Multicast for Ubiquitous Streaming of Multimedia Content to Mobile Terminals

Network Architecture and Protocols

Doctoral Dissertation by

Mariann Hauge

Submitted to the Faculty of Mathematics and Natural Sciences at the University of Oslo in partial fulfilment of the requirements for the degree Doctor Scientiarum (Dr. Scient.) in Computer Science

March 2007

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
Nr. 624

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2007.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

ABSTRACT

The Universal Mobile Telecommunication Services (UMTS) network was envisioned to carry a wide range of new services; however, the first UMTS release was not designed to efficiently support multimedia content. In this thesis we analyse several mechanisms, and suggest architectural changes to improve UMTS's capacity for a subset of the multimedia services; high-bandwidth group services.

In our initial work we have suggested how IP multicast protocols can be used in the UMTS network to reduce the required network capacity for group services. This proposal was one of many suggestions for the evolving Multimedia Broadcast/Multicast Service (MBMS) architecture for UMTS.

The next technique we have suggested and analysed is a new wireless channel type named the "sticky-channel"; this channel is intended for sparsely populated multicast groups. The sticky-channel is able to stick to mobile multicast members in the boarder area of neighbouring radio cells, thus some base stations does not need to broadcast the multicast data. Consequently, the total number of broadcast channels needed to cover a given area is reduced. There is a marginal reduction of required resources with this technique.

In the main part of our work we have studied heterogeneous multihop wireless access for multicast traffic in the UMTS network. In a heterogeneous wireless access network, the wireless resources needed to distribute high-bandwidth group services, can be shared among co-operating network technologies. Mobile terminals with a UMTS interface and an IEEE 802.11 interface are readily available, consequently a heterogeneous network with UMTS and 802.11 links will be easy to deploy. We have described a heterogeneous architecture based on those wireless technologies. In this architecture, the range of a UMTS radio channel is reduced, and local IEEE 802.11-based Mobile Ad Hoc Networks (MANETs) forward the data to users located outside the coverage of the reduced UMTS channel. The wireless resources required to transmit a data packet are proportional to (at least) the square of the distance the packet must travel, thus a reduction in the channel range releases a significant amount of UMTS radio resources. Detailed simulation results showed acceptable service quality when the UMTS broadcast channel range is more than halved.

Finally we have studied whether Forward Error Correction (FEC) at the packet-level on multicast flows could improve the performance of the heterogeneous wireless access network. There is a marginal improvement. Most of the protection brought by the FEC code has been used to repair the increased packet-loss introduced by the FEC overhead.

PREFACE

This dissertation is submitted to the Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, in partial fulfilment for the degree of Dr. Scient. My work for this dissertation started in August 1999. Part of the work has been done as a full-time student, and part of the work as a part time student. The timeframe spanning the dissertation project also includes a pregnancy followed by 12 months maternity leave, and 4 months leave to finish a project for my employer (at the time) and part funder; Ericsson AS.

The work was carried out at University Graduate Center at Kjeller (UniK) under supervision of Professor Øivind Kure at the Norwegian University of Science and Technology, Trondheim and Professor Pål Spilling at UniK. Due to Pål Spilling's retirement from his position as a professor, Professor Knut Øvsthus at Bergen University College, and Associate Professor II Paal Engelstad at the Department of Informatics, University of Oslo, have also been assigned as my supervisors in the final year of the thesis work.

The research has been part of the Future Communication Systems (FUCS) program, partly funded by the Research Council of Norway. Ericsson AS funded the remaining half of my project. The work has also been supported by UniK and the Norwegian Defence Research Establishment (FFI).

As an employee of Ericsson, I wanted to focus my research on wireless communication for mobile telecommunication networks. I chose the Universal Mobile Telecommunication Services (UMTS) network as the framework for my research, and decided to narrow my focus on UMTS's low capacity for high-bandwidth group services. The remainder of this dissertation presents the result of this work; analysis of several mechanisms intended to improve UMTS's performance for these types of service.

ACKNOWLEDGMENTS

First and foremost, I want to thank my technical advisor Professor Øivind Kure for his excellent supervision of my research work. He has been a valuable and motivating discussion partner, for all my work. Øivind was always available when I needed his comments. Thank you!

I am grateful to my (former) advisor Professor Pål Spilling and my former employer Ericsson AS for giving me the opportunity to pursue my PhD through the Future Communication Systems (FUCS) program at the University Graduate Center at Kjeller (UniK).

I would like to thank Dr. Frank Y. Li at UniK for his support, feedback and valuable suggestions for my research. I have also gained much insight from discussions with former and present PhD students at UniK, especially Andreas Hafslund, Eli Winjum and Lars Landmark.

The efficient and friendly administrative support I have been given by the staff at UniK, have been appreciated.

This dissertation is dedicated to my family. First and foremost my dedication goes to my mother Klara, who did not live to see my PhD work finished, and to my father Tørres. They taught me to appreciate challenges and to stubbornly keep working with a problem until it can be solved. Those two values were the most important qualities I needed to see my research work through.

I appreciate the support from my brother Trond and my niece Turid who have both proofread and given comments to part of my work. Trond has also given me frequent Adobe© support for high quality pdf documents. Turid has provided LaTeX support for the cases when I was forced to use LaTeX for my paper submissions.

This dissertation is also dedicated to my nearest family; Arild and our daughter Helene (she was born halfway through my PhD work). It goes without saying that Arild's support has been essential for my work. I look forward to having more spare time to spend at home with them.

Oslo, March 25, 2007

Mariann Hauge

LIST OF PUBLICATIONS

The author of this thesis is the primary author of papers A through F and the research report G (appended as Part II of the thesis). Papers A through F are co-authored with the external, technical supervisor. He has been a valuable discussion partner for the work presented and has also provided structural suggestions to how the content could be presented in a clear and precise manner. The multicast protocols proposed in paper D has benefited from technical discussion with a larger set of co-authors. The research report G provides detailed description of a slightly modified version of one of the multicast protocols proposed in paper D.

The author of this thesis has contributed to papers H and I as a discussion partner. For paper H the author has also provided technical expertise of the IEEE 802.11 standard to form the basis for applying the proposed empirical approach.

All papers (except the research report) have been published at peer-reviewed conferences and workshops.

PAPER A: M. Hauge and Ø. Kure, “Multicast in 3G Networks: Employment of Existing IP Multicast Protocols in UMTS,” *In proceedings of the 5th ACM International Workshop on Wireless Mobile Multimedia 2002 (WoWMoM'02)*, Atlanta, USA, 2002, pp. 96-103.

PAPER B: M. Hauge and Ø. Kure, “Sticky Point-to-multipoint Channel for Multicast in UMTS,” *In proceedings of the 8th International Conference on Cellular and Intelligent Communications 2003 (CIC'03)*, Seoul, South Korea, 2003.

PAPER C: M. Hauge and Ø. Kure, “Multicast Service Availability in a Hybrid 3G-cellular and Ad Hoc Network,” *In proceedings of the 1st International Workshop on Wireless Ad-hoc Networks 2004 (IWWAN'04)*, Oulu, Finland, 2004.

PAPER D: M. Hauge, A. Hafslund, F.Y. Li and Ø. Kure, “Multicast-service Distribution on a Cellular Network Assisted by Local Ad Hoc Networks,” *In proceedings of the 3rd Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net'04)*, Bodrum, Turkey, 2004, pp. 68 - 80.

PAPER E: M. Hauge and Ø. Kure, "A Heterogeneous Cellular and Ad Hoc Network Architecture for Multipoint Streaming: A detailed performance analysis," *In proceedings of the 3rd International Conference on Mobile Computing and Ubiquitous Networking (ICMU'06)*, London, UK, 2006. (A version of this paper has been accepted for publication in a special issue of the Information Processing Society of Japan (IPSJ) Journal, featuring selected papers from ICMU 2006.)

PAPER F: M. Hauge and Ø. Kure, "Evaluation of Packet-level FEC with Multicast Streaming for a Heterogeneous 3G-Cellular and Ad Hoc Network," *In proceedings of the European Symposium on Mobile Media Delivery 2006 (EuMob'06)*, Alghero, Sardinia, Italy, 2006.

REPORT G: M. Hauge, "Multicast in a Hybrid Cellular and Ad Hoc Network: Specification of an Ad Hoc Routing Protocol with Cellular Assistance," Department of Informatics, University of Oslo, Norway, *Research Report no. 334,* March 2006, http://www.duo.uio.no.

Related papers:

PAPER H: F.Y. Li, M. Hauge, A. Hafslund, Ø. Kure and P. Spilling, "Estimating Residual Bandwidth in 802.11-based Ad Hoc Networks: An Empirical Approach," *In proceedings of the 7th International Symposium on Wireless Personal Multimedia Communications 2004 (WPMC'04)*, Abano Terme, Italy, 2004, pp. 471-476.

PAPER I: F.Y. Li, A. Hafslund, M. Hauge, P. Engelstad, Ø. Kure and P. Spilling, "Dilemma of Using High Datarate in IEEE 802.11b Based Multihop Ad Hoc Networks," *In proceedings of the The 3rd IASTED International Conference on Communications, Internet, and Information Technology 2004 (CIIT'04)*, St. Thomas, US Virgin Islands, 2004, pp. 231-237.

LIST OF FIGURES

List of Terms and Acronyms

1xEV-DO	1x EVolution-Data Optimized
3G	3^{rd} generation of mobile networks
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
4G	4^{th} generation of mobile networks
AAA	Authentication, Authorization, and Accounting
AODV	Ad-hoc On-demand Distance Vector
AuC	Authentication Centre
APN	Access Point Name
ARS	Ad-hoc Relay Station
AS	Autonomous System
ASM	Any Source Multicast
BCMCS	Broadcast/Multicast Services (for CMDA2000)
BM-SC	Broadcast Multicast - Service Center
CBT	Core Based Tree
CDMA2000	CDMA2000 is a family of third-generation (3G) mobile telecommunications standards mainly used on the American continent. CDMA2000 is standardized by 3GPP2.
CeNALAN	Cellular Network Assisted by Local Ad Hoc Networks
CGF	Charging Gateway Function
CN	Core Network
DVB	Digital Video Broadcasting
DVB-H	DVB for handheld terminals
EIR	Equipment Identity Register
Erasure code	An encoding of a block of data packets that allow the receiver to reconstruct a certain number of erasures (lost packets).
FACH	Forward Access Channel
FEC	Forward Error Correction
Flooding	A mechanism to distribute a data packet to all nodes in a network. All nodes in the network broadcasts the incoming packet on all of its interfaces, once for each packet.
GGSN	Gateway GPRS Support Node

GPRS	General Packet Radio Service
HDR	High Data Rate
HLR	Home Location Register
HSDPA	High Speed Downlink Packet Access
IEEE	Institute of Electrical and Electronics Engineers
IEEE 802.11	Wireless LAN standard developed by the IEEE
IEEE 802.16	Wireless MAN standard developed by the IEEE
IETF	Internet Engineering Task Force. A standardisation body for IP networking.
IGMP	Internet Group Management Protocol
IP	The Internet Protocol
IST	Information Society Technologies
J-Sim	Component based network simulator, coded in Java©
LAN	Local Area Network
MAC	Medium Access Control
MAN	Metropolitan Area Network
MANET	Mobile Ad hoc NETwork
MBC	MBMS Bearer Context
MBGP	Multicast Border Gateway Protocol
MBMS	Multimedia Broadcast/Multicast Service (for UMTS)
MLD	Multicast Listener Discovery
MPR	Multipoint Relay
MSDP	Multicast Source Discovery Protocol
Node-B	The UTRAN base station
OLSR	Optimized Link State Routing
Path-loss	The attenuation undergone by an electromagnetic wave in transit from a transmitter to a receiver in a telecommunication system.
PIM-SM	Protocol Independent Multicast - Sparse Mode
QoS	Quality of Service
RAN	Radio Access Network
Relay	Repeater for a wireless data-channel.
RNC	Radio Network Controller
SDMB	Satellite Digital Multimedia Broadcast
SGSN	Serving GPRS Support Node

Softstate	Temporary state information that is deleted when a timer expires. The timer is refreshed before timeout if the state information is still valid and should be maintained.
SSM	Single Source Multicast
SSM	Source Specific Multicast
TDD	Time Division Duplex
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network

CONTENTS

PART I

PART II

PAPER A

PAPER B

PAPER C

RESEARCH REPORT G

PART I

INTRODUCTION

1 Introduction

1.1 Background

The first release of the 3rd generation (3G) of mobile networks (e.g., Universal Mobile Telecommunication Services (UMTS) [2, 48]) was just completed when this thesis work started. UMTS was designed to provide wireless access to the existing Internet services as well as UMTS specific services. A wide range of multimedia content was predicted to be an important set of new services. However, the UMTS network described in the first release did not have enough resources to support several concurrent multimedia streams.

A UMTS channel could provide maximum data-rates ranging from 64Kb/s - 2Mb/s depending of the environment. In later releases, the High Speed Downlink Packet Access (HSDPA) channel has been introduced to provide maximum data-rates up to 11Mb/s. However, the highest data-rates are only available to users that experience a good channel quality. More important, the maximum available data-rate to the user also defines the total available radio resources in each cell/sector, thus each radio cell/sector can support only a few high data-rate users simultaneously.

One example of a high-data rate service - streaming video of a sports event - requires a channel capacity of at least 256kb/s to provide adequate small screen video quality [30]. The resources required to support a wireless channel with a given bandwidth is proportional to (at least) the square of the channel range. Consequently, a bandwidth of 256kb/s might be available in the close vicinity of the base station, however it will be difficult to support at the cell border. As wireless Internet access increase in popularity, we believed that the demand for high data-rate connections would rapidly exceed the available capacity.

The network capacity can be increased by deploying more base stations. However, such a solution has several disadvantages: First, there is a worldwide scarcity of frequencies suitable for outdoor mobile radio communication. Second, infrastructure equipment for the UMTS radio access network is expensive. Thus, smaller cells increase the cost/bit significantly.

UMTS is an evolving network architecture that is being standardised in several stages. Its capabilities are still being extended. New research intended to enhance UMTS's capacity for high data-rate services would therefore be of value.

We have focused on a subset of the high data-rate services; group services (i.e., services where the same data is sent to a group of subscribers). We believed that efficient network support for this resource-greedy service type, would become one of the enablers for a successful future wireless network. During this work there has been an increasing focus on group services from the research community [46]. There has also been an increasing focus on heterogeneous wireless networking as a mean to increase the capacity of 3G and beyond networks. In heterogeneous wireless networks the resources needed to transmit high-bandwidth services can be shared among the cooperating network technologies. One example is the Satellite Digital Multimedia Broadcast (SDMB) [86] technology developed in the European IST project MAESTRO [70]. The SDMB system is intended to complement the UMTS network with broadcast capacity for multimedia services.

Multicast is one alternative delivery method for group services. It is used on the Internet for e.g., video conferences, radio and video streaming, and game playing. Its advantages are bandwidth savings over thinner links, and reduced resource consumption in servers. Multicast improves the network's resource consumption by transmitting a packet requested by several users, only once on each link. The first UMTS release allowed a user to join an existing Internet group service; however, the network did not support multicast distribution. The service was delivered to the receiver through a point-to-point tunnel between the UMTS gateway and the mobile node. Clearly, this was an inefficient solution wasting limited UMTS networ resources.

We have formulated and analysed several schemes and architectures that improves UMTS's capacity for high data-rate group services.

1.2 Motivation and Research Overview

The path of our work was formed as the research progressed. In the following we describe the motivation and intermediate findings that set the course of our research. This description includes a brief overview of our work. The main contributions are presented in Section 3.2

The research for this thesis has involved several technologies and protocols. We have studied physical layer through to network layer for IEEE 802.11-based [51, 52] ad hoc networks and the infrastructure-based UMTS [2, 48] network. The common theme in all our work has been analysis of methods to reduce the UMTS resources required to deliver high-bandwidth group services to the consumers, and thereby increase the availability of such services.

When our work started, 3GPP had just begun its work on the Multimedia Broadcast/Multicast Service (MBMS) [3, 104] architecture. MBMS would introduce efficient distribution of multicast and broadcast traffic in the UMTS network. In this context, we considered it worthwhile to study the applicability of the existing IETF IP multicast protocols for MBMS (Paper A). Reuse of IP multicast protocols in UMTS would reduce the complexity in application gateways, and reduce the time and cost for development of new protocols, as opposed to the design and implementation of new UMTS-specific protocols. We performed an analysis of the simple IP multicast solution provided with UMTS Release-99, and two IP multicast architectures suggested by us. This work showed that a multicast architecture based on IP multicast mechanisms, could be a possible solution for the evolving MBMS.

In our IP multicast work, we studied multicast transport in the complete UMTS network except the wireless link between the mobile terminal and the base station (the radio cell). However, the bottleneck in most UMTS network deployment is the capacity of the radio cell. We have therefore focused on solutions to reduce the use of wireless resources in the radio cell (for multicast traffic), in the remainder of our work.

In Paper B we analysed and compared three different channel types (unicast, broadcast, and sticky-channel broadcast) for multicast distribution in the UTRAN (UMTS Terrestrial Radio Access Network) [6] radio cell. The motivation for this study was the following: It is cost-effective to use a one-to-many (broadcast) channel for multicast distribution in the wireless cell for densely populated groups; however, the unicast channels utilize the radio resource much more efficiently than a broadcast channel. A broadcast channel has to be robust enough to reach all the multicast receivers. The resource cost is defined by the multicast receiver which at any instant experiences the worst channel quality. This robustness requires much wireless resources. Neither are fast power adjustments nor packet scheduling based on channel quality available for a broadcast channel in contrast to a unicast channel. Thus it is best to choose wireless unicast to each of the multicast members in sparsely populated groups.

As a third alternative to existing unicast and broadcast, we suggested a physical channel type named sticky-channel where the mobile multicast terminal in a wide cell-border region “sticks” to a broadcast channel from the neighbouring cell. This channel was based on the observation that during resource planning of the UMTS network it is recommended to build your network such that a terminal is covered by two or more base stations in at least 40% of its connection time (e.g., [59]). The purpose of the sticky-channel was to reduce the total wireless resource consumption in a given region by eliminating the need for multicast transmission from a subset of all base stations deployed to serve the region. The sticky-channel showed a marginal reduction in wireless resources for sparsely populated multicast groups compared with the broadcast and unicast options.

To further reduce the UTRAN radio resources needed to distribute high-bandwidth group services, we identified heterogeneous wireless networking as one feasible method. In our opinion, the 4th generation (4G) of mobile networks will most likely consist of a multitude of wireless standards that cooperate to form the 4G radio access network.

In Paper C we suggested a heterogeneous multihop wireless access network for multicast traffic. In a heterogeneous wireless access network, the cost to distribute a high-bandwidth group service can be shared among the cooperating technologies, e.g., the range of a UMTS radio channel can be reduced and local IEEE 802.11-based Mobile Ad Hoc Networks (MANETs) [25] can be used to forward the data to users located outside the reduced UMTS channel (see Figure 1 for an example). Since the path-loss on the radio channel is proportional to the n^{th} power of distance where n ranges between 2 and 4 [82], a reduction in the channel range can free a significant amount of UTRAN radio resources, and thus increase the network capacity.

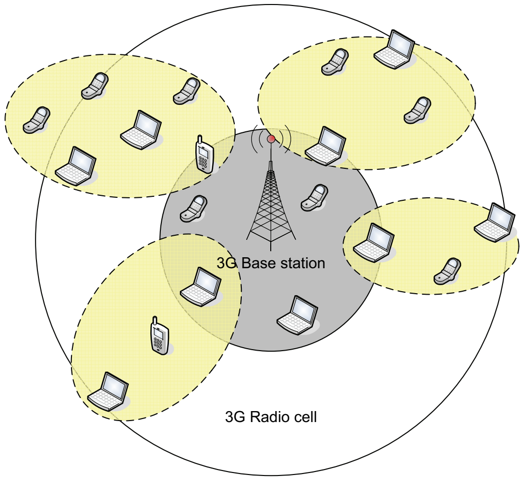

Figure 1: This figure shows a heterogeneous wireless access network. The range of the UMTS broadcast channel has been reduced to cover the small grey area, and MANETs are used to forward the multicast data to terminals in the remaining area of the radio cell.

When we started our work with the heterogeneous wireless access network, MANETs was becoming adequately mature, and mobile terminals with both IEEE 802.11 and UTRAN transceivers were readily available. Thus a combination of 3G cellular networks and IEEE 802.11-based MANETs was a feasible and easy-to-deploy architecture.

Mobile Ad Hoc Networks (MANETs) [25] operate independently of a fixed or preplanned infrastructure. The networks are autonomous and may be isolated, or interconnected to other networks through gateways. Each node in a MANET is also a router. A stand-alone routing protocol for an ad hoc network dynamically forms multihop routes from source to destination. Nodes in a MANET can be highly mobile, consequently, the presence of links and thus the network topology, is constantly changing. A MANET routing protocol must therefore be able to handle rapid route changes.

In Paper D we sketched two multicast routing protocols for the heterogeneous multihop access network; one distributed protocol with some central support and one fully centralised protocol. The distributed protocol is described in detail in Research Report G. We chose to design new multicast protocols for our heterogeneous architecture to allow for efficient exploitation of the available centralised infrastructure (UMTS). The other option was to modify existing MANET multicast protocols. Some central support improves the efficiency of MANET multicast routing. Our choice allowed us to build a distributed routing protocol based on (what we believed to be) the best components from several popular MANET protocols and introduce some central support to this design.

Encouraging results from the preliminary analysis of the heterogeneous wireless access network performed in Paper C and Paper D motivated for a detailed analysis of the architecture and the distributed multicast protocol (Research Report G). The analysis is presented in Paper E.

Traditionally the telecommunication industry has had a high focus on network reliability and quality of service (QoS) whereas best-effort data networks rely on redundant network capacity to support some service quality. In resent years QoS mechanisms has become available for data networks as well, however large differences still exists between different network types. UMTS is a reliable network with efficient QoS mechanisms, whereas the MANETs are unreliable best-effort networks. Thus an important aspect of our detailed analyses of the heterogeneous wireless access network has been to study the service quality in the heterogeneous multihop network in comparison with the standard, reliable, one-hop, wireless UTRAN channel. We used this analysis to identify the maximum MANET size (and thus shortest UTRAN channel range) that provided an acceptable service quality. The heterogeneous architecture allowed a reduction of the UTRAN broadcast channel to approximately 45% of the cell range.

During our work with the heterogeneous wireless access network, we observed that most packet-loss was loss of single packets or short sequences of two or three packets. This led us to believe that forward error correction (FEC) used as an erasure code on the multicast packets could improve the performance of the heterogeneous network. This could allow larger MANETs and therefore, a further reduction of the range of the UMTS multicast radio channel (and consequently also a reduction in the required radio resources).

In packet-level FEC (FEC as an erasure code), a number of redundant packets are generated for each block of data packets. The redundant packets are used at the receiver to regenerate a number of lost packets. Thus a FEC encoded multicast stream can be transmitted on a path with a higher packet loss ratio. Consequently, FEC could allow longer ad hoc paths and thus shorter UMTS channel range in our heterogeneous architecture. An analysis of multicast with FEC in the heterogeneous wireless access network showed some throughput improvement, but at a very high bandwidth cost (Paper F).

1.3 Thesis Outline

The thesis is organized in two parts. Part I is an introduction to the areas of which our work depends, and to the areas where the thesis contributes, whereas Part II consist of a set of published articles that presents the results of our research.

Part I

After a brief introduction that describes the background, motivation and outline of the thesis, chapter 2 gives an overview and presents related work of the research fields coved by our work. Chapter 3 describes our research methods, a summary of the main contributions and gives a short discussion of the work.

The list of figures and the list of terms and acronyms given in the beginning of the thesis are restricted to Part I. Likewise, since each article includes a reference list, the reference list found at the end of Part I, is exclusive to this part of the thesis.

Part II

Part II consists of the following six research papers and one research report:

- Multicast in 3G Networks: Employment of Existing IP Multicast Protocols in UMTS
- Sticky Point-to-multipoint Channel For Multicast in UMTS
- Multicast Service Availability in a Hybrid 3G-cellular and Ad Hoc Network
- Multicast-service Distribution on a Cellular Network Assisted by Local Ad Hoc Networks
- A Heterogeneous Cellular and Ad Hoc Network Architecture for Multipoint Streaming: A detailed performance analysis
- Evaluation of Packet-level FEC with Multicast Streaming for a Heterogeneous 3G-Cellular and Ad Hoc Network
- Multicast in a Heterogeneous Cellular and Ad Hoc Network: Specification of an Ad Hoc Routing Protocol with Cellular Assistance

2 Overview and Related Work

This chapter presents an overview of the research areas covered by this thesis. The purpose is not to give a comprehensive overview of the different areas, but to provide necessary background information, and describe important work related to our research.

The basics of IP multicast are presented in Section 2.1. Multicast is a common component in all our work, and the two multicast architectures for UMTS that we have suggested and analysed were based on IP multicast mechanisms. In Section 2.2 we present the Multimedia Broadcast/Multicast Service (MBMS) [3] for UMTS; MBMS is the current multicast architecture for UMTS. Section 2.3 gives an overview of multicast routing for Mobile Ad Hoc Networks (MANETs). MANET multicast mechanisms are used in the multicast protocols we have suggested for our heterogeneous wireless access network. The area of multihop and heterogeneous wireless access networking for 3G and beyond is covered in Section 2.4. We have also studied packet-level forward error correction (FEC) for multicast on the heterogeneous wireless architecture; section 2.5 describes the use of packet-level FEC in wireless networks.

2.1 Internet Protocol (IP) Multicast

Multicast distribution over the Internet is based on the model introduced by Stephen Deering [27]. It is intended for one-to-many, few-to-many and many-to-many communication. Multicast improves the network's bandwidth budget by transmitting a packet requested by several users only once on each link. Consequently its advantages are bandwidth savings over thinner links, and reduced resource consumption in servers.

Multicast sessions use a particular class of IP addresses, class D. Originally the portion of the address range that was intended for global multicast groups, was an unregulated and flat address space. Thus any multicast group could ask for any multicast address, which might result in address collisions between groups. A portion of this range has later been assigned to GLOP addressing [71]; each Autonomous System (AS) is given a short range of static global multicast addresses in the GLOP address space. The Multicast Address-Set Claim (MASC) Protocol [80] provides one suggestion for how to handle global dynamic multicast address allocation for the remaining range, however this mechanism is not widely used on the Internet.

In the IP multicast model, sources need not be members of the multicast group, and can dynamically join or leave sessions. The multicast receivers can also dynamically join or leave a session. The receivers do not need to know the sources in the session, and likewise, the sources do not need to know the receivers in the session.

Typically, nodes that want to join a multicast session inform the local router of this intent via the Internet Group Management Protocol (IGMP) [19], or Multicast Listener Discovery (MLD) [96]. Local routers that serve multicast members, join the multicast distribution tree maintained by a multicast routing protocol. Interdomain multicast routing is most commonly handled by the Multicast Border Gateway Protocol (MBGP) [13], with additional support from the Multicast Source Discovery Protocol (MSDP) [32] for shared distribution tree protocols. The multicast routing protocols for intradomain routing can be classified into three groups:

- Shared distribution tree protocols
- Source specific tree protocols
- Stateless multicast.

Shared Distribution Tree Protocols

In shared tree multicast routing protocols, all sources for one multicast session use the same distribution tree. A core node is selected to root the multicast tree for a multicast session (Figure 2). New multicast members join the multicast tree by sending a join message towards the root of the common tree, and are subsequently attached to the tree. Using a shared multicast tree has the disadvantage that packets are distributed to the multicast group along paths that can be much longer than the shortest paths from sources to receivers. The advantage is that routers only need to maintain one state object for each multicast group, and signalling traffic is required to maintain only one tree. Shared distribution tree protocols are also referred to as any source multicast (ASM).

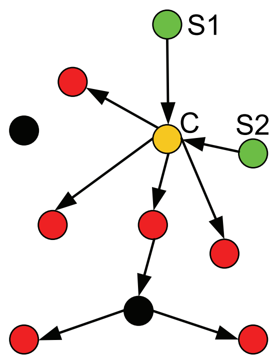

Figure 2: The figure shows shared tree routing. Each source (S1 and S2) sends the packets to the core node (C) which in turn forwards the data onto the multicast tree.

Protocol Independent Multicast-Sparse Mode (PIM-SM) [33] and Core Based Trees (CBT) [12] are the two most popular shared tree routing protocols. In CBT, multicast packets from a source that is not a multicast member, are tunnelled to the core node which in turn forwards the data on the multicast tree. CBT maintains bidirectional trees; thus source data from a multicast member are distributed on the tree directly by the source. PIM-SM builds a unidirectional tree from the core to the multicast members. In PIM-SM, a new source always sends the encapsulated multicast packet to the core (Rendezvous Point), which in turn forwards the data on the multicast tree. Next, the core router performs a source specific join towards the source, to attach the source node to the distribution tree and avoid the encapsulation and (possibly) inefficient paths. PIM-SM also allows receivers to switch to a source-based shortest path tree.

PIM-SM maintains temporary multicast state information (softstate) in the routers. The softstate is refreshed with periodic join messages addressed to the core node, from the routers that serve multicast members. CBT keeps hardstate information that is maintained with acknowledged join-requests and quit-requests. The shared multicast tree model is most useful in a few-to-many scenario for sparse member distributions.

A variant of PIM-SM, called Bi-directional Protocol Independent Multicast (BIDIR-PIM) [37], is being specified, this protocol is intended for many-to-many scenarios.

Source Specific Distribution Tree Protocols

The alternative to shared trees is to build source specific trees from each source (Figure 3). A source specific tree trades complexity in the form of a large state space in routers and a larger number of links, for an efficient shortest path distribution tree for the multicast data. Source Specific Multicast (SSM) [47, 17] represents this routing type. SSM is in reality the source specific parts of PIM-SM [33]. These protocols are also referred to as single source multicast (SSM). SSM is most useful for one-to-many scenarios for sparse member distributions.

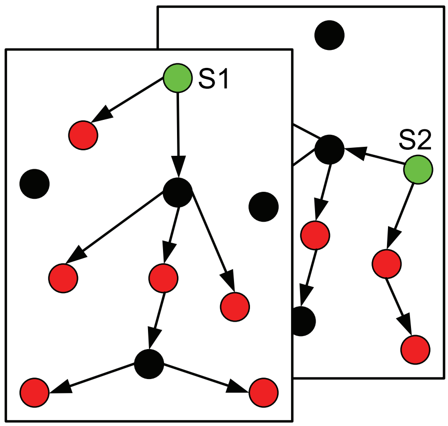

Figure 3: The figure pictures source specific multicast routing. One multicast tree is established for each source (S1 and S2). The red nodes are multicast members.

A different type of source specific tree multicast protocols is represented with Protocol Independent Multicast-Dense Mode (PIM-DM) [7]. This protocol type is intended for groups with high multicast member density. In PIM-DM, multicast sources periodically flood the multicast data in a given domain. Routers which are not interested in the multicast data explicitly prune their branch of the distribution tree. Due to the periodic flooding, these protocols are not efficient for sparse multicast groups, and do not scale well with increasing group size.

Stateless Multicast

Stateless multicast is a different approach to multicast distribution. This protocol type does not maintain the multicast group states in the routers. The addresses (or some coded address) of all receivers must be present in the header of a stateless multicast data packet. Upon reception of a multicast packet, the router consults its unicast protocol and chooses the correct network interface for the next hop towards each of the receivers. The packet is replicated, and the address list updated when different next-hop nodes are required to reach all the receivers.

Obviously this type of multicast does not scale well with increasing group size; however it, scales very well with increasing number of small multicast groups. Stateless multicast must be explicitly supported by all routers the multicast packet passes on its way to the destinations; this is the main disadvantage with this protocol type. This protocol type is therefore most useful when the multicast source is connected to the rest of the network via narrow-bandwidth links. In this case only the narrow-bandwidth routers must support stateless multicast. The multicast packet is expanded to multiple unicast in a gateway between the narrow-band network and the rest of the network.

The IETF draft "Explicit Multicast (Xcast) Basic Specifications" [18] describes the general operation of this type of protocol. Differential Destination Multicast (DDM) [55] is an example of an Xcast protocol.

2.2 UMTS Multimedia Broadcast/Multicast Service

In the first UMTS release, multicast support in the UMTS standard was an optional solution that provided access to a multicast service by means of unicast tunnelling. The IP multicast routing was terminated in the GGSN (Gateway GPRS Support Node) between UMTS and the Internet (ref. Figure 4). GGSN also served as an Internet Group Management Protocol (IGMP) [19] designated router, and performed IGMP signalling on point-to-point data channels with the mobile terminals that wanted to be multicast members. In this UMTS architecture only the GGSN and the mobile terminal were multicast aware. This solution enabled a mobile terminal to access an existing multicast service on the Internet, but did not exploit the potential network resource gain associated with multicast routing.

The Multimedia Broadcast/Multicast Service (MBMS) [3, 104] was introduced in the UMTS standard (Release-6) to improve the efficiency of group communication in UMTS. Our work presented in [43] and [45] suggested possible solutions to some parts of the MBMS architecture. The final MBMS design is split into the MBMS Bearer Service and the MBMS User Service.

The MBMS Bearer Service

The MBMS Bearer Service includes a Multicast and a Broadcast Mode. The advantage of the MBMS Bearer Service compared to the original UMTS bearer services is that a multicast routing mechanism is introduced. One MBMS packet flow is replicated (when needed) by the UMTS network routing nodes: GGSN (Gateway GPRS Support Node), SGSN (Serving GPRS Support Node) and RNCs (Radio Network Controller) (ref. Figure 4).

In MBMS each multicast group is identified by a multicast IP address and an Access Point Name (APN). The APN represents a specific GGSN, and is set as the source of the data packets for the multicast group. In contrary to IP multicast, MBMS allow only one known source for each group. Also, unknown multicast members are not allowed in MBMS; a User Equipment (UE) must register a join with the Broadcast Multicast - Service Center (BM-SC) to join a multicast group. This modified multicast model, where the source and all the receivers are

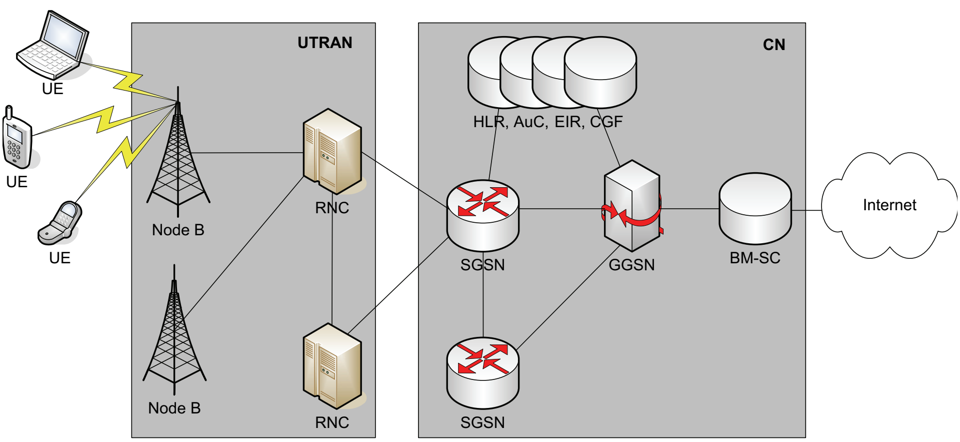

Figure 4: This figure shows the UMTS (Universal Mobile Telecommunications System) elements participating in the MBMS architecture. CN (Core Network) is a high capacity backbone network, while UTRAN (UMTS Terrestrial Radio Access Network) incorporates the low capacity radio links.

known, eases the mechanisms for security and the mechanisms for source and receiver charging. The MBMS architecture addresses to some extent the multicast deployment problem issues raised in [28].

MBMS uses a query-and-response mechanism similar to unsolicited join and leave of IGMP [19], both for group management and for establishment of multicast forwarding trees. A comparison between the MBMS mechanism and IGMP can be found in [103]. When a user equipment wants to join a multicast group, it sends a join to the GGSN, which in turn verifies with the BM-SC whether this user is allowed to join. The GGSN receives the address of the APN that administers the multicast group, from the BM-SC. A positive acknowledge to this join initiates a sequence of MBMS signalling messages. These messages create/modify the MBMS Bearer Context (MBC) state in all UMTS nodes on the path between the mobile terminal (UE) and the APN for the multicast group. The MBC stores (among other parameters) the multicast IP address, and the downlink interfaces that serve multicast members.

When a mobile multicast member wants to leave the multicast group, the mobile terminal (UE) sends a leave message to the GGSN. This message unsubscribes the user from the multicast group, and removes the multicast state associated with this user in the network nodes.

When a mobile multicast member performs a handover to a new base station, this member must be associated with a new link in the multicast tree. The standard UMTS signalling messages that support node mobility have been augmented to include information to support multicast tree maintenance.

MBMS creates a standard multicast forwarding tree to distribute the multicast data in the Core Network and part of the radio network (UTRAN). The multicast tree spans the GGSNs, SGSNs and the RNCs.

The method for multicast data distribution from the RNCs via the base stations (Node-B) to the mobile user equipment, is selected based on the number of multicast subscribers associated with each radio cell/sector. If there are few members located in a cell/sector, a normal point-to-point unicast connection is setup between the RNC and the user equipment. Otherwise a common path is established between the RNC and the base station, and a broadcast channel is allocated for the wireless link to the mobile terminal (UE).

MBMS may use an advanced counting scheme to decide the approximate number of multicast members associated with a radio cell/sector. Based on this number, the network chooses whether zero, one, or more dedicated (i.e. unicast) radio channels lead to a more efficient resource usage than one common (i.e. broadcast) radio channel.

The Forward Access Channel (FACH) was selected to be the common point-to-multipoint wireless channel in the MBMS RAN (Radio Access Network) [5, 77]. FACH uses soft-combining and selective-combining to improve the reliability of the data reception and thus allow a reduction in the transmission power for the broadcast channel. Selective- and soft-combining will both combine the broadcast transmissions from adjacent base stations. Soft-combining is used to combine received power before channel decoding while selective-combining decodes the signal from each base station independently and compares the results before it chooses the one that has the highest probability of being correct.

The MBMS User Service

The MBMS User Service is basically the MBMS Service Layer. It offers a streaming and a download delivery method. The Streaming Delivery method can be used for continuous transmissions like Mobile TV services. The Download Method is intended for "Download and Play" services. To increase the transmission reliability, an application layer FEC code may be used. Further, a file repair service may be offered to complement the download delivery method.

2.3 Multicast for Mobile Ad Hoc Networks (MANETs)

A Mobile Ad Hoc Network (MANET) [25] is a multihop wireless network. It is a self-configuring network of mobile routers (and associated hosts) connected by wireless links. The routers are free to move randomly and organise themselves arbitrarily. The network's wireless topology may therefore change rapidly and unpredictably. Such a network may operate in a stand alone fashion, or it may be connected to the Internet.

The tree reorganization in MANETs is more frequent than in conventional wired networks, since the multicast protocols have to respond to network dynamics in addition to group dynamics. Consequently, multicast protocols designed for fixed networks do not support the dynamics of MANETs very well. The multicast protocols suggested specifically for MANETs can be classified in four categories [24]: Tree-based protocols, meshed-based protocols, hybrid protocols, and stateless multicast. In addition to these four types, we include multicast by means of broadcast in our discussion. Geographic multicast protocols have emerged as a sixth category; however, since we did not consider geographic multicast for our work, these protocols are therefore not included in this introduction. Neither did we study energy-efficient protocols, nor multicast protocols that attempt to provide quality of service guarantees.

The tree-based protocols are based on the IP multicast protocols for fixed networks. These protocols strive to create an optimal multicast distribution tree where the multicast data is distributed to all members with a minimum number of link broadcasts. These protocols are designed to handle some mobility; however, as the node mobility increases, the multicast throughput decreases (and the signalling traffic increases). A basic tree-based protocol is not able to repair broken links quickly enough in a highly mobile network.

Mesh-based protocols were introduced to increase the multicast distribution trees' robustness to node mobility. These protocols introduce some redundancy in the multicast distribution tree; when a link is broken in a mesh tree, the multicast data will (in many cases) continue to flow on a redundant link. This allows the protocol to continue forwarding multicast data while the broken link is being repaired. Clearly the multicast distribution is not optimal on a mesh since the data might travel on parallel paths to the multicast members; however, this inefficiency is traded for better multicast throughput in highly mobile networks.

The hybrid multicast protocols attempt to get the most out of both the tree-based and the mesh-based protocols by combining the two.

In multicast by means of broadcast (flooding), multicast data is distributed with flooding. In this case there is no need for a multicast routing protocol to maintain a multicast distribution tree, thus there is no signalling overhead. However, there will clearly be an overhead due to a high number of redundant packet transmissions; this redundancy is reduced as the density of multicast members increase. The method is very robust for mobility (it represents a fully redundant mesh).

Stateless multicast make use of the unicast routing protocol, thus the unicast protocol's robustness to node mobility is important for the performance of this multicast type. No multicast signalling is required, but all addresses of the multicast members must be listed in the header of each data packet. Stateless multicast is therefore efficient only for small multicast groups.

2.3.1 Tree-based Multicast Routing Protocols

Tree-based multicast is the traditional multicast forwarding mechanism used in fixed networks. Many of the protocols designed for ad hoc multicast is based on this structure. Tree-based mechanisms can be source initiated or receiver initiated, source-based trees or shared trees. Routing information can be maintained proactively or on-demand. Due to the dynamic nature of these protocols, most of them store the current spanning tree information in temporary softstates.

We have chosen to briefly describe the following tree-based multicast protocols:

- AMRIS [99] was one of the first stand-alone MANET multicast protocols. It uses ID numbers to identify a node's position in the tree hierarchy and consequently avoid routing loops.
- MAODV [85] and MOLSR [60] are multicast extensions to two popular unicast routing protocols. MAODV was also one of the first MANET protocols, it introduced sequence numbers to avoid routing loops. MOLSR exploits the efficient proactive routing mechanisms in OLSR [23].
- ADMR [54] gets the most out of passive signalling; it uses all overheard traffic to maintain multicast routing information, thus this is a protocol with low signalling overhead.
- ABAM [93] is an example of a protocol that uses a criteria different from the shortest path to establish the multicast tree. ABAM performs well i highly mobile networks because it chooses to route the multicast data over stable routing paths (paths between nodes with little mobility).
- ACMRP [76] is a tree-based protocol with a loose tree structure, it paves the way for a discussion of tree-based versus mesh-based protocols.

Ad hoc Multicast Routing Protocol Utilizing Increasing Id-numbers (AMRIS) [99]

This is one of the first stand-alone MANET multicast protocols. AMRIS constructs a common distribution tree for all sources in a group. The tree construction is initiated by a message from an elected main source. This message is flooded to all nodes in the network, and it assigns an id number to each node; the id number indicates the node's level in a forwarding tree. Multicast members reply with a unicast join request to the neighbour with the lowest id number. In this way, the join request propagates along the reverse path towards the multicast tree. Each network node sends a periodic one-hop message to announce its presence and current id number for all active multicast groups. Broken links are detected based on this message: The downlink node (with highest id number) performs a local link repair, where the id number is used to avoid routing loops. Multicast state information times out based on missing neighbour messages.

This protocol does not do any refresh of the distribution tree; after a sequence of link repairs, the forwarding tree might be less optimal. Neither does the protocol specify how to handle partition and reunion of the tree, or how to elect the main source.

Multicast Ad-hoc On-demand Distance Vector (MAODV) [85]

MAODV is an extension to the popular unicast AODV [79], and uses AODV's route discovery mechanisms to find a path to the multicast distribution tree. One common multicast distribution tree is maintained for each active multicast group. A node that wants to join a multicast group floods a route-request (unicast can be used if the node has a route to the group), and on-tree group members reply to the request along the reverse path of the route request. Eventually the joining node activates one, of possibly many, replies. The first multicast member becomes the leader of the group, and periodically floods a group-hello message to maintain the forwarding tree.

All nodes monitor the radio channel for neighbour traffic and thus detect lost neighbours (links) within a timeout period. The node downstream of the break attempts a link repair with a range limited flood of the route-request message. The node upstream of the break waits some time for a possible reconnection before it removes the state information associated with the downlink node, and performs a tree prune in case it is now a non-member leaf node. MAODV has mechanisms to handle partition and reunion of the multicast tree.

Multicast Optimized Link State Routing (MOLSR) [60]

MOLSR is a suggested extension to the popular proactive unicast protocol, OLSR [23]. This protocol maintains one tree for each source (source-specific trees), and the tree formation is initiated by the source with a flood of the source-claim message. Note that the flooding is done by the OLSR Multipoint Relays (MPRs), and is thus less resource consuming than basic flooding. Multicast members respond to the source-claim with a one-hop parent-claim message. The message follows the path to the source as specified in the routing table, but it does not necessarily follow the reverse path of the source-claim.

Periodic floods of the source-claim message with confirm-parent responses are used to maintain the forwarding tree. In addition, changes in the unicast topology can trigger changes in the multicast routing tree. If any node on the tree detects a change in the next hop towards a multicast source, it joins the new hop and may leave the previous parent (otherwise a soft-state mechanism will eventually remove the old link).

There is no need for local multicast link repair since the unicast protocol will find an alternative route. Neither does a source-based tree need to handle partition and reunion of the distribution tree. The protocol has mechanisms to handle routers that are not multicast enabled. Due to the proactive nature of the protocol, there is a high overhead in most scenarios; the advantage is a more robust protocol in highly mobile networks.

Adaptive Demand-Driven Multicast Routing Protocol (ADMR) [54]

ADMR attempts to reduce the periodic active signalling traffic to a minimum. This protocol maintains a distribution tree for each source; the source starts the establishment of the multicast tree by flooding the first multicast packet (including an ADMR header) on the network. Multicast members respond with a join message on the reverse path to establish the tree. Subsequent multicast members attach to the tree with a flooded multicast solicitation message, wait for a response and validate one of the response messages in a three-way-handshake.

ADMR monitors the traffic pattern of the multicast stream. Based on that, the protocol can detect link breaks in the tree, as well as sources that have become inactive, and branches which are no longer needed. A limited number of "keep-alive" packets are transmitted in temporary breaks in the source data. The downlink node of a broken link attempts a link repair with a flooded reconnect message, and the source responds with a reply. If receivers experience frequent link breaks, the source is informed, and the source floods a few multicast data packets to start a refresh of the multicast forwarding tree.

Associativity-Based Multicast Routing (ABAM) [93]

ABAM creates source-specific trees where the trees are constructed based on link stability rather than hop distance. Each node keeps track of the link stability to each of its neighbours by registering the reliability of periodic one-hop beacons. A three-way handshake is used to create a multicast tree. Each source floods the network with a multicast query; nodes receiving the query will append the link stability, signal strength, power life, etc., to the query message and subsequently rebroadcast the message. Multicast receivers chooses the multicast query that has traversed the most stable route and unicast a reply on this path. Finally the source validates the tree by sending a setup message along the paths of the multicast tree.

Local repair is used to maintain the tree. The repair can be either a branch repair, a sub-tree repair or a full tree repair (in case the source node loses its connection to the rest of the three). All repairs are performed with a three-way handshake in a query-reply-validate sequence.

This protocol achieves a high throughput in mobile networks since it chooses stable links for the forwarding tree. The disadvantage is a lower multicast efficiency due to larger paths. There is also the risk that the most stable paths become congested since all sources and groups tend to choose the same paths.

The Adaptive Core Multicast Routing Protocol (ACMRP) [76]

ACMRP is a shared tree protocol. The tree establishment is initiated by the source, and it uses a periodic request (flooding) - reply, strategy to maintain the distribution tree. In this protocol there is one dynamic core, which is responsible for the formation and maintenance of the forwarding tree. One common tree is used by all the sources in a multicast group.

All nodes on the network keep a table of the current core for each group. The first source of a group takes the role as the core. Periodically this role is migrated to a node closer to the center of the multicast network, to improve the efficiency of the common multicast distribution tree compared to source-based trees. The core initiates a limited link state signalling sequence to identify a node closer to the center of the group that should take over the core role. Thus the core role converges towards the best location in the group.

The authors classify this protocol as mesh-based; however, there is no explicit formation of redundant links in the multicast distribution tree. The broadcast nature of a radio network with a common channel is exploited to provide some redundancy. A node on the forwarding tree accepts new multicast packets from any of its neighbours. In tree-based protocols on fixed networks, a node accepts a packet only from its defined uplink to avoid loops. When a MANET node accepts a packet from any of its links, it will receive a duplicate copy of the multicast data packet in the cases when there are two or more branches of the multicast distribution tree within radio range of each other. This redundancy can be exploited to improve throughput if each multicast packet can be uniquely identified, and each node keeps a history (cache) of received data packets. Several tree-based multicast protocols exploit this redundancy.

2.3.2 Mesh-based Multicast Routing Protocols

The mesh-based protocols provide richer connectivity compared with tree-based structures. Some level of redundancy is added to the forwarding trees, thus these trees will in many cases have an alternative path to bypass a broken link. These protocols trade low overhead and efficient multicast routing for better connectivity (and thus throughput). The relatively large overhead in the mesh protocols consists of signalling to create and maintain the mesh, and redundant data packets. Mesh-based mechanisms can be source initiated or receiver initiated, source-based meshes or shared meshes. Most of them maintain temporary softstates.

We have chosen to briefly describe the following mesh-based multicast protocols:

- CAMP [34] is an efficient mesh protocol with little signalling overhead. It uses a common tree to reduce the signalling traffic, and is developed from the well known CBT [12] protocol for fixed networks.
- ODMRP [61] is very robust in highly mobile networks. This protocol has shown high throughput in several performance studies (e.g., [62]). The protocol has a high overhead, therefore we also present DCMP [26], which is an extension to ODMRP, with reduced overhead.

The Core-Assisted Mesh Protocol (CAMP) [34]

CAMP extends the operation of the Core-Based Tree (CBT) [12] protocol with mesh functionality for wireless mobile networks. The protocol depends on an underlaying proactive link state unicast protocol, and this is one of the reasons why it has little signalling overhead. CAMP establishes a common mesh for all sources in the multicast group. The protocol is receiver initiated in the sense that multicast members unicast a join towards the core to connect to the multicast tree. The address of one (or several) core(s) is distributed in group membership reports. If no core is known, the receiver floods the join to all nodes in the network; a node that already is a multicast member can respond to the join and create a path to the mesh.

Each multicast member continuously consults the unicast routing table to check if the multicast packets arrive from its neighbour that is on the reverse path towards the source. If not, then the multicast member sends a message on the shortest path towards the source to include this path in the mesh. The new path will be part of the common tree that is used by all sources, thus the number of sources, and their locations, define the level of redundancy in the network. Old links in the mesh are removed upon a timeout.

Mesh maintenance is handled by the underlaying unicast protocol's modification of the path towards the source. Members leave the multicast tree with a quit notification message. The role of the core is to reduce the network signalling and to provide a first/redundant path to the mesh. The protocol has mechanisms to handle mesh partition and reunion. The core does not need not be a multicast member.

The On-Demand Multicast Routing Protocol (ODMRP) [61]

This is a popular mesh protocol. It is a stand-alone protocol that does not require underlying unicast routing. One common mesh is formed for each multicast group. Each source establishes a source specific forwarding tree by flooding a join-query to all nodes in the network. Multicast members respond with a reply. The resulting mesh is the union of all the source specific forwarding trees, thus the level of redundancy in the mesh is dependant on the number of sources and their location relative to the receivers. Each source periodically floods a join-query to refresh the mesh; thus group membership and forwarding state is maintained through softstate. Members do not send an explicit leave when they want to leave the multicast group, in stead the connection is left to timeout. If a link break is detected, a local temporary repair is performed until the next flooding of the join-query establishes a new tree.

This protocol provides good connectivity, and consequently good throughput, in highly mobile networks at the cost of a high signalling load and some redundant data transmission. Due to the periodic flood of the join-query from each source, the signalling load is high. To reduce this load, the protocol provides an extension for passive clustering (similar to MPRs in OLSR [23]) to reduce the flooding overhead.

To further reduce the signalling overhead, the Dynamic Core Based Multicast Routing Protocol (DCMP) [26] has been proposed. DCMP is ODMRP with some modifications. The basic operation of ODMRP is kept, and in addition DCMP introduces some signalling to organize the group sources in three categories: active sources, core active sources and passive sources. The active sources behave like ODMRP sources, and the core active sources operate as ODMRP sources for nearby passive sources. Thus in situations with many sources, this protocol variant reduces the level of redundancy in the group mesh, and consequently the overhead is reduced. An active source is allowed to be core for at most a defined number of passive sources, and a passive source is allowed to be at most a defined number of hops away from the core source. The number of active sources in the group, and thus the level of redundancy, can be adjusted with those two constants.

2.3.3 Hybrid Tree-based and Mesh-based Protocols

The tree-based multicast protocols provide high data forwarding efficiency at the expense of low robustness, whereas mesh-based multicast protocols provide better robustness at the expense of higher forwarding overhead and increased network load. Thus, there is a possibility that a hybrid multicast solution may achieve better performance by combining the advantages of both tree-based and meshed-based approaches.

We have chosen to describe the following hybrid multicast routing protocols:

- AMRout [102] uses unicast tunnels to connect multicast members. Non-member routers does not need to be multicast enabled. The multicast distribution tree is created from an underlaying mesh.
- MCEDAR [89] uses a combination of MAC mechanisms and routing mechanisms to establish efficient multicast forwarding. The routing protocol generates a mesh whereas the MAC mechanism optimize the mesh to a tree structure.
- MANSI [87] is an interesting resent protocol that uses random probing packets to improve the multicast forwarding. MANSI identifies areas with high mobility and creates a mesh for these paths.

The Ad Hoc Multicast Routing Protocol (AMRout) [102]

AMRout creates a mesh of source and group members. Non-members are not allowed to be relay in the mesh, thus group members and sources are interconnected with unicast tunnels to form a connected network. A subset of the connection points (tunnels) are dynamically selected to form a multicast distribution tree. This protocol relies on an underlying unicast protocol to maintain the unicast tunnel connections. In addition, the mesh interconnection of the member trees provides some robustness.

Each group in the network has one core that is responsible for discovering new group members and create/maintain the multicast forwarding tree. The first member of a group is the core. Periodically the core floods a join request to all nodes in the network, to find other members. The members respond with an acknowledge to establish a tunnel. The core also multicasts a tree-create message on the mesh to create a distribution tree within the mesh. All nodes that receive a redundant copy of the message send a tree-create-nac to the previous hop to prune that mesh link from the distribution tree. The protocol has mechanisms to handle mesh partition and reunion. The core role is often migrated to a new multicast member (partly to avoid that all tunnels are terminated by a single node).

The Multicast Core-Extraction Distributed Ad hoc Routing Protocol (MCEDAR) [89]

MCEDAR extends the CEDAR unicast mechanisms. CEDAR relies on a proactively maintained set of cluster heads (called cores in [89]), which are used to optimize flooded traffic in a similar manner as the MPRs in the more recent OLSR [23]. In CEDAR a broadcast is performed with reliable unicast transmission between the cluster heads, and a modified MAC layer is used to optimize and suppress some of the redundant unicast traffic.

For each multicast group, MCEDAR extracts a sub-graph of the core-graph to function as the routing infrastructure. The sub-graph is a mesh structure that has a defined robustness factor R. Once the multicast-graph is formed for a multicast group, data forwarding is done on the multicast-graph using the optimized flooding mechanism. This is a receiver initiated protocol where a new multicast member requests its dominator (chosen cluster head) to join the multicast group. The dominator floods (with optimized flooding) a join request and waits for an acknowledge from an on-tree node. If the new member receives acknowledge from more uplink nodes than the robustness factor R, some of the connections are pruned with an explicit leave. The underlying modified MAC layer ensures that one subset of the multicast-graph that represents a source-based multicast tree is used to forward the multicast data. The protocol supports local link repair for the cases where no redundant mesh path is available to bypass a broken link.

A Multicast Routing Algorithm with Swarm Intelligence (MANSI) [87]

The resent MANSI is a protocol with a weak mesh structure. It creates an effective low-cost tree-based forwarding tree, and it also identifies regions with high link-failure rate and establishes a local mesh for those areas. The protocol establishes a core-based shared tree for each group. The first group source takes the role of the core and floods a periodic core-announce message to all nodes in the network. Members respond with join requests along the reverse path.

When the first forwarding tree is established, the protocol uses proactive “swarm intelligence” to improve the paths. This means that all multicast members, except the core, periodically unicast a probe packet (called an ant) to survey a limited area for a shorter or more efficient path towards the core. Each ant packet makes probabilistic decisions of the direction to take, collects information from the visited nodes and returns to the sender with the information. Each node maintains a table where several of its next-hop nodes are listed, linked with information about the cost associated with a path to the core via the relevant node. Upon receiving a core-announce, each member responds with a join request via the neighbour that is listed with the lowest cost. If the node has experienced a high link failure rate, a second join request is sent via the neighbour that is listed with the next best cost to establish a local mesh. The ant packets are also used to do local repair.

2.3.4 Multicast by Means of Broadcast (Flooding)

It is costly to keep a multicast distribution tree for a MANET up to date. The possibly high mobility of the nodes, and the unreliably of the wireless links, require a high frequency of signalling messages to maintain the multicast distribution tree. It might therefore be cost-effective to flood the multicast messages in some situations. The trade-off is unnecessary broadcast of the multicast messages to nodes that are not multicast members versus the signalling cost to maintain the multicast mesh or tree.

A basic flooding mechanism has a high overhead in many multicast scenarios. Intensive flooding can result in a network with several properties that lead to an increase in packet collisions. These properties are collectively referred to as *the broadcast storm problem* in [94]. Smarter flooding protocols avoid this problem to some extent.

In the following we describe SMF [69] which uses optimized cluster-based flooding to reduce the overhead.

Simplified Multicast Forwarding for MANET (SMF) [69]

Simplified Multicast Forwarding (SMF) is one example of a solution that uses flooding to distribute multicast. SMF utilizes clustering to reduce the number of redundant broadcasts that are inherently connected with flooding on common radio channels. The clustering mechanism selects a subset of a node's one-hop neighbours that provide connectivity to all two-hop neighbours of the nodes. This subset of neighbours is then selected to broadcast a flooded message. This mechanism allows all nodes to receive the flooded message, but reduces the total number of transmissions.

SMF can be integrated with unicast protocols that do similar clustering to reduce the number of signalling messages (e.g., the Multipoint Relay (MPR) selection in OLSR [23]), or it uses its own mechanism to perform the relay selection. Thus SFM require some signalling to perform an efficient flood of the multicast messages; however, this signalling is not as expensive as multicast tree maintenance. The signalling creates a framework that reduces the number of redundant packets that are transmitted during the flood of the messages. Clearly, the optimization of the flooding mechanism will also reduce the multicast redundancy and thus its robustness to mobility.

2.3.5 Stateless Multicast

In stateless multicast, the addresses of all members of the multicast group must be described in the header of each multicast packet. Clearly this mechanism does not scale well with increasing group size; however, this protocol type can provide an efficient mechanism to traverse narrow bandwidth links close to the multicast source for small multicast groups.

Stateless multicast depends on the unicast routing protocol; there is in theory no difference between stateless multicast for fixed networks versus for Mobile Ad Hoc Networks (MANETs). The differences in the network behaviour for these two network types are handled by choosing the appropriate unicast protocol and are thus hidden from the multicast mechanism. In reality there will be differences due to different performance of the unicast routing protocols for MANETs versus fixed networks.

We have chosen to describe DDM [55] here because this protocol also provides an optional statefull version to optimize the protocol performance.

Differential Destination Multicast (DDM) [55]

DDM is a receiver initiated stateless multicast protocol. The receivers explicitly join a multicast source/group pair with a unicast join message to the source. Likewise, the receivers explicitly leave the session. The protocol can operate in two modes, a stateless mode or a statefull mode (or a combination thereof). In the stateless mode, the protocol is a pure stateless multicast protocol where the identification of all group members is coded in the header of each packet. Each router that is a next hop receiver of the packet consults its unicast routing table to identify the next-hop router for all the receivers (associated with this next-hop node) listed in the packet header. If the receivers are reached via different next hops, the router regenerates the packet header such that the receivers are listed in blocks, each associated with a unique next hop. The membership information stored by the source is softstate based, requiring the session members to send a new join message to refresh their multicast membership.

In the statefull mode of the protocol, each router stores the multicast destinations that must be served by this router. The subsequent packet header only lists the possible changes in the destinations associated with a specific next-hop router. This version of the protocol requires some local signalling, and it keeps some state information; however, the state information is introduced to optimize the protocol, it is not required for the protocol to work.

2.3.6 Summary

It is impossible to find a MANET multicast protocol that is perfect (little overhead and high efficiency and throughput) for all levels of node mobility, different topologies and traffic patterns. In Figure 5 we provide an approximate representation of the area where the different multicast protocol types are most effective.

Stateless multicast is very efficient for small multicast groups, but does not scale well to larger groups. This protocol is also useful in situations where a multicast source is connected to the rest of the network with a narrow-bandwidth link. The underlying unicast protocol defines the level of mobility that this multicast type can handle.

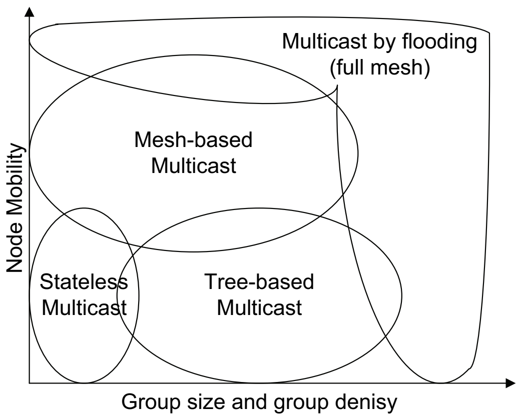

Figure 5: The figure compares the different multicast types with respect to group size/density and the level of mobility.

Tree-based multicast scales well with group size, but shows increasing packet loss with increasing mobility. Mesh-based protocols scale very well with increasing group size and are more robust for scenarios with high mobility, at the cost of some redundant data packets and a high signalling overhead. The hybrid mesh-based and tree-based protocols will typically fit in the area between the mesh-based and tree-based protocols in the graph. Lee, et al. presents a performance comparison of representative multicast protocols from tree-based, mesh-base and hybrid protocol classes in [62]. The mesh-based protocols perform best in most scenarios.

Multicast by means of flooding is most useful for groups with a high multicast density, and for groups with very high mobility. The redundant packet overhead is low for multicast groups with high multicast member densities, and flooding is the only solution that is robust for very high node mobility. Kunz presents a performance comparison of tree-based, mesh-based and flooding protocols in [57], and shows that an optimized flooding protocol performs best in most scenarios. This analysis is done for low multicast bitrates.

The throughput of a flooding protocol will decrease as the traffic load increases. Flooding requires much network capacity (also the optimized flooding mechanisms), thus the rate of packet collisions increases rapidly as the traffic increases (*The broadcast storm problem* [94]).

2.4 Dynamic Wireless Access Networks for UMTS

The standard wireless access network for UMTS is a one-hop radio link on one of the two radio technologies specifically defined for UMTS. A flexible heterogeneous and/or multihop architecture can potentially increase the availability and capacity of the access network.

We expect that the 4th generation (4G) of mobile networks will consist of many different wireless technologies and architectures. This view is also taken by the European IST project Ambient Networks [10, 73]. The main purpose of this project is to create an architecture and mechanism to support efficient interaction between different network technologies and designs. Mechanisms are required to support collaboration on many different levels, from interaction between autonomous networks to multihop heterogeneous wireless access.

When Mobile Ad Hoc Networks (MANETs) are introduced to the wireless access network for 3G and beyond, some challenges associated with MANETs must be acknowledged [81]. One issue is to enforce some level of quality of service (QoS) in a largely unpredictable ad hoc network. Varying channel capacity, a shared medium access and unpredictable network connectivity due to a dynamic mobile network topology all make it difficult to predict channel availability and quality. Another important problem is security [72]. Relay nodes can easily eavesdrop information, delete messages, inject erroneous messages, or impersonate a node. A malicious node that transmits incorrect routing messages can destroy connectivity in the complete MANET.

The mobile nodes' willingness to relay traffic to other nodes is also an open issue. Nodes may refuse to relay packets for other nodes; the reasons for this might include lack of trust, desire to save battery power, or desire to spare the available bandwidth for their own traffic. An important question is how to price the service offered by the auxiliary network providers; an incentive scheme is necessary for nodes relaying messages on behalf of other nodes [91].

Most of the architectures we describe in this overview touch the mentioned issues only briefly. To some extent, these issues have been addressed by studies that focus on integration and feasible business models for multihop and/or heterogeneous network (e.g., [11]). The task of finding solutions to these problems can, to some extent, be eased by coexistence with an infrastructure-based cellular network.

The authors of [20] define the integration of one-hop cellular access networks with heterogeneous wireless access and multihop wireless access in two stages:

- Stage one describes the case where other radio networks are used as access networks for the 3G architecture.

- Stage two describes the evolution of the 3G cellular one-hop radio network into a multihop access network.

Both approaches are expected to augment the capacity and coverage of the 3G RANs (Radio Access Networks). In our work we chose to study a heterogeneous multihop access network for multicast traffic. Thus most of this introduction is concentrated on stage two type architectures. In subsection 2.4.1 we briefly describe architectures that fit the stage one classification. Multihop wireless access networks (stage two) are presented in two subsections: section 2.4.2 for architectures designed for unicast traffic, and section 2.4.3 for architectures designed for multicast traffic.

2.4.1 Heterogeneous Access Networks

In the following we briefly describe some interesting architectures where other network technologies are used as access network for 3G networks.

IEEE 802.11 Wireless Local Area Network (WLAN) hot spots have been integrated with the UMTS architecture, and thus UMTS-enabled hot spots can be used as access networks to the UMTS infrastructure. This integration has been standardised by 3GPP; descriptions can be found in [9, 4].

Another example can be found in [49], where the authors propose the cooperation between a cellular network and a multihop WLAN with access points wired to the cellular infrastructure. The WLAN provides capacity while the cellular network provides coverage.

In the Satellite Digital Multimedia Broadcasting system (SDMB) [86], the broadcast channel is intended to enhance the download capacity of 3G and beyond cellular systems. It is compliant with the 3G MBMS [3] architecture. Terrestrial repeaters can be used to cover blind spots, and the 3G cellular network can be used to provide the uplink channel from the mobile device. The satellite radio interface is identical to the UTRAN WCDMA, except for some unavoidable differences such as the frequency band and timing issues.

The Digital Video Broadcast for Handheld terminals architecture (DVB-H) [31] should also be mentioned in this context. If DVB-H becomes popular and is ubiquitously deployed, the transmitters will most likely often be co-located with cellular base stations. Furthermore, dual mode phones with both cellular and DVB-H transceivers exist, thus, cooperation between the two technologies will be beneficial.

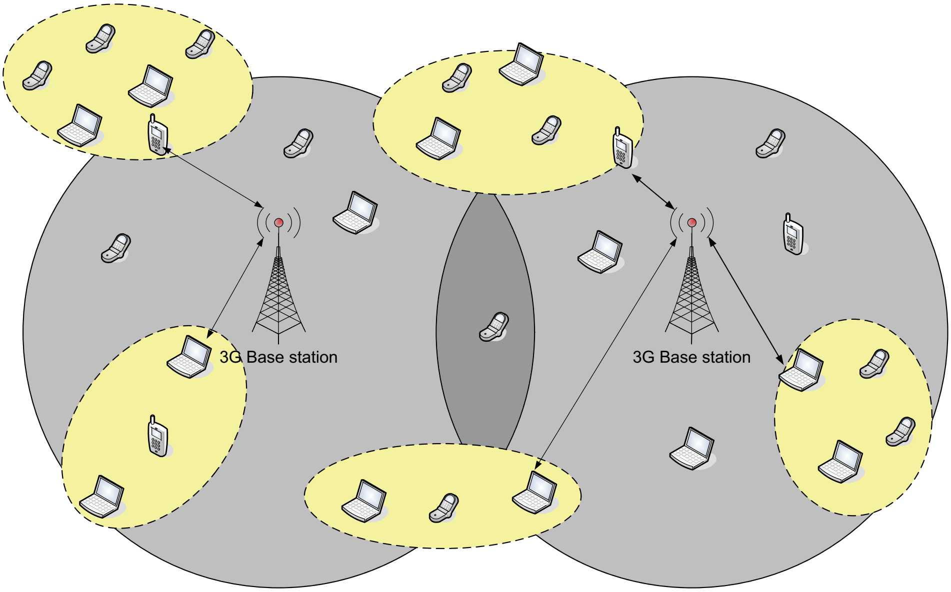

Figure 6: This figure pictures different ways to integrate MANETs with 3G cellular networks. The depicted scenarios represent increased connectivity beyond cellular range, ad hoc networks for load balancing between cells and ad hoc networks to provide high bandwidth to the cell border.

2.4.2 Multihop Wireless Access Networks for Unicast

The multihop wireless access architectures are designed to solve a myriad of different problems with the standard 3G access networks. Some focus on higher total throughput in a 3G cell, while others attempt to increase the range of high-bandwidth channels, yet others are used to increase the coverage of the cellular network, and some relay traffic from one overloaded cell to a lesser loaded neighbour cell. Several of these cases are shown in Figure 6.

Several proposals exist where the same radio interface is used both for the one-hop (cellular) communication and for the multihop (MANET) communication, but most of the recent multihop wireless access network proposals involve 3G networks and MANETs based on some IEEE 802.11 product.

In the following description we provide a snapshot of the architectures suggested for this field. It is not possible to give a complete overview since this is an active research area, where new proposals appear frequently.

Opportunity Driven Multiple Access (ODMA) Figure 84

ODMA was an early architecture that was discussed for 3G by 3GPP. ODMA attempted to provide high-bandwidth channels to terminals located near the cell border, by allowing mobile stations closer to the base station to act as relays for the data-flow. The design was meant for unicast traffic to nodes within signalling coverage of the 3G base station. ODMA is a single transceiver architecture that uses a WCDMA TDD channel for both the cellular and the multihop channels. The standardisation was discontinued due to a high complexity compared with the available gain.

Intelligent Relaying (IR) for Future Personal Communication Systems [38]

Similar to ODMA, IR is also a single transceiver architecture intended for CDMA TDD. The purpose of this design is to provide a connection to the base station for terminals in the shadow from direct base station coverage, and to terminals outside the range of the base station. The architecture also intends to reduce transmission power by using a multihop route with shorter hops than direct mobile to base station communication.

Route selection is performed by the base station. Periodically each terminal that is willing to be relay, reports a list of neighbours. The base station uses this list as well as measured interference, time slot and CDMA spreading factor information to perform route selection.

The architecture is most efficient for lightly loaded networks with sparsely deployed base stations.

The Self-Organizing Packet Radio Ad Hoc Networks with Overlay (SOPRANO) [105]

SOPRANO is also a single transceiver architecture. The authors propose a complete heterogeneous architecture in that sense that most nodes participate in the multihop model. A cell-splitting technique is used, where many nodes have routing functionality. Mobile nodes route their packets towards the router with the minimum path-loss channel, in the direction of the base station. The reverse path is used for packets addressed to the mobile node. Base stations, as well as mobile terminals, reduce their transmission power. The range of the cellular network can also be increased with this architecture.

The small virtual cell centred on the base station becomes the bottleneck in this architecture since all flows must be terminated by the base station. Several mechanisms are suggested to trade increased total bandwidth for the reduced transmission power, e.g., to route traffic to neighbouring base stations to provide load balancing. All routers within three- to four-hop range must be synchronised to efficiently utilize the WCDMA-TDD time slots.

The Multihop Cellular Network (MCN) [64]

Similar to SOPRANO, MCN also performs cell-splitting, and it is assumed that the same wireless channel is used in mobile to mobile communication as in mobile to base station communication. The transmission power is set to a level that balances the number of hops in the multihop path with the number of parallel channels (frequency reuse). A routing protocol is proposed [50], that allows a direct path for inter-cell communication and uses the base station as a gateway for intra-cell communication.

The Unified Cellular and Ad Hoc Network Architecture (UCAN) [67]

UCAN is one of the most complete heterogeneous wireless access architectures. Parts of this work for unicast traffic resemble our work with multicast traffic. It was published at about the same time as we published the first analysis of the multicast architecture.

UCAN includes a mechanism for secure crediting to motivate mobile users to relay traffic to other receivers. The architecture uses dual transceivers terminals, with a 3G transceiver (in this case represented by High Data Rate (HDR): 1xEV-DO (CDMA2000) Figure 14) and an IEEE 802.11b/g [52] transceiver. The main purpose of the design is to improve the throughput when the standard 3G channel experiences poor channel conditions. The service is intended to be used for unicast traffic to nodes within signalling range of the 3G network.

The scheme identifies gateway clients (called proxies in [67]) that serve as the gateway between the 3G transmission and an ad hoc multihop path. A proactive proxy discovery protocol is proposed as well as a reactive discovery protocol. Nodes that experience bad channel quality use one of these MANET protocols to discover a gateway with good channel conditions within the range of a given maximum number of hops. The multihop path to the gateway is then used for the downlink transmission.

The authors report the possibility for a high increase in total throughput. However it must be noted that this performance is based on the assumption that the MANET has sufficient capacity.

The Integrated Cellular and Ad hoc Relaying System (iCAR) [100]

iCAR is also a fairly complete architecture that is analysed based on line-switched voice-calls; however, the ideas are applicable also for packet-switched data traffic. In this architecture, special Ad-hoc Relay Stations (ARSs) are deployed in the cellular network; these are small movable (but not mobile) units. These devices all have a connection with the cellular base station, and they can connect to other ARSs within range using a WLAN transceiver (e.g. IEEE 802.11). Mobile devices can connect to an ARS on a WLAN interface or to the base station on a cellular interface. The main purpose of the ARS is to relay unicast traffic from a congested cell to a nearby cell with available capacity; however, the ARS can also be used to increase the coverage of the heterogeneous radio access network. The cellular network executes a central controlled routing protocol for the ARSs. The ARSs inform the cellular network of their ARS neighbours, and based on the neighbour information, the cellular network calculates the ARSs' routing table for the multihop paths to neighbouring base stations.

Whenever the cellular network detects that a mobile terminal needs a relying route, it informs the terminal of neighbouring base stations with available capacity. Next the mobile terminal queries the neighbouring ARSs for information of their routes to any of these base stations, and the best route is chosen. This architecture also allows mobile terminals to establish a connection to other terminals via an ARS link (or multihop link), not involving a base station.

This architecture with dedicated relays is more predictable and reliable than the MANET equivalent, on the other hand it is more costly and less dynamic. The performance (in this case represented by reduced call blocking probability) is improved with the iCAR architecture, but the improvement is naturally highly dependant on the number of ARSs and their positions.

The Mobile-Assisted Data Forwarding for Wireless Data Networks (MADF) [101]

Similar to the iCAR architecture, MADF attempts to balance the load between adjacent cells. Different from iCAR, MADF uses mobile terminals as relay. The MADF architecture can use either in-band relay channels, where the same transceiver is used for relaying as for communication with the base stations, or an out-of-band channels (e.g., IEEE 802.11) can be used for the relay traffic. MADF is able to relay traffic from terminals outside the range of the cellular base station.

The Two-Hop-Relay Architecture [97]

Another approach is taken in this architecture. As the name implies, the wireless path is restricted to a maximum of two hops from the mobile terminal to the cellular base station. This is also a dual transceiver heterogeneous architecture where some WLAN transceiver (e.g., IEEE 802.11) is used in addition to the 3G transceiver. The relay can either be a dedicated relay as in the iCAR proposal (description on p. 32), or it can be a mobile terminal that has reported its willingness to be relay. The architecture is mainly intended to increase the capacity of unicast traffic on the cellular system, but can also be used to increase the coverage of the system, and allow terminals with a single WLAN transceiver to connect to the cellular system via the dual transceiver relay. Thus, the system can be thought of as a heterogeneous radio access network extension to the integration of WLAN hot spots, as an alternative wireless access network to the cellular system [9].

The authors chose the two-hop limitation on the wireless path lengths as a trade-off for the routing complexity and unreliably involved with larger MANETs. In Two-Hop-Relay there is no MANET routing. The relay gateway periodically announces its presence on the WLAN channel, and this message also includes the capacity of the relay gateway's cellular channel, and the ID of the associated base station. The relay nodes use a terminal mobility protocol like Mobile IP [78] (slightly modified) to support the terminals that wish to use the two-hop-relay path.

Ad Hoc Routing for Cellular Coverage Extension (ARCE) [36]

ARCE describes an architecture and routing protocol for a cellular network that is assisted by MANETs to increase cellular coverage. The architecture is intended only for coverage extension, and does not support peer-to-peer traffic.

A MANET protocol collects connectivity and path information. This information is used by the base station to calculate one or more possible MANET paths from a mobile station, to terminals that are located inside cellular coverage. The MANET source receives the path information from the base station and uses a source routing mechanism to forward data to a gateway node within cellular coverage.

Communication with terminals outside cellular range must be initiated by the terminal. outside coverage. This terminal can usually find a path to a node inside cellular coverage with a limited flooded query. The network on the other hand, must initiate a flooded query from many nodes associated with many or all base stations to find a path to a terminal outside range. Consequently, the architecture requires that the mobile terminal initiates all communication.

The Cellular Aided Mobile Ad Hoc Network (CAMA) [15]

CAMA uses the cellular network to improve ad hoc network performance. The other proposals we have described all use MANETs as a tool to improve the 3G network performance.

CAMA uses the cellular network for out-of-band signalling to a CAMA agent that is co-located with the cellular infrastructure. The CAMA agent assists with routing and security. It is also envisioned that the CAMA MANETS can off-load the cellular network with some high-bandwidth multimedia traffic for the cases where a MANET node is connected to the Internet.

2.4.3 Multihop Wireless Access Networks for Multicast

To our knowledge, we were the first to suggest and analyse a heterogeneous multihop access network for multicast traffic [44, 40, 41]. We believed the possible capacity gain associated with a multihop heterogeneous network could be potentially much better for multicast traffic than for unicast traffic, due to the signalling overhead associated with multihop heterogeneous wireless networking.

Later (at least) two more heterogeneous multicast architectures has bee published. These two are described below.

Enhancing Cellular Multicast Performance Using Ad Hoc Network [75]
In [75] the authors propose a scheme where local IEEE 802.11-based MANETs are introduced to solve the receiver heterogeneity problem for cellular multicast transmission. The work is motivated by the UCAN architecture (description on p. 32). In this case the heterogeneous architecture is intended for the situation where a cellular broadcast channel distributes a service with a given bitrate. For high-bandwidth services it is likely that some receivers experience bad channel conditions and suffer a high packet loss (if they receive any packets at all). These receivers establish a MANET path to a receiver (proxy) with a better channel condition. The setup is done on a receiver by receiver basis, thus the MANET is used in unicast mode. The cellular network performs the MANET routing. A node with bad channel conditions informs the cellular network, which in turn uses the channel conditions of the neighbouring nodes and their 802.11 bandwidth to identify the best proxy (within three hops of the receiver). The 802.11 channel interference associated with paths that are within radio range of each other is used as a parameter in the path calculation. The proposal does not explain how the cellular network knows the 802.11 connectivity and position of the mobile terminals.

The Integrated Cellular and Ad Hoc Multicast (ICAM) [16]
This proposal is an extension to the UCAN architecture (description on p. 32) for unicast traffic. It is an interesting proposal where the channel interference in the 802.11 MAC protocol is

taken into account in a different manner than in the above proposal. A node that experiences a deteriorating wireless channel condition, starts to search for a gateway (proxy) on the 802.11 channel. Terminals with better channel conditions include their address and available cellular bandwidth in the search packet, and forward the packet. The terminal also notifies the base station of the query. After a short timeout, the base station will have received one or many path messages from the potential gateways, and have a picture of the partial node topology around the receiver with bad cellular channel condition. The base station then chooses a gateway and a path to the receiver; the cellular bandwidth of the gateway is part of the path calculation. Paths that already exist to neighbouring receivers are favoured (for multicast transmission on the ad hoc network). Lastly the routing algorithm attempts to setup MANET paths that do not interfere with each other. Thus the gateway with the best channel condition might not be chosen if another gateway provides a path that does not interfere with other existing MANET paths to members of the multicast group. This fairly complicated algorithm attempts to reduce the interference on the MANET, and thus assure that it is the cellular network that sets the bandwidth limitation and not the assisting ad hoc network.

2.4.4 Summary

We have given a fairly complete snapshot of heterogeneous wireless access network proposals intended for 3G and beyond. It is difficult to compare these architectures since they are all designed to solve different subset of a larger set of challenges associated with standard 3G wireless access.

Typically, the proposals with a single transceiver are based on WCDMA TDD. These are complex (i.e., require tight time synchronisation and detailed resource management), but will provide a somewhat predictable service. The proposals with a 3G transceiver and an IEEE 802.11 transceiver are much simpler but also more unpredictable. The proposals that require 3G signalling connection to all mobile terminals can improve the MANET routing with some infrastructure support, while the proposals that attempt to extend the range of the 3G service must rely on distributed MANET routing. For the time being, the dual transceiver proposals are not able to support reliable transmission, and can therefore only be used for best effort services.

3GPP's standardisation of the single transceiver architecture, ODMA (description on p. 30), was discontinued, partly due to complexity reasons. We believe the standardisation efforts of multihop wireless access networks for 3G and beyond, will be continued with the much simpler and easy-to-deploy solutions using dual transceivers in two disjoint frequency bands.

2.5 Forward Error Correction for Wireless Multicast

One of the methods for improving the quality of a wireless connection is Forward Error Correcting (FEC) codes. FEC codes can be used to correct bit errors in a data packet, or as an erasure code where lost packets are reconstructed. With FEC erasure codes (e.g. Reed-Solomon [98]) some redundancy is added to the data stream at the source, allowing the receiver to reconstruct a certain number of lost packets. The same redundant packets allow different receivers to reconstruct different sets of lost packets. This property makes the scheme very well suited for broadcast and multicast traffic.

As pictured in Figure 7, the input to a FEC encoder is some number *k* of equal length source symbols. The FEC encoder generates some number *n* - *k* of parity symbols, that are of the same length as the source symbols, and these parity symbols are placed into packets for transmission. The number of parity symbols placed into each packet can vary on a per packet basis, or a fixed number of symbols (often one) can be placed into each packet. Also, each packet header contains sufficient information to identify the particular code symbols in the payload of that packet. For a *systematic* FEC code, the source packets are preserved and all generated parity symbols are placed in the additional redundant packets. All well known FEC erasures

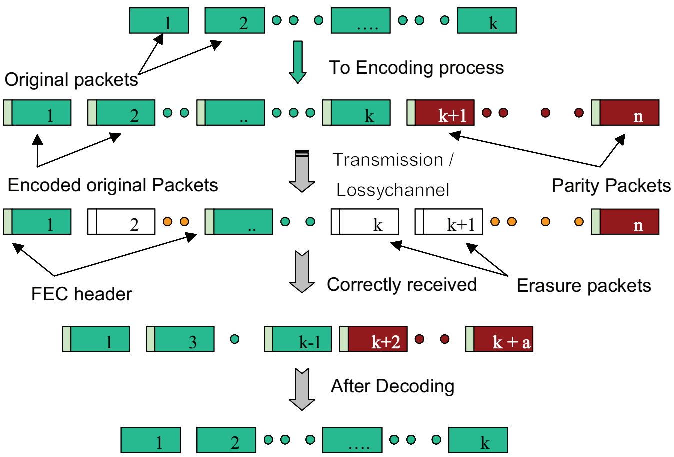

Figure 7: The figure (ref. [22]) shows the encoding and decoding process for an ideal systematic FEC code.

codes are *systematic*.

The data stream is partitioned into source blocks consisting of *k* elements, and the redundant symbols are used to protect each source block from lost packets. An *ideal* FEC code (e.g., Reed-Solomon [98]) can reconstruct an exact copy of the *k* source packets from any *k* of the *n* code packets.

Many reliable multicast protocols intended for unreliable fixed networks incorporate FEC coding. A framework for efficient use of FEC for multicast is specified by IETF [65, 66]. A FEC mechanism will also be included for broadcast and multicast transmission in several infrastructure-based one-hop wireless media, such as the Multimedia Broadcast/Multicast Service (MBMS) [3, 77] standardised for 3G-UMTS networks, the European Digital Video Broadcasting (DVB) standard for broadcasting to mobile handheld terminals (DVB-H) [29], and the CDMA2000 High Data Rate (HDR) Broadcast/Multicast Services (BCMCS) [8]. FEC is also part of the reliable mechanism in the Reliable Multicast data Distribution Protocols (RMDP) [83].

Both the MBMS and DVB-H standards use the very flexible, but not fully *ideal* Raptor codes [88]. The Raptor codes are very dynamic in the manner that the level of FEC protection (redundancy) is flexible and can be adjusted on demand, based on e.g., feedback of the current wireless channel condition. The algorithm's decoding complexity is also much lower than for the very well known Reed-Solomon codes. Experimentation with FEC codes on these type of wireless systems report encouraging results (e.g., [1]).

It is not evident that utilization of FEC codes will be successful in a multihop wireless network, since the resilience available with FEC is available at the sacrifice of wireless bandwidth. A multihop wireless network with a common channel (and no central scheduling control) is extremely sensitive to network load and congestion.

The use of packet level FEC on MANETs has not been fully analysed. The reason for this might be the mentioned properties which indicates that FEC protection will not be effective on MANETs. A similar concern is stated in [56]. The author claims that FEC-based multicast protocols for ad hoc networks are not practical because the protocols need to know the worst case packet loss to generate enough redundant data (to provide a fully reliable transmission). Thus, the FEC-based protocols increase the network traffic, also when the loss-rate is low. For the case of a fully reliable channel, these arguments are certainly correct; however, the situation where FEC is used to *improve* the transmission is a different one.

An interesting Robust Multicast Routing in Mobile Ad hoc Networks (RoMR) scheme is proposed in [68]. In this proposal, *k* multicast packets are encoded as *n* packets with a FEC algorithm. Furthermore, *n* independent multicast distribution trees are established. Each of the *n* encoded multicast packets are assigned to one of the *n* multicast trees, and transmitted on this tree. When the multicast group members receive *k* of the *n* packets, the data block can be decoded. Simulation results show that this protocol provides a robust multicast delivery compared with the chosen basic multicast implementation (MOLSR [60]). However, it is not clear from the results how important the FEC encoding is for the improved delivery.

As part of our work we studied the efficiency of FEC protection for multicast on a heterogeneous Mobile Ad Hoc Network (MANET) based on a 3G cellular link and IEEE 802.11 channels. This study is reported in [42].

3 Methods, Contribution and Discussion

3.1 Methods and Assumptions

A wireless heterogeneous data network is extremely complex and do not lend itself well to theoretical analysis. Instead packet-level, event-driven simulation studies are usually carried out to study the performance of network components, algorithms and protocols, and their interaction. If theoretical results exist, these are often used to define the upper or lower performance bound for the studied system and thereby set a goal towards which a system, protocol, etc. can be compared.

A network simulator attempts to approximate the behaviour of an actual network, but under controlled forms where the result of different parameters can be studied. A network simulator incorporates different levels of abstraction to approximate the actual system. Thus there are large differences between different simulators. This is clearly shown in [21]. Naturally the accuracy of the results from a simulation is tightly coupled with the level of abstraction and complexity of the simulator. This is why network simulation results should be used to indicate trends for a system rather than to give exact results. Simulation results can be reproduced and are therefore very useful for the research community for comparison with other proposals. Simulation results is by far the most used research method in the field of wireless communication [58].

The authors of [58] also point out the most common pitfalls of network simulation. Often the simulation results are not statistically sound, or the researchers fail to show the important parameters in their simulation such that it is impossible to understand if their conclusions are valid. If the researcher is aware of the limitations associated with network simulation, then simulation is a very useful research tool.

We wrote a simple static simulator in Java© to roughly identify performance trends for two of the mechanisms we analysed in our work. This is a static simulator without any mobility or network protocol functionality. The simulator allowed us to capture a sequence of snapshots of the network topology for different network sizes and thus provided an environment to calculate e.g., transmission coverage, network load, ad hoc connectivity, path lengths in number of hops, etc. for a range of topologies.

We chose the network simulator J-Sim [53] for detailed simulation of the heterogeneous wireless network architecture. A detailed description of the J-Sim simulation environment can be found in [95]. We have carefully tried to avoid the pitfalls associated with the network simulation as described in [58] to ascertain the quality of our simulation results.

A series of open source network simulators with support for wireless ad hoc networking are available (e.g., ns2 [92], GloMoSim [35], J-Sim [53], OMNeT++ [74]). We chose J-Sim for the following reasons: It was a Java©-based simulator that provided flexibility in the sense that we could run the simulator on a variety of available platforms and operating systems. The component-based architecture provided a flexible and well arranged set of building blocks to model the heterogeneous network. The component structure also eased the introduction of a new component (in our case, a multicast routing protocol) to the simulator. Additionally, the authors of the simulator claimed that the environment performed well for simulation of large networks with many nodes.

The main drawbacks of the J-Sim simulator was that the wireless extension to the simulator was fairly new, thus there was a high probability for latent bugs. Additionally, little work was reported on this simulation environment yet. However, since we were the first (to our knowledge) to suggest a multihop heterogeneous architecture to reduce the multicast load on UMTS wireless channels, the architecture did not lend itself to easy comparison with other proposals.

Ideally, real test-bed experiments should be performed to validate the simulation results. However, for the work presented in this thesis we neither had enough time, nor the resources available to implement and perform high quality test-bed experiments.

3.2 Contribution

As stated in the introduction, the main focus of our work has been to find ways to reduce the resources required by the Universal Mobile Telecommunications System (UMTS) [2] to deliver high-bandwidth group services; and thereby increase the availability of such services to mobile users. In the following we present our main contributions and briefly describe the chosen research methods for each study.

3.2.1 IP Multicast Architecture for MBMS (Paper A)

The main contribution of this work is a study of how the commonly used IP multicast mechanisms and protocols could be utilized in the Multimedia Broadcast/Multicast Service (MBMS) [3, 104] architecture. The work is reported in (Paper A) [43].

In this study we suggested two multicast architectures, one that used IP multicast protocols in the UMTS core network, and unicast in UTRAN, and one that used IP multicast routing throughout the UMTS network (excluding the radio cell). We analysed these proposals with respect to the following issues:

- Minimized network resource consumption
- Multicast group management
- Data privacy and integrity
- Charging mechanisms
- Sender and receiver mobility handling

The analysis was performed on a conceptual level involving UMTS mechanisms, but without going into details of the UMTS protocols. Our work showed that a multicast architecture based on IP multicast mechanisms could be a viable solution for the evolving MBMS. IP multicast protocols in MBMS context is to a large extend able to solve the multicast deployment problem issues raised in [25].

The study included a simple spread-sheet calculation of the protocol efficiency for two different multicast service types. Figure 8 show how an architecture with IP multicast protocols can reduce the UMTS network traffic for two group-service types (details of the studied scenarios can be found in Paper A).

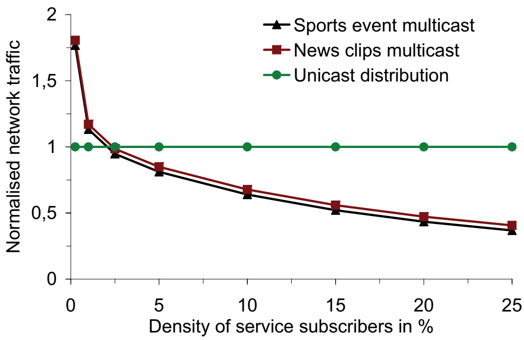

Figure 8: Network traffic generated by multicast distribution compared to unicast distribution.

We concluded that the IP protocols were able to support most of the requirements for UMTS multicast. Introduction of IP multicast would need some modifications to existing UMTS signalling (especially for the case where multicast routing is used both in the core network and in the radio network), thus the multicast mechanisms would be best provided, integrated with the UMTS-specific protocols.

In the years following this work, 3GPP has finalized the MBMS architecture for UMTS Release-6. 3GPP chose to reuse much of the IP mechanisms, but not the protocols. All multicast support is merged with existing UMTS protocols or introduced in new UMTS protocols created explicitly for the MBMS architecture.

3.2.2 A Wireless Channel Type for Multicast in MBMS (Paper B)

The main contribution of the work described in this subsection, is an analysis of a broadcast channel type for the last hop in a UMTS network; the sticky-channel. We also compared the sticky-channel with unicast and normal broadcast for multicast distribution in the radio cell for sparsely populated multicast groups. The work is presented in (Paper B) [45].

The sticky-channel exploited coverage redundancy given by radio planning to reduce the total number of broadcast channels needed to cover a give area. This channel would stick to terminals in neighbouring cells and thereby make the broadcast channel in some of the neighbouring cells superfluous.

The wireless resource analysis was done based on the formulas for cell capacity and transmission power presented in [90]. We wanted to study the three different channel types (unicast, normal broadcast and sticky-channel) for a large network, thus we wrote a simple simulator in Java© that allowed us to calculate coverage and network load for the different channel types on a variety of different topologies. The simulator code (version 1.0) is available at UniK - University Graduate Center, and can be distributed upon request.

Figure 9 shows how often the sticky-channel is utilized in one of the studied scenarios. All multicast terminals in the light red sectors are served by a sticky-channel from a dark red neighbour sector (more information about the scenarios can be found in Paper B.

Even though the sticky-channel is applicable for many of the sectors in a large network (ref. Figure 9), the load analysis showed only a small reduction in the average load placed on a sector by a multicast service. The sticky-channel is not as useful as expected due to the high interference cost associated with the increased range of the sticky-channels.

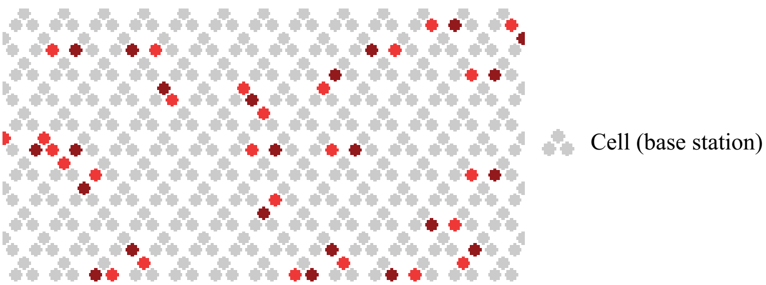

Figure 9: This figure pictures a large cellular network. A cell (base station) is represented with three bullets (sectors). The dark-red bullets portray sectors using sticky point-to-multipoint channels. Red bullets represent sectors covered by a neighbour's sticky-channel. Light grey bullets represent no transmission, unicast and normal multicast

The Forward Access Channel (FACH) was chosen as the broadcast channel for the MBMS architecture. FACH uses mechanisms that do the complete opposite of what our sticky-channel did; it uses soft-combining and selective-combining (description on p. 15) to improve the reliability of the data (and consequently allow a reduction in the transmission power for the broadcast channel). FACH may use any redundant broadcast channel from neighbouring base stations to improve the channel reception, whereas we attempted to remove the multicast transmission from some base stations by increasing the power from the neighbouring base station, and let this channel stick to multicast members in the neighbouring cell. The MBMS solution is clearly more flexible and efficient than our suggested channel type, but it also introduces more complexity in the transceivers and requires tight synchronization between neighbouring base stations.

3.2.3 The Heterogeneous Wireless Network Architecture for Multicast Traffic (Paper C, Paper D, Paper E and Research Report G)

The main part of our work studied heterogeneous wireless access networking for multicast traffic. To our knowledge, we were the first to suggest such multihop architecture to improve the availability of cellular multicast distribution.

In Paper C we present a conceptual heterogeneous network architecture. The architecture allows a reduction of the radio resources required in the 3G cellular network (UMTS) by reducing the range of a 3G broadcast channel. Local IEEE 802.11-based ad hoc networks are introduced to forward the data onto users located outside the 3G broadcast range. A short range UTRAN broadcast channel is transmitted with low power and thus gives less interference for other channels in the same and neighbouring cell. Therefore, a short range channel allows space for more parallel channels compared with a long-range channel of the same bitrate.

For the study reported in Paper C, we extended the core of the Java© simulator, written for the sticky-channel study (Paper B), to analyse the connectivity for the proposed heterogeneous network architecture for a variety of different topologies and transmission ranges (both for the cellular part of the network and for the MANET part of the network). The simulator code (version 2.0) is available at UniK - University Graduate Center, and can be distributed upon request.

Due to an increasing focus on group services from the research community and the service

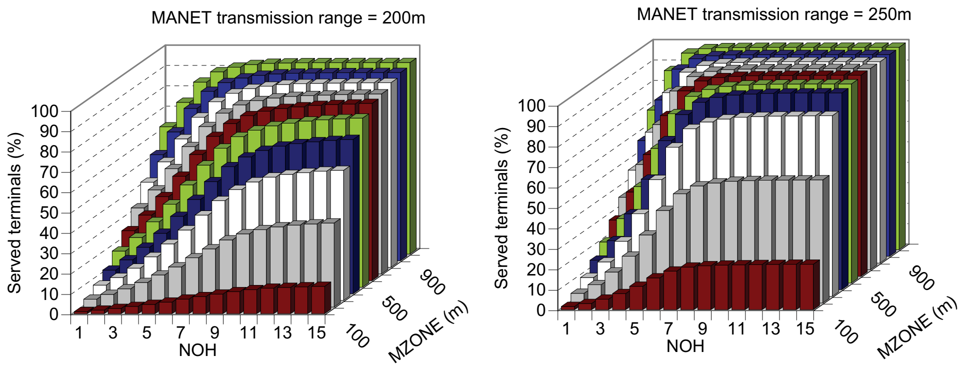

Figure 10: The figures show the trade-off between the radius of the cellular broadcast channel (MZONE), the maximum path length (in number of hops, NOH) between a multicast member and the gateway node within cellular coverage, and the required coverage (served terminals in%). This is shown for MANET transmission rages of 200m and 250m.

providers, the scenarios we studied for the heterogeneous access architecture involve larger groups and a higher density of multicast members than the scenarios in our previous work.

The analysis presented in Paper C identified the most important trade-offs associated with efficient establishment of a connected heterogeneous multihop access network for multicast traffic. The main trade-off is linked to the following four parameters: The transmission range of the MANET channel, the radius of the cellular broadcast channel (MZONE), the maximum number of hops allowed in the MANET paths, and the required service coverage (in percent of complete coverage). In this study we calculated the optimal network for each snapshot of the network topology. We did not consider the effect of mobility and the inability of a routing protocol to find the optimal paths. Figure 10 pictures the mentioned trade-off for MANET transmission ranges of 200m and 250m. The simulation is performed on large cells (radius = 1500m) and with 300 mobile terminals in each cell (details of the simulation setup can be found in Paper C). We concluded from the simulation that formation of ad hoc networks to support cellular multicast communication could be feasible in many scenarios, e.g., for MANET transmission range of 250m, approximately 95% of the multicast members are covered with a cellular broadcast channel covering only 500m and support by MANETs with path lengths of maximum 6 hops (ref. Figure 10).

In Paper D [40] we present two multicast routing protocols designed for the heterogeneous wireless access network. We provided two protocols (one fully centralized, and one distributed with some central support) to prepare for a performance comparison of the two different protocol types on the heterogeneous network. The distributed protocol is based on mechanisms from the leading MANET protocols. In addition we have introduced some centralized support.

We did a simple spread-sheet calculation of estimated overheads on both channel types for the two proposed protocols. As expected the calculation showed that most of the signalling traffic is sent on the cellular channel for the centralised protocol, and on the MANET channels for the distributed protocol.

A thorough network simulation studying the performance of the wireless network architecture and the distributed multicast routing protocol (updated and described in details in Research Report G [39]) is presented in Paper E [41]. The purpose of the heterogeneous access network was to reduce the range of a UMTS broadcast channel to a minimum and use ad hoc networks to cover the cell area outside the range of the broadcast channel, while maintaining a fair service quality. However, a large MANET with long multihop paths will have a higher packet loss ratio, larger jitter and delay than a smaller network.

We implement the distributed multicast protocol in the J-Sim simulation environment [53], and thus modelled our network in this environment. Details about the simulation parameters are given in the “Simulation Setup and Results” section of Paper E. The multicast protocol code, simulation script and analysis tools bundled with a few modifications to J-Sim 1.3 core functionality is available at UniK - University Graduate Center, and can be distributed upon request.

In the network analysis we look at average throughput, throughput fairness, packet loss characteristics, the multicast signalling overhead, and the amount of resources required of passive relays. We studied these parameters to find the trade-off between the 3G coverage (and consequently the required wireless resources in UTRAN) and the MANET size (and thus the service quality). The evaluation is complemented by a sensitivity analysis of factors that could affect the efficiency of the trade-off such as node mobility, different traffic patterns, and multicast member densities. An important part of the work was to compare the service quality in the heterogeneous multihop network with the expected quality of the reliable one-hop cellular channel.

Figure 11 shows the number of received packets in percent of sent packets for 4 different cellular broadcast ranges. MZONE1 is the shortest were the reduced broadcast range is only 20% of the full cell range (MZONE2 = 33%, MZONE3 = 47% and MZONE4 = 60%). The heterogeneous wireless architecture allowed a reduction of the UMTS broadcast channel to less than 50% of the cell range (MZONE3), for a challenging scenario with large radio cells,

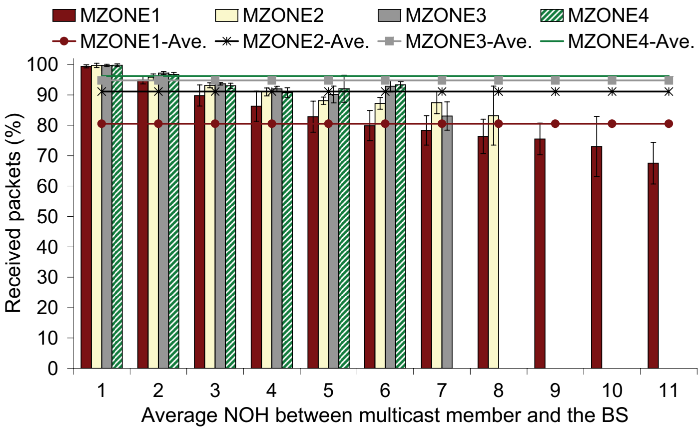

Figure 11: The figure shows the distribution of total received packets (%) as a function of the path length. The horizontal lines shows average received packets (%) for all terminals.

while supporting a fair service quality. For this case the average number of received packets are more than 95% whereas received packets for the worst case MANET size (NOH) are approximately 85% (ref. Figure 11).

This reduction in the UMTS broadcast range free significant resources, thus network operators will be able to support several concurrent best-effort type group services in situations where a conventional UMTS network would not have enough available radio resources.

These results are based on the assumption that the MANET part of the heterogeneous network has sufficient capacity to support the multicast bandwidth provided by one or more UMTS channels. Thus until better QoS mechanisms are available for MANETs, such architecture must be used with admission control and support only best effort traffic.

Throughout our work with the network simulation of the heterogeneous wireless access network we made several improvements to the distributed routing protocol described in Paper D. To document these improvements, we have described the final protocol in Research Report G [39]. This report describes the operation of the protocol as well as the message formats, and suggests default values for the protocol parameters. Thus the report provides enough information to do a correct implementation of the distributed multicast protocol with some central support.

3.2.4 Multicast with Forward Error Correction for the Heterogeneous Wireless Network Architecture (Paper F)

In our last work we studied the effect of Forward Error Correction (FEC) at the packet-level for multicast traffic on the heterogeneous wireless network architecture. The main contribution of this work is to provide a better understanding of the use of FEC for multicast in MANETs (Paper F) [42].

We performed the study on the heterogeneous architecture described in the previous work (Paper E) [41]. We also used the J-Sim simulation environment prepared for Paper E. Details about the simulation parameters are given in the "Simulation Setup and Results" section of Paper F. The multicast protocol code, simulation script and analysis tools bundled with a few modifications to J-Sim 1.3 core functionality is available at UniK - University Graduate Center, and can be distributed upon request.

The UMTS link was assumed to be reliable and available for the total simulation time, thus the observed results were all a consequence of the MANET behaviour.

We observed the following: FEC improves the multicast throughput for unsaturated MANETs; however, the improvement comes at a high bandwidth-cost. E.g., from Figure 12 we see that 20% FEC overhead improves the throughput from ca. 93.2% received packets to ca 95.7% whereas 40% FEC improves the throughput from ca 93.6% to 98.%. The main reason for the low utilization of the FEC resilience can be deducted from Figure 13. The figure shows that the packet loss in the MANET increases with more than 50% when 20% FEC overhead is added to the dataflow. Thus much of the protection available with the FEC code was used to repair the extra packet loss introduced with the FEC overhead. The increased bandwidth intensified

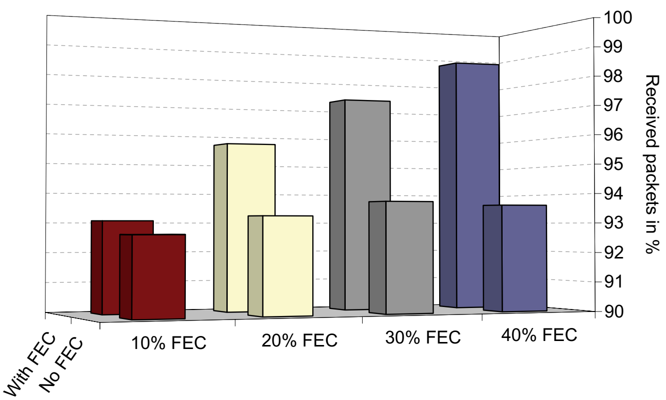

Figure 12: The figure shows the reliability of a multicast flow with and without FEC, as a function of the FEC overhead.

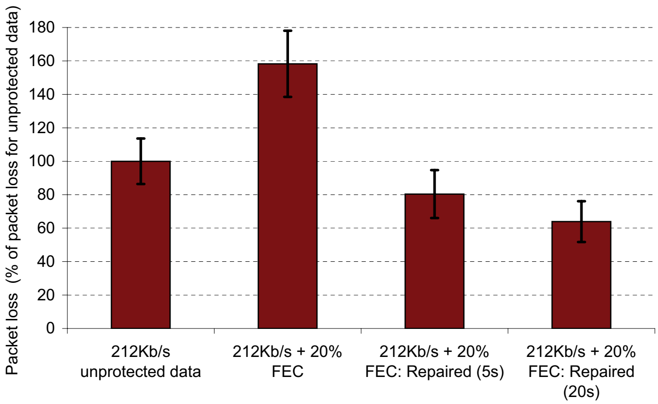

Figure 13: The figure compares the packet loss for data + 20% FEC before and after repair, with the packet loss for the same unprotected data stream. FEC repair based on two FEC encoding block sizes are shown: block sizes equivalent to 5s and 20s of the CBT stream.

the packet loss frequency in the MANET. A multihop mobile network with a common channel is extremely sensitive to network load and congestion because of the hidden terminal problem. In a multihop ad hoc network, the available payload bandwidth is typically about 1/4 to 1/7 of the given maximum link capacity of the common channel [63]. In other words, the FEC overhead on a MANET link will put 4-7 times that load on the common MANET channel.

The simple FEC scheme we studied here can be useful when the MANET bandwidth is abundant, but should not be used for ad hoc networks with high load.

3.3 Discussion

3.3.1 Future Work

For the heterogeneous wireless access architecture there are several open issues that we would like to pursue further. We suggested two routing protocols for this architecture, one distributed and one centralized (Paper D). As part of the work presented here, we have done a detailed study of the distributed version (Paper E). We would like to make a similar study of the centralized version and compare the performance of the two routing protocols in future works.

In our work with ad hoc multicast routing we observed a high number of parallel paths in the multicast distribution trees that was not disjoint, thus we would like to improve the multicast routing for the distributed protocol, and we think the concepts from the multicast routing algorithm with swarm intelligence (MANSI) [87] could be useful in the heterogeneous multicast protocol.

In our most recent work we studied forward error corrected (FEC) multicast flows on the heterogeneous wireless network architecture (Paper F). In this work we did not utilize the dynamics available with some FEC codes. The dynamic codes allow different FEC overheads for each block size, thus the FEC encoding can be associated with the current radio link quality. We would like to pursue the study of multicast with FEC on the heterogeneous architecture further to analyse a situation where FEC with variable overhead is used based on feedback of the current channel quality.

3.3.2 Conclusion

All our work is done as a proof of concept with the level of detail needed for such studies. Consequently many aspects and details that must be available for possible interworking with UMTS are left out. Issues such as Authentication, Authorization and Accounting (AAA) mechanisms, reliability, integration with the UMTS-specific protocols and interworking between IEEE 802.11 transmission and UTRAN transmission are all mechanisms that are required if any of our analysis and suggested mechanism are to become integrated with 3G and beyond networks.

Some of the work is finalized in that sense that the analysis was done for the initial phase of the Multimedia Broadcast/Multicast Service (MBMS) [3, 104] architecture for Universal Mobile Telecommunications System (UMTS) [2]. 3GPP has made its choice for the foundations of the MBMS architecture. An evolution of the basics for this architecture must be continued based on the design specified in UMTS Release-6. Our work associated with the MBMS platform has been compared with 3GPP's choice in Section 3.2.1 for the IP multicast mechanisms for UMTS (Paper A), and in Section 3.2.2 for the sticky-channel option for transmission in the UMTS radio cell (Paper B).

Our work with the heterogeneous wireless architecture is intended for the evolving MBMS architecture. In Release-6, 3GPP standardised the way in which an IEEE 802.11 hot spots can be used as an access network to the UMTS network [9, 4]. There is no standard as of yet, that describes the next step where several wireless technologies cooperate in a multihop fashion to provide the wireless access to the UMTS network. An attempt called Opportunity Driven Multiple Access (ODMA) [84] was introduces for standardisation several years ago, but was stopped (partly due to complexity reasons). This proposal was intended for the TDD-CDMA networks where the same transceiver was used both for communication with the base station and in the multihop path. We believe a heterogeneous wireless architecture based on two radio technologies (much simpler than ODMA) for a greedy service type where a high gain is possible (e.g., multicast) similar to our design (Paper E and Paper F) could be an input to a new standardisation effort for this type of networking.

The heterogeneous wireless access network uses multihop wireless paths. An important challenge inherently associated with such networks is the question of how to motivate fellow mobile users to provide some of their limited bandwidth, battery and processing capacity to relay data traffic to other users. Mechanism to motivate and, recompense the mobile users that allow relaying must be decided by the network and service providers.

To ease the motivation of the mobile users, our architecture only use relays in the IEEE 802.11 MANET, all mobile nodes connected to the UMTS broadcast channel are multicast members. Our architecture also shows an evenly distributed, and fairly low load on the MANET relays. We believe that most users connected to the UMTS network will not, at the same time, be using their 802.11 interface for heavy data traffic Therefore it will most likely be possible to motivate many users to provide some MANET relay capacity for fellow mobile nodes. Furthermore, as more people install equipment in their home, with dedicated mobile phones (e.g., burglary alarms and weather stations) available MANET capacity in these redundant phones might also be made available for relaying traffic for others.

References

[1] 3GPP, "Efficient FEC code for reliable MBMS user services," TSGS#27, Tokyo, Japan, (05)0164, 14-17 March, 2005, http://www.3GPP.org.

[2] 3GPP specifications of Universal Mobile Telecommunications System (UMTS) available at: http://www.3gpp.org.

[3] 3GPP, "TS 22.246: Multimedia Broadcast/Multicast Service (MBMS) user services; Stage 1," V6.3.0, 2006, http://www.3gpp.org.

[4] 3GPP, "TS 23.234: 3GPP system to Wireless Local Area Network (WLAN) Interworking; System Description," V6.10.0, 2006, http://www.3gpp.org.

[5] 3GPP, "TS 25.346: Introduction of the Multimedia Broadcast Multicast Service (MBMS) in the Radio Access Network," V6.9.0, 2006, http://www.3gpp.org.

[6] 3GPP, "TS 25.401: UTRAN Overall Description," V7.1.0, 2006, http://www.3gpp.org.

[7] A. Adams, J. Nicholas and W. Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)," *RFC3973*, March 2006, http://www.ietf.org.

[8] P. Agashe, R. Rezaiifar and P. Bender, "CDMA2000® High Rate Broadcast Packet Data Air Interface Design," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 83-89, 2004.

[9] K. Ahmavaara, H. Haverinen and R. Pichna, "Interworking Architecture Between 3GPP and WLAN Systems," *IEEE Commun. Mag.*, vol. 41, no. 11, pp. 74-81, 2003.

[10] Ambient Networks, IST project 507134, http://www.ambient-networks.org.

[11] I. Armuelles, T. Robles, H. Chaouchi, I. Ganchev, M. O'droma and M. Siebert, "On Ad Hoc Networks in the 4G Integration Process," *in Proc. Med-Hoc-Net*, Bodrum, Turkey, 2004.

[12] A. Ballardie, "Core Based Trees (CBT version 2) Multicast Routing," *RFC 2189*, September 1997, http://www.ietf.org.

[13] T. Bates, R. Chandra,D. Katz and Y. Rekhter, "Multiprotocol Extensions for BGP-4" *RFC4760*, January 2007, http://www.ietf.org.

[14] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana and A. Viterbi, "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users," *IEEE Commun. Mag.*, vol. 38, no. 7, pp. 70-77, 2000.

[15] B. Bhargava, X. Wu, Y. Lu and W. Wang, "Integrating Heterogeneous Wireless Technologies: A Cellular Aided Mobile Ad Hoc Network (CAMA)," *MONET*, vol. 9, no. 4, pp. 393-408, 2004.

[16] R. Bhatia, L. Li, H. Luo and R. Ramjee, "ICAM: integrated cellular and ad hoc multicast," *IEEE Transaction on Mobile Computing*, vol. 5, no. 8, pp. 1004-1015, 2006.

[17] S. Bhattacharyya, "An Overview of Source-Specific Multicast (SSM)" *RFC3569*, July 2003, http://www.ietf.org.

[18] R. Boivie, N. Feldman, Y. Imai, W. Livens, D. Ooms and O. Paridaens, "Explicit Multicast (Xcast) Basic Specifications," *draft-ooms-xcast-basic-spec-09.txt*, December, 2005, http://www.ietf.org.

[19] B. Cain, S. Deering, I. Kouvelas, B. Fenner and A. Thyagarajan, "Internet Group Management Protocol, Version 3," *RFC 3376*, October 2002, http://www.ietf.org.

[20] D. Cavalcanti, D. Agrawal, C. Cordeiro, B. Xie and A. Kumar, "Issues in Integrating Cellular Networks WLANs, and MANETs: A Futuristic Heterogeneous Wireless Network," *IEEE Wireless Commun.*, vol. 12, no. 3, pp. 30-41, 2005.

[21] D. Cavin, Y. Sasson and A.e. Schiper, "On the Accuracy of MANET Simulators," *in Proc. POMC*, Toulouse, France, pp. 38-43, 2002.

[22] M. Chipeta, M. Karaliopoulos, B.G. Evans and R. Tafazolli, "On the use of packet-level FEC and data carousels for the delivery of broadcast/multicast services to mobile terminals," *in Proc. VTC-Spring*, Stockholm, Sweden, pp. 2394-2399, 2005.

[23] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol," *RFC3626*, October 2003, http://www.ietf.org.

[24] C.d.M. Cordeiro, H. Gossain and D.P. Agrawal, "Multicast over Wireless Mobile Ad Hoc Networks: Present and Future Directions," *IEEE Network*, vol. 17, no. 1, pp. 52-59, 2003.

[25] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations," RFC2501, January 1999, http://www.ietf.org.

[26] S.K. Das, B.S. Manoj and C.S.R. Murthy, "A Dynamic Core Based Multicast Routing Protocol for Ad hoc Wireless Networks," *in Proc. MobiHoc'02*, Lausanne, Switzerland, pp. 24-35, 2002.

[27] S. Deering and D. Cheriton, "Multicast Routing in Datagram Inter-Networks and Extended LANs," *ACM TOCS*, vol. 8, no. 2, pp. 85-110, 1990.

[28] C. Diot, B.N. Levine, B. Lyles, H. Kassem and D. Balensiefen, "Deployment Issues for the IP Multicast Service and Architecture," *IEEE Network*, vol. 14, no. 1, pp. 78-89, 2000.

[29] DVB., "IP Datacast over DVB-H: Content Delivery Protocols," DVB Document A101, December 2005, http://www.dvb-h-online.org.

[30] E. Ekudden, U. Horn, M. Melander and J. Olin, "On-demand mobile media—A rich service experience for mobile users," *Ericsson Review*, vol. 78, no. 4, pp. 168-177, 2001.

[31] G. Faria, J.A. Henriksson, E. Stare and P. Talmola, "DVB-H: Digital Broadcast Services to Handheld Devices," *Proc. IEEE*, vol. 94, no. 1, pp. 194-209, 2006.

[32] B. Fenner and D. Meyer, "Multicast Source Discovery Protocol (MSDP)" *RFC3618*, October 2003, http://www.ietf.org.

[33] B. Fenner, M. Handley, H. Holbrook and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)," *RFC4601,* August 2006, http://www.ietf.org.

[34] J.J. Garcia-Luna-Aceves and E.L. Madruga, "The Core Assisted Mesh Protocol," *IEEE JSAC*, vol. 17, no. 8, pp. 1380-1394, 1999.

[35] Global Mobile Information Systems Simulation Library (GloMoSim) available at: http://pcl.cs.ucla.edu/projects/glomosim.

[36] I. Gruber and H. Li, "Cellular-Ad Hoc Network Interoperation for Coverage Extension," in Proc. MWC, Shanghai, China, pp. 513-516, 2004.

[37] M. Handley, I. Kouvelas, T. Speakman and L. Vicisano, "Bi-directional Protocol Independent Multicast (BIDIR-PIM)," *draft-ietf-pim-bidir-08.txt*, October 2005, http://www.ietf.org.

[38] T.J. Harrold and A.R. Nix, "Performance Analysis of Intelligent Relaying in UTRA TDD," *in Proc. VTC-Fall*, Vancouver, Canada, pp. 1374-1378, 2002.

[39] M. Hauge, Multicast in a Heterogeneous Cellular and Ad Hoc Network: Specification of an Ad Hoc Routing Protocol with Cellular Assistance, Department of Informatics, University of Oslo, Norway, Research Report no. 334, 2006, http://www.duo.uio.no.

[40] M. Hauge, A. Hafslund, F.Y. Li and Ø. Kure, "Multicast-service Distribution on a Cellular Network Assisted by Local Ad Hoc Networks," *in Proc. Med-Hoc-Net*, Bodrum, Turkey, pp. 68-80, 2004.

[41] M. Hauge and Ø. Kure, "A Heterogeneous Cellular and Ad Hoc Network Architecture for Multipoint Streaming: A detailed performance analysis," *in Proc. ICMU*, London, UK, 2006.

[42] M. Hauge and Ø. Kure, "Evaluation of Packet-level FEC with Multicast Streaming for a Heterogeneous 3G-Cellular and Ad Hoc Network," *in Proc. EuMob*, Alghero, Sardinia, Italy, 2006.

[43] M. Hauge and Ø. Kure, "Multicast in 3G Networks: Employment of Existing IP Multicast Protocols in UMTS," *in Proc. WoWMoM*, Atlanta, USA, pp. 96-103, 2002.

[44] M. Hauge and Ø. Kure, "Multicast Service Availability in a Hybrid 3G-cellular and Ad Hoc Network," *in Proc. IWWAN*, Oulu, Finland, 2004.

[45] M. Hauge and Ø. Kure, "Sticky Point-to-multipoint Channel for Multicast in UMTS," *in Proc. CIC*, Seoul, South Korea, 2003.

[46] L. Henden, et al., "Broadcast and multicast - a vision on their role in future broadband access networks (BMC Vision)," EU/IST FP6 - Broadcast/Multicast Cluster (BMC), Brussels, Belgium, January, 2005, http://www.cordis.lu/ist/ct/proclu/c/broadcast.htm.

[47] H. Holbrook and B. Cain, "Source-Specific Multicast for IP," *RFC4607*, August 2006, http://www.ietf.org.

[48] H. Holma and A. Toskala, *WCDMA for UMTS*, John Wiley & Sons, 2nd ed., September, 2002.

[49] H.-Y. Hsieh and R. Sivakumar, "Internetworking WWANS and WLANS in next generation wireless data networks," *in Proc. 3Gwireless*, San Francisco, CA, USA, 2002.

[50] Y.-C. Hsu and Y.-D. Lin, "Base-Centric Routing Protocol for Multihop Cellular Networks," *in Proc. GLOBECOM*, Taipei, Taiwan, pp. 158-162, 2002.

[51] IEEE, "IEEE 802, part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.

[52] IEEE, "IEEE 802, part 11 standard amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band," IEEE Std 802.11g™-2003, 2003.

[53] J-Sim available at: http://www.j-sim.org.

[54] J.G. Jetcheva and D.B. Johnson, "Adaptive Demand-Driven Multicast Routing in Multi-Hop Wireless Ad Hoc Networks," *in Proc. MobiHoc*, Long Beach, CA, USA, pp. 33-44, 2001.

[55] L. Ji and M.S. Corson, "Explicit Multicasting for Mobile ad hoc Networks," *MONET*, vol. 8, no. 5, pp. 535-549, 2003.

[56] T. Kunz, "Multicasting in mobile ad-hoc networks: achieving high packet delivery ratios," *in Proc. CASCON*, Toronto, Ontario, Canada, pp. 156-170, 2003.

[57] T. Kunz, "Multicast Versus Broadcast in a MANET," *in Proc. ADHOC-NOW 2004, (LNCS 3158),* Vancouver, British Columbia, Canada, pp. 14-27, 2004.

[58] S. Kurkowski, T. Camp and M. Colagrosso, "MANET Simulation Studies: The Incredibles," *Mobile Computing and Communications Review*, vol. 9, no. 4, pp. 50-61, 2005.

[59] J. Laiho, A. Wacker and T. Novosad, *Radio Network Planning and Optimisation for UMTS*: Wiley, 2002.

[60] A. Laouiti, P. Jacquet, P. Minet, L. Viennot, T. Clausen and C. Adjih, "Multicast Optimized Link State Routing," INRIA, *Research Report RR-4721*, February 2003, http://www.inria.fr/rrrt/rr-4721.html.

[61] S.J. Lee, W. Su and M. Gerla, "On-Demand Multicast Routing Protocol in Multihop Wireless Mobile Networks," *MONET*, vol. 7, no. 6, pp. 441-453, 2002.

[62] S.-J. Lee, W. Su, J. Hsu, M. Gerla and R. Bagrodia, "A Performance Comparison Study of Ad Hoc Wireless Multicast Protocols," *in Proc. INFOCOM*, Tel-Aviv, Israel, pp. 565-574, 2000.

[63] J. Li, C. Blake, D.S.J.D. Couto, H.I. Lee and R. Morris, "Capacity of Ad Hoc Wireless Networks," *in Proc. MobiCom*, Rome, Italy, pp. 61-69, 2001.

[64] Y.-D. Lin and Y.-C. Hsu, "Multihop Cellular: A New Architecture for Wireless Communications," *in Proc. INFOCOM*, Tel-Aviv, Israel, pp. 1273-1282, 2000.

[65] M. Luby, L. Vicisano, J. Gemmell, L. Rizzo, M. Handley and J. Crowcroft, "Forward Error Correction (FEC) Building Block," *RFC3452*, December 2002, http://www.ietf.org.

[66] M. Luby, L. Vicisano, J. Gemmell, L. Rizzo, M. Handley and J. Crowcroft, "The Use of Forward Error Correction (FEC) in Reliable Multicast," *RFC3453*, December 2002, http://www.ietf.org.

[67] H. Luo, R. Ramjee, P. Sinha, L.E. Li and S. Lu, "UCAN: A Unified Cellular and Ad-hoc Network Architecture," *in Proc. MobiCom*, San Diego, CA, USA, pp. 353-367, 2003.

[68] G.H. Lynn, "RoMR: Robust Multicast Routing in Mobile Ad-Hoc Networks," PhD thesis, the Graduate Faculty of the School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA, USA, 2003, http://pittcat.pitt.edu.

[69] J. Macker, "Simplified Multicast Forwarding for MANET," *draft-ietf-manet-smf-03.txt*, October 2006, http://www.ietf.org.

[70] MAESTRO, European IST project, http://ist-maestro.dyndns.org/maestro.

[71] D. Meyer and P. Lothberg, “GLOP Addressing in 233/8”, *RFC3180*, September 2001, http://www.ietf.org.

[72] N. Milanovic, M. Malek, A. Davidson and V. Milutinovic, "Routing and security in mobile ad hoc networks," *Computer*, vol. 37, no. 2, pp. 61-65, 2004.

[73] 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 Commun.*, vol. 11, no. 2, pp. 14-22, 2004.

[74] OMNeT++, Discrete Event Simulation System available at: http://www.omnetpp.org.

[75] J.C. Park and S.K. Kasera, "Enhancing Cellular Multicast Performance Using Ad Hoc Networks," *in Proc. WCNC*, New Orleans, LA, USA, pp. 2175-2181, 2005.

[76] S. Park and D. Park, "Adaptive Core Multicast Routing Protocol," *Kluwer Wireless Networks*, vol. 10, no. 1, pp. 53-60, 2004.

[77] S. Parkvall, E. Englund, M. Lundevall and J. Torsner, "Evolving 3G Mobile Systems: Broadband and Broadcast Services in WCDMA," *IEEE Commun. Mag.*, vol. 44, no. 2, pp. 30-36, 2006.

[78] C.E. Perkins, "IP Mobility Support for IPv4," *RFC3344*, August 2002, http://www.ietf.org.

[79] C.E. Perkins, E.M. Belding-Royer and S.R. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," *RFC3561*, July 2003, http://www.ietf.org.

[80] P. Radoslavov, D. Estrin, R. Govindan, M. Handley, S. Kumar and D. Thaler, "The Multicast Address-Set Claim (MASC) Protocol", *RFC2909*, September 2000, http://www.ietf.org.

[81] R. Ramanathan and J. Redi, "A brief overview of ad hoc networks: challenges and directions," *IEEE Commun. Mag*., vol. 40, no. 5, pp. 20-22, 2002.

[82] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, Englewood Cliffs, NJ, pp. 105 - 176, 2002.

[83] L. Rizzo and L. Vicisano, "RMDP: an FEC-based Reliable Multicast protocol for wireless environments," *ACM Mobile Computing and Communications Review*, vol. 2, no. 2, pp. 23-31, 1998.

[84] T. Rouse, I. Band and S. McLaughlin, "Capacity and Power Investigation of Opportunity Driven Multiple Access (ODMA) Networks in TDD-CDMA Based Systems," *in Proc. ICC*, New York, USA, pp. 3202-3206, 2002.

[85] E.M. Royer and C.E. Perkins, "Multicast Operation of the Ad-hoc On-Demand Distance Vector Routing Protocol," *in Proc. MobiCom*, Seattle, WA, USA, pp. 207-218, 1999.

[86] C. Selier and N. Chuberre, "Satellite Digital Multimedia Broadcasting (SDMB) system presentation," *in Proc. IST Mobile Summit*, Dresden, Germany, 2005.

[87] C.-C. Shen and C. Jaikaeo, "Ad Hoc Multicast Routing Algorithm with Swarm Intelligence," *MONET*, vol. 10, no. 1-2, pp. 47-59, 2005.

[88] A. Shokrollahi, "Raptor Codes," Digital Fountain, Tech. Rep. DF2003-06-001, June 2003, http://www.digitalfountain.com.

[89] P. Sinha, R. Sivakumar and V. Bharghavan, "MCEDAR: Multicast Core-Extraction Distributed Ad hoc Routing," *in Proc. WCNC*, New Orleans, LA, USA, pp. 1313-1317, 1999.

[90] K. Sipilä, Z.-C. Honkasalo, J. Laiho-Steffens and A. Wacker, "Estimation of Capacity and Required Transmission Power of WCDMA Downlink Based on a Downlink Pole Equation," *in Proc. VTC2000-Spring*, Tokyo, Japan, 2000.

[91] H. Tewari and D. O'Mahony, "Multiparty micropayments for ad hoc networks," *in Proc. WCNC*, New Orleans, Lousiana, USA, pp. 2033-2040 vol.3, 2003.

[92] The Network Simulator - ns-2 available at: http://www.isi.edu/nsnam/ns.

[93] C.-K. Toh, G. Guichal and S. Bunchua, "ABAM: on-demand associativity-based multicast routing for ad hoc mobile networks," *in Proc. VTC-Fall*, Boston, MA, USA, pp. 987-993, 2000.

[94] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Wirel. Netw.*, vol. 8, no. 2/3, pp. 153-167, 2002.

[95] H.-Y. Tyan. Design, Realization and Evaluation of a Component-based Compositional Software Architecture for Network Simulation. PhD thesis, Department of Electrical Engineering, The Ohio State University, 2002.

[96] R. Vida and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6," *RFC 3810*, June 2004, http://www.ietf.org.

[97] H.-Y. Wei and R.D. Gitlin, "Two-Hop-Relay Architecture for Next-Generation WWAN/WLAN Integration," *IEEE Wireless Commun.*, vol. 11, no. 2, pp. 24-30, 2004.

[98] S.B. Wicker and V.K. Bhargava, Reed-Solomon Codes and Their Applications. New York: John Wiley & Sons Inc., 1999.

[99] C.W. Wu and Y.C. Tay, "AMRIS: A Multicast Protocol for Ad hoc Wireless Networks," *in Proc. MILCOM*, Atlantic City, NJ, USA, pp. 25-29, 1999.

[100] H. Wu, C. Qiao, S. De and O. Tonguz, "Integrated Cellular and Ad Hoc Relaying Systems: iCAR," *IEEE J-SAC*, vol. 19, no. 10, pp. 2105-2115, 2001.

[101] X. Wu, S.-H.G. Chan, B. Mukherjee and B. Bhargava, "MADF: Mobile-Assisted Data Forwarding for Wireless Data Networks," *J. Commun. and Netw.*, vol. 6, no. 3, pp. 216-225, 2004.

[102] J. Xie, R.R. Talpade, A. Mcauley and M. Liu, "AMRoute: Ad Hoc Multicast Routing Protocol," *MONET*, vol. 7, no. 6, pp. 429-439, 2002.

[103] G. Xylomenos, "Group Management for the Multimedia Broadcast/Multicast Service," *in Proc. IST Mobile Summit*, Dresden, Germany, 2005.

[104] G. Xylomenos, V. Vogkas and G. Thanos, "The multimedia broadcast/multicast service," *Wireless Communication and Mobile Computing*, 2007. In press.

[105] A.N. Zadeh, B. Jabbari, R. Pickholtz and B. Vojcic, "Self-Organizing Packet Radio Ad Hoc Networks with Overlay (SOPRANO)," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 149-157, 2002.