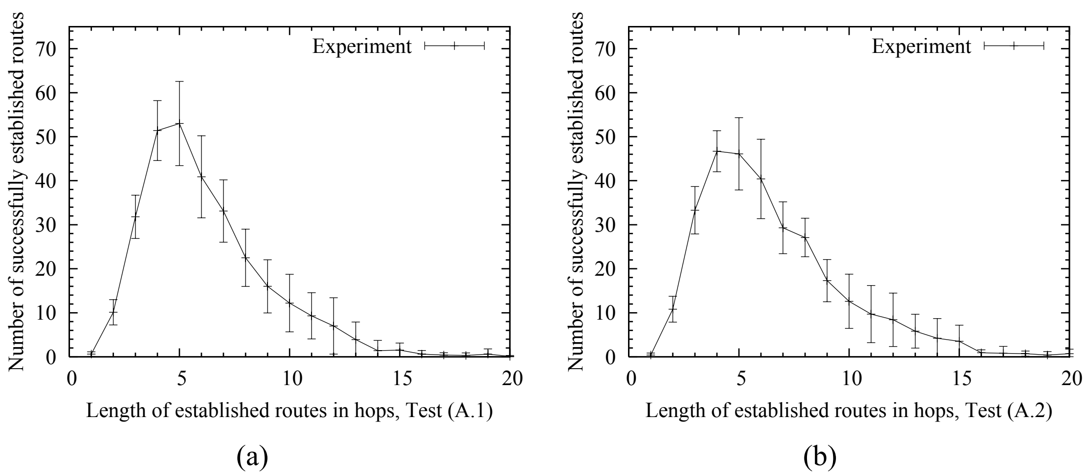


Figure 119: Route-length distribution for malicious nodes. We show (a) AODV in Test (A.1) and (b) gossip-*p* enhanced AODV for Test (A.2) for 2% black holes.

Table 16: Experimental Parameter Set for Misbehavior Model Verification.

Test	Test (A.1), AODV	Test (A.2), Gossip-*p* enhanced AODV
Gossip probability	1	0,7
Node misbehavior	2% of nodes are black holes, $m = 10$.	2% of nodes are black holes, $m = 10$.

The common parameters for the test-series A.* are as follows.
Simulation: 10 replications per experiment, 500 s simulation time.
Nodes density/placement: n =500 nodes are randomly placed in a simulation area of $(3742.92\text{m})^2$ The node density is constantly M=7. The x and y coordinates are independently and identically uniformly distributed.
Mobility: Random waypoint mobility with 0-2m/s.
Traffic: The source/destination pairs are randomly selected and 50 nodes participate in the communication (25 sources). We use constant bitrate traffic with UDP as transport protocol. The traffic rate is chosen to be 4 packets/s with 512 bytes each per flow.
Physical layer parameters: Propagation model = free space, transmission power = 7dBm, transmission range r = 249.862m, r_0 = 94.438m, r_1 = 176.679m.
Link layer parameters: MAC 802.11b with Distributed Coordination Function (DCF), max. transmission rate = 11 Mbit/s.
AODV parameters: Local repair = deactivated, reply by intermediat = deactivated,
ERS = activated, hello messages = deactivated, net diameter = 35.

$$h^{\square}{}_{max} = h_{max}\sqrt{\frac{1}{m+1}} = 31.8\sqrt{\frac{1}{11}} \approx 9.6,$$

$$h^{\circ}{}_{max} = h_{max}\sqrt{\frac{2}{\pi(m+1)}} = 31.8\sqrt{\frac{2}{11\pi}} \approx 7.67\,.$$

Our results confirm the prediction that the drop would occur around $h_{max}/2$, which gives $h = 4$ or $h = 5$ as observed in the simulation. The calculation of the loss the *black hole* introduces is performed using:

$$d = h(d)r_1 \text{ and } \frac{h^{\circ}{}_{max}}{2} \approx 3.8 \Rightarrow d = 0.181\,.$$

Integration of the probability measure function gives:

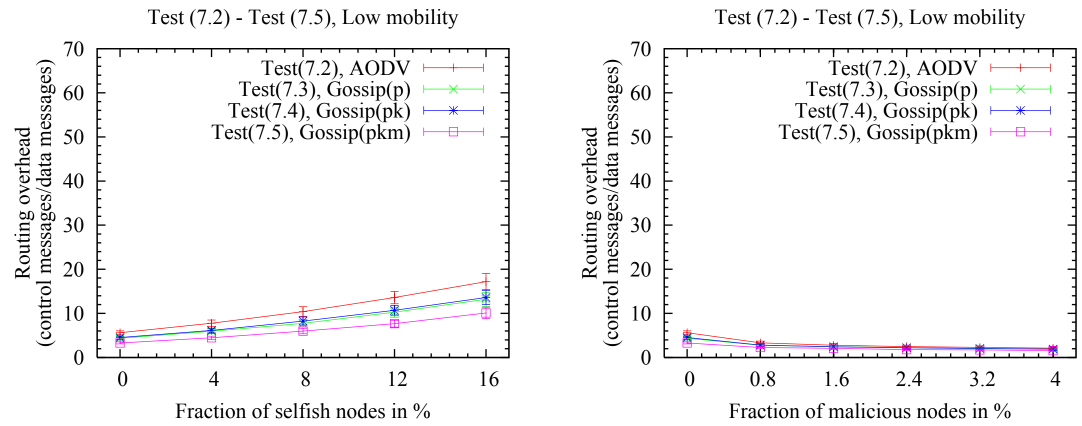
$$P_d(\xi) = \int_0^{0.181} p_d(\xi)d\xi = \int_0^{0.181} (1-q_{ers})^{\frac{d\theta_{ers}}{r_1}} 2\xi(\xi^2 - 4\xi + \pi)d\xi \approx 0.08217 = 8.22\%.$$

This is an accurate prediction for our experimental results which are: $P_d(d) = 0.0811$ with standard deviation $\sigma = 0.04462$.

Additional Simulation Results

In the following, we present additional plots for the simulation study described in Chapter 7. There, a detailed interpretation of results has already been given. Hence, we here only present the figures without additional explanation. The factors that distinguish the shown test from the tests presented in Chapter 7 are printed in italics.

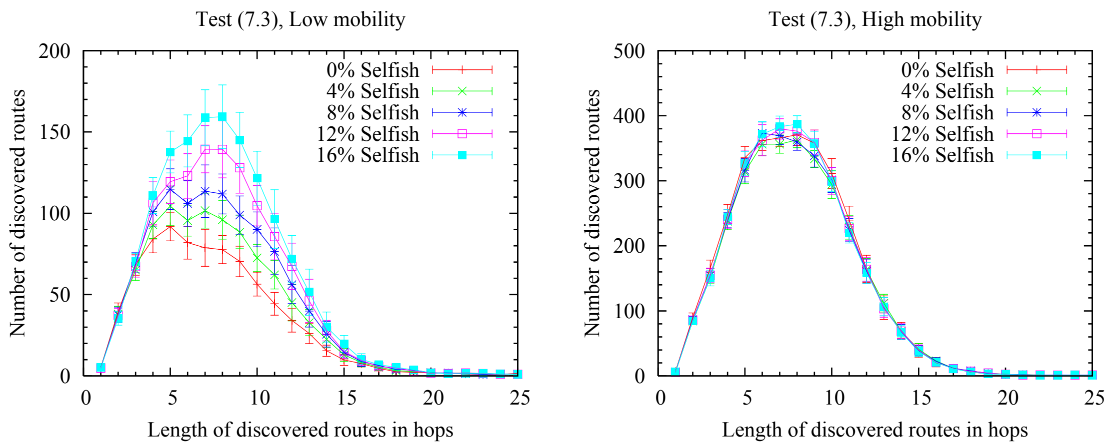
- Figure 120 shows the routing overhead for *low node mobility* for all tests.



(a) (b)

Figure 120: Routing overhead for low node mobility, Test (7.2) to Test (7.5). Our measurements show (a) selfish node and (b) malicious nodes. The error bars show the 95% confidence intervals.

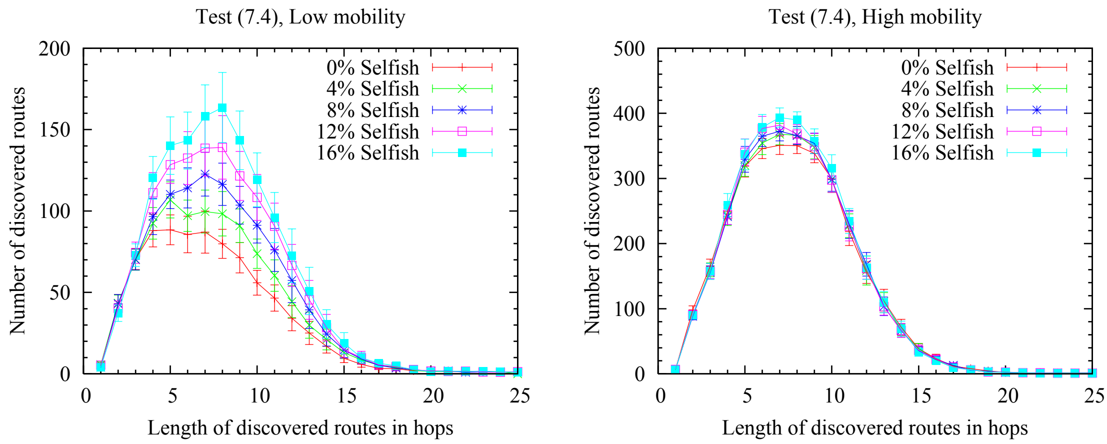
- Figure 121 shows the route-length distribution for selfish nodes for *gossip-p enhanced AODV* (Test (7.3)).



(a) (b)

Figure 121: Route-length dist. of gossip-*p* enhanced AODV + selfish nodes, Test (7.3). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals. Please note the different y-axis.

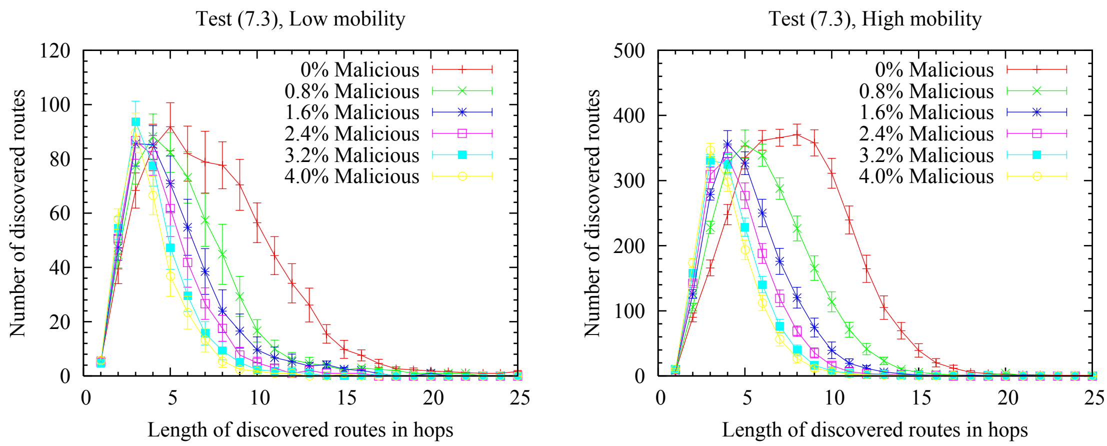
- Figure 122 shows the route-length distribution for selfish nodes for *gossip-pk enhanced AODV* (Test (7.4)).



(a) (b)

Figure 122: Route-length dist. of gossip-*pk* enhanced AODV + selfish nodes, Test (7.4). Our measurements show (a) low node mobility and (b) high node mobility. The error bars show the 95% confidence intervals. Please note the different y-axis.

- Figure 123 shows the route-length distribution for malicious nodes for *gossip-p enhanced AODV* (Test (7.3)).



(a) (b)

Figure 123: Route-length dist. of gossip-*p* enhanced AODV + malicious nodes, Test (7.4). Our measurements show (a) low node mobility and (b) high node mobility with 95% confidence intervals. Please note the different y-axis.

Notation of Formulae

For the convenience of the reader we keep the standard notation of the respective mathematical areas as far as possible. However, since we explore different fields of mathematics, such as graph theory and probability theory, this leads to a few overlapping variables. Throughout the individual chapters, there is no overlap in notation, though. In the following our notation of formulae is described.

Notation of Formulae in Chapter 3

L	set of locations
B	set of user roles (behavior)
Z	set of zones
C	set of cells
M	set of traffic classes
A	size of the area of investigation in square meters
p, r, k, $f_b^{\ u}(t)$, $f_b^{\ a}(z)$, $f_b^{\ i}$	modeling parameters and variables
$U(t)$	number of users at time t
$\tilde{U}_b$	sustained number of actively communicating users of behavior b
$\bar{U}_b$	overall user activity in user hours for behavior b for the interval $[t_1, t_2]$
$W_{m,c}(t)$	workload for cell c and traffic class m at time t

Notation of Formulae in Chapter 4 and Chapter 5

$G = G(V, E)$	graph G with a set of vertices V, $\lvert V\rvert = p$ and a set of edges E, $\lvert E\rvert = q$
i, j, k, u, v	vertices of G
$\{i, k\}$	edge of G that joins i and k, $l = (i, k) \in E$ directed edge from i to k
N^i	set of neighbors of node i
r_j^i	rate of traffic entering the network at i for destination j in bit/s
t_j^i	total traffic at i destined for j in bit/s
ϕ_{jk}^i	routing parameter, that is fraction of t_j^i form i to j over (i, k)
f_{ik}	expected traffic flow on (i, k) in bit/s
C_{ik}	capacity of (i, k) in bit/s
D_{ik}	delay of (i, k) in bit/s
$D'_{ik} = \partial D_{ik}/\partial f_{ik}$	marginal delay from i to j
D_j^i	marginal distance from i to j
D^T	total network wide delay
S_j^i	successor set of i for destination j
l_k^i	cost of link (i, k)
$SG_j(t)$	routing graph for a destination j
λ_{ij}, ψ, A_j^i, B_j^i	modeling parameters and variables
FD_j^i	feasible distance from i to destination j
LSM	link-state map
M	set of traffic classes, $\lvert M\rvert = g$

$MRT[m]$	multipath routing table for class m
t_l	time interval for calculation of the successor set
t_s	time interval for calculation of the routing parameter ϕ
α, β, γ	modeling parameters and variables of Algorithm-M^2DR
$C_{perceived}$	perceived link capacity

Notation of Formulae in Chapter 6 and Chapter 7

A	size of the investigated area. for our model is A dimensionless, because we operate on a unit square
N	set of nodes, $\|N\| = n$
x, y	coordinates of a node
r	transmission range of a node
r_1	average routing progress per hop
d	Euclidean distance between two nodes
M	average degree of the *network*
$h(d)$	length of a route in hops
h_{ers}	number of hops for expanding ring search
$\theta, \theta_f, \theta_{ers}$	elongation of routes, for flooding, for expanding ring search
σ	reduction of route query due to reply by intermediate
$p_d(\xi)$	probability density function (pdf) for d
$P_d(\xi)$	cumulative probability distribution (cdf) for d
s_x, s_y	side lengths of an arbitrary rectangle
α, β, ξ	random variables

q	loss probability, $1 - q$ denotes the corresponding success probability
$pm_d^{success}(\xi)$	probability measure for *successfully discovered* routes out of $p_d(\xi)$
f_{in}, f_{sn}, f_{mn}	fraction of inactive, selfish, and malicious nodes
h_{max}	maximum route-length in hops for cooperating nodes
$h^{\square}{}_{max}$, $h^{\circ}{}_{max}$	maximum route-length in hops for a square area, for a circular area

List of Abbreviations

3G	Third Generation (here in context of cellular networks)
3GPP	Third Generation Partnership Project [3GPP04]
ACK	Acknowledgement
Algorithm-*DLY*	Algorithm-minimum-DeLaY routing
Algorithm-*DMP*	Algorithm-Dynamic MultiPath routing
Algorithm-*DST*	Algorithm-minimum DiSTance routing
Algorithm-M^2DR	Algorithm-Multiclass Minimum-Delay Routing
Algorithm-*MDR*	Algorithm-Minimum-Delay Routing
Algorithm-*SMP*	Algorithm-Static MultiPath routing
Algorithm-*WMP*	Algorithm-dynamic Weighted MultiPath routing
AODV	Ad hoc On-demand Distance Vector routing [PBRD03]
B3G	Beyond Third Generation (here in context of cellular networks)
BSS	Base Station Subsystem (here in context of cellular networks)
CBR	Constant Bit Rate
CN	Core Network
CONFIDANT	Cooperation Of Nodes, Fairness In Dynamic Ad-hoc NeTworks[BB02b].
CoV	Coefficient of Variance
CR	Core Router
DiffServ	Differentiated Services [BBC+98]
DSR	Dynamic Source Routing [JMH04]

E2E	End to End
EGW	Edge Gateway
e/i-ratio	external/internal-ration
ERS	Expanding Ring Search
GGSN	Gateway General Packet Radio Service Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
iCAR	integrated Cellular and Ad hoc Relaying system [WQDT01]
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IFIP	International Federation for Information Processing
IMT-2000	International Mobile Telecommunications 2000
IntServ	Integrated Services [BCS94]
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISDN	Integrated Service Digital Networks
ITU	International Telecommunication Union
LAN	Local Area Network
LFI	Loop Free Invariant
LSM	Link State Map
MAC	Medium Access Control
MANET	Mobile Ad hoc Network
MDVA	Distance-Vector Multipath Routing Algorithm [VGLA01]
MPLS	Multi Protocol Label Switching
MN	Mobile Node
MobQoS	Mobility aware Quality of Service Routing [HSHS04]
MRT	Multipath Routing Table
NIRA	New Internet Routing Architecture [Yan03]
ns-2	(Berkeley) Network Simulator-2

OLSR	Optimized Link State Routing [CJ03]
OSPF	Open Shortest Path First [Moy98]
OSPF-ECMP	Open Shortest Path First - Equal Cost MultiPath [Moy98]
OSPF-OMP	Open Shortest Path First - Optimized MultiPath [Vil02]
P2P	Peer to Peer
PSTN	Public Switched Telephone Network
QoS	Quality of Service
Q-OSPF	Quality of Service aware Open Shortest Path First [AWK+99]
RAN	Radio Access Network
RAP	Radio Access Point
RAR	Radio Access Router
RAS	Radio Access Server
RERR	Route ERRor message
RFC	Request For Comments
RREP	Route REPly message
RREQ	Route REQuest message
RNC	Radio Network Controller
RNS	Radio Network Subsystem
SEAD	Secure Efficient Distance Vector Routing in Mobile Wireless Ad Hoc Networks[HJP02]
SGSN	Serving General Packet Radio Service Support Node
SNF	Split Naming/Forwarding Network Architecture [JFA03],
TCP	Transmission Control Protocol
TRANSIMS	TRansportation ANalysis SIMulation System [LAN04]
TRR	Terminode Routing [Bla02]
TTL	Time To Live
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestial Radio Access Network

VoIP	Voice over IP
VWL	Virtual Wireless Link
WAP	Wireless Application Protocol
WLAN	Wireless Local Area Network
WR	Wireless Relay
WRR	Wireless Relay Router
WWW	World Wide Web

List of Figures

List of Tables