IST-2003-507581 WINNER

D3.1

Description of identified new relay based radio network deployment concepts and first assessment by comparison against benchmarks of well known deployment concepts using enhanced radio interface technologies

Contractual Date of Delivery to the CEC:	*31.10.2004*
Actual Date of Delivery to the CEC:	**03/11/2004**
Author(s):	Abdulkareem Adinoyi, John Boyer, Stefano Brunazzi, Byron Bakaimis, Luca Coletti, Carl Eklund, David D. Falconer, Adrian Florea, Mariusz Glabowski, Jeremy Gosteau, Patrick Herhold, Reza Hoshyar, Tim Irnich, Paulo Jesus, Niklas Johansson, Rafal Krenz, Yajian Liu, Matthias Lott, Lino Moretti, Jussi Ojala, Afif Osseiran, Ilhem Ouachani, Ralf Pabst, Pietro Pellati, Quiliano Perez, Mahmudur Rahman, Nicola Riato, Adriano Santos, Malte Schmidt, Daniel Schultz, Kai-Erik Sunell, Tommy Svensson, Bernhard Walke, Michał Wódczak, Halim Yanikomeroglu
Participant(s):	*AU, CTH, CU, EAB, MOT, NOK, PTIN, PUT, SAGO, SEUK, SM, SMC, TID, TUD, UniS, VOD*
Workpackage:	*WP3 - Radio Network Deployment Concepts*
Estimated person months:	*104 PM*
Security:	**Restricted**
Nature:	**R**
Version:	1.0
Total number of pages:	275

Abstract:

In D3.1 the first concept for a WINNER radio network deployment mainly focused on relay based concepts are proposed. Further the concept of an radio and protocol architecture is presented that takes up the relaying approach as well as the idea of having different WINNER radio interface modes to serve different user and usage scenarios in an optimal manner. Thereby the presented protocol concept is based on generic and mode specific user an control plane functionalities, where the common functions for different modes are generalised and provide a common interface towards upper layer functions/protocols. Further some relaying concepts for urban hot area coverage including mesh networks are presented accompanied by promising simulation results. In addition to the relaying concepts benchmarks have been produced to allow a comparison of new concepts against today's systems.

Keyword list:

Relay based deployment, multi mode radio network architecture, generic link layer, mode convergence protocol, logical node architecture, fixed relay station, cooperative relaying, OFDMA, mesh network

Disclaimer:

Authors

Partner	Name	Phone / Fax / e-mail
Carleton University	Abdulkareem Adinoyi	Phone: +1-613-520-2600 Ext 1579 Fax: +1-613-520-5727 e-mail: adinoyi@sce.carleton.ca
Carleton University	Mahmudur Rahman	Phone: +1-613-520-2600 Ext 1579 Fax: +1-613-520-5727 e-mail: mmrahman@sce.carleton.ca
Carleton University	David D. Falconer	Phone: +1-613-520-5722 Fax: +1-613-520-5727 e-mail: ddf@sce.carleton.ca
Carleton University	Halim Yanikomeroglu John Boyer Adrian Florea	Phone: +1-613-520-5734 Fax: +1-613-520-5727 e-mail: halim@sce.carleton.ca
Chalmers University of Technology	Tommy Svensson	Phone: +46 31 772 1823 Fax: +46 31 771 1782 e-mail: tommy.svensson@s2.chalmers.se
Ericsson AB	Niklas Johansson	Phone: +46 8 508 778 60 Fax: +46 8 757 57 20 e-mail: niklas.j.Johansson@ericsson.com
Ericsson AB	Afif Osseiran	Phone: +46 585 32670 Fax: + 46 8 757 57 20 e-mail: Afif.Osseiran@ericsson.com
Ericsson AB	Kai-Erik Sunell	Phone: +46 8 757 35 61 Fax: +46 8 757 57 20 e-mail: kai-erik.sunell@ericsson.com
Motorola S.A.S.	Jeremy Gosteau	Phone: +33169352564 Fax: +33169354801 e-mail: jeremy.gosteau@motorola.com
Motorola S.A.S.	Ilhem Ouachani	Phone: +33169352559 Fax: +33169354801 e-mail: ilhem.ouachani@crm.mot.com

Motorola S.A.S.	Pietro Pellati	Phone: +33169352509 Fax: +33169354801 e-mail: pellati@crm.mot.com
Nokia Corporation	Carl Eklund	Phone: +358-718036566 Fax: +358-718036851 e-mail: carl.eklund@nokia.com
Nokia Corporation	Jussi Ojala	Phone: +358-718036248 Fax: +358-718036210 e-mail: jussi.ojala@nokia.com
Portugal Telecom Inovação SA	Paulo Jesus	Phone: +351 234 403 386 Fax: +351 234 424 160 e-mail: paulo-j-jesus@ptinovacao.pt
Portugal Telecom Inovação SA	Adriano Santos,	Phone: +351 234 403 386 Fax: +351 234 424 160 e-mail: it-a-santos@ptinovacao.pt
PUT	Rafał Krenz	Phone: +48 61 665 26 14 Fax: +48 61 665 25 72 e-mail: krenz@et.put.poznan.pl
PUT	Mariusz Głąbowski	Phone: +48 61 665 26 14 Fax: +48 61 665 25 72 e-mail: mglabows@et.put.poznan.pl
PUT	Michał Wódczak	Phone: +48 61 665 26 14 Fax: +48 61 665 25 72 e-mail: mwodczak@et.put.poznan.pl
RWTH Aachen University	Tim Irnich	Phone: +49 241 80 27928 Fax: +49 241 80 22242 e-mail: tim@comnets.rwth-aachen.de
RWTH Aachen University	Ralf Pabst	Phone: +49 241 80 25828 Fax: +49 241 80 22242 e-mail: pab@comnets.rwth-aachen.de
RWTH Aachen University	Daniel Schultz	Phone: +49 241 80 27916 Fax: +49 241 80 22242 e-mail: dcs@comnets.rwth-aachen.de

RWTH Aachen University	Bernhard Walke	Phone: +49 241 80 27910 Fax: +49 241 80 22242 e-mail: walke@comnets.rwth-aachen.de
Samsung	Byron Bakaimis	Phone: +44 (0) 1784 428 600 Fax: +44 (0) 1784 428 624 e-mail: byron.bakaimis@samsung.com
Siemens AG	Malte Schmidt	Phone: +49 2871 91 3963 Fax: +49 2871 91 3387 e-mail: malte.schmidt@siemens.com
Siemens AG	Matthias Lott	Phone: +49 (0)89 636 45136 Fax: +49 (0)89 636 45591 e-mail: matthias.lott@siemens.com
Siemens Mobile Communications	Luca Coletti	Phone: +39 2437 7489 Fax: +39 2437 7989 e-mail: luca.coletti@siemens.com
Siemens Mobile Communications	Nicola Riato	Phone: +39 2437 7418 Fax: +39 2437 7989 e-mail: nicola.riato@siemens.com
Siemens Mobile Communications	Stefano Brunazzi	Phone: +39 2437 7100 Fax: +39 2437 7989 e-mail: stefano.brunazzi@siemens.com
Siemens Mobile Communications	Lino Moretti	Phone: +39 2437 7255 Fax: +39 2437 7989 e-mail: lino.moretti@siemens.com
Telefonica Investigacion y Desarollo	Quiliano Perez	Phone: +34 91 3129801 Fax: +34 91 3374402 e-mail: qpt@tid.es
TU Dresden	Patrick Herhold	Phone: +49 351 463 35286 Fax: +49 351 463 37255 e-mail: herhold@ifn.et.tu-dresden.de
University of Surrey	Yajian Liu	Phone: +44 1483 686015 Fax: +44 1483 686011 e-mail: Y.Liu@surrey.ac.uk

University of Surrey	Reza Hoshyar	Phone: +44 1483 689489 Fax: +44 1483 686011 e-mail: R.Hoshyar@surrey.ac.uk
Vodafone	Adam Pollard	Phone: +44 1635 672479 Fax: +44 1635 676147 e-mail: adam.pollard@vodafone.com

Executive Summary

This document aims at the description and first evaluation of new innovative deployment concept and related protocol and network architectures to deploy a new WINNER air interface.

The WINNER radio interface is assumed to operate in different modes to serve different environments, e.g. rural or urban scenario, in an optimal way. To support these WINNER radio interface operating modes the protocol architecture has to provide means to generalise the interface to the higher layers and to allow a smooth interworking between the different modes. All these requirements are part of the WINNER radio interface architecture dealing mainly with layer 2 and 3 functionalities as described in this document. As next step a logical node architecture is described as basis for a new radio network architecture that comprise also the relay as new network element.

After the introduction of the harmonising radio interface architecture some concept of how to deploy a new broadband radio access network are described and partly evaluated, based on existing radio interface technologies. After a short overview on how an enhanced conventional single hop deployment of these systems might look like a substantial description of innovative new deployment concepts based on the idea of relaying is given. The introduced concept comprises mainly the idea of fixed relays, but gives also an introduction on how to use and benefit from mobile relays. The relay concepts further distinguish between homogeneous and heterogeneous relays and are including mesh networks.

As starting point for the chapter on new relay based deployment concept the mechanisms of Weighted Spectral Efficiency is described that a mean to evaluate deployment concept in a fair manner. The described mechanism of Weighted Spectral Efficiency comprises the costs for the deployment as well as the spectral efficiency. Following the mechanism some concepts on how to integrate relays are described based on MAC frame based protocols like, e.g. IEEE 802.16a or HiperLAN/2. Next to the concept description some simulation results on the traffic performance of a MAC frame based relay concept is presented. These results show the advantage of the relaying in order to deploy broadband systems in urban area where many obstacle are to overcome as well as in free space scenarios to extend the range of the access point. Another promising concept described in D3.1 is the concept of cooperative relaying, which uses relays as a kind of virtual antenna array in order to increase the link budget. The described concepts for heterogeneous relays are related to the protocol concepts described earlier in the document as a heterogeneous relay is also a form of interworking between two WINNER radio interface modes and can therewith benefit from the concepts to harmonise the radio interface operating modes. The relay based concepts are further supported by routing and forwarding strategies as described in this deliverable.

Finally the annex provides some benchmarks to allow a comparison of the new deployment concepts against existing systems. Further the annex contains some more detailed description and enhancements to relay based deployment concepts.

1. Introduction

The definition and development of new radio access network deployment concepts considering the potentials and requirements of the newly developed common radio interface technologies and the identified spectrum and related propagation characteristics is an important part of the new ubiquitous radio system concept.

The very high data rates envisioned for wireless systems beyond 3G (B3G) in reasonably large areas do not appear to be feasible with the conventional cellular architecture due to two basic reasons. First, the transmission rates envisioned for B3G systems are two orders of magnitude higher than those of 3G systems. This demand creates serious power concerns since it is well known that for a given transmit power level, the symbol (and thus bit) energy decreases linearly with the increasing transmission rate. Second, the spectrum that is considered for WINNER will be located well above the 2 GHz band used by the 3G systems. The radio propagation in these bands is significantly more vulnerable to non line-of-sight conditions, which is the typical mode of operation in today's urban cellular communications.

The brute force solution to these two problems is to significantly increase the density of base stations, resulting in considerably higher deployment costs that would only be feasible if the number of subscribers also increased at the same rate. This seems unlikely, with the penetration of cellular phones already high in developed countries. On the other hand, the same number of subscribers will have a much higher demand in transmission rates, making the aggregate throughput rate the bottleneck in future wireless systems. Under the working assumption that subscribers will not be willing to pay the same amount per data bit as for voice bits, a drastic increase in the number of base stations does not seem economically justifiable.

It is obvious from the above discussion that more fundamental enhancements are necessary for the very ambitious throughput and coverage requirements of future systems. Toward this end, in addition to advanced transmission techniques and collocated antenna technologies, some major modifications in the wireless network architecture itself that will enable effective distribution and collection of signals to and from wireless users are required. The integration of multi-hop capability into conventional wireless networks is perhaps the most promising architectural upgrade. Basically, rate can be traded for range of coverage and vice versa. Even more, such an architectural upgrade not only promises to reduce deployment costs, but also to enhance greatly the overall network performances compared to traditional cellular network concepts.

WINNER will consider multiple radio interface operating modes to cover the different types of scenarios in an optimal manner. The types of scenarios can be roughly divided into wide area and short range. The outcome of first research work towards new, scalable radio interfaces is planned to lead to a convergence and then allow for integrating the concepts for the different scenario into a common ubiquitous radio system. This issue will be discussed in Chapter 2 where concepts for a WINNER radio network architecture will be proposed that take different WINNER RAT modes as well as the relay based deployment concepts into account. The concept of generic link layer (GLL) is adopted to generalize all common link layer functions for different modes and provide a common interface towards upper layer functions/protocols. Besides that the Modes Convergence Protocol (MCP) has been introduced to enable a well structured design of B3G protocols of layer 2 and 3 according to the various operating modes by means of cross layer optimisation as well as cross plane management. The general common view of the GLL and MCP concepts is that the radio interface functions should be divided into generic user plane and control plane functions such that independent scalability, dimensioning and evolution of the user and control planes is achieved. An initial partitioning of functions into generic user plane and control plane functions is included in the proposal as well as their placement in the protocol architecture is discussed. A flexible logical node architecture model is also proposed to reduce the inter node interface definition, standardization and maintenance effort.

In Chapter 3 enhancement to current conventional single hop deployments will be discussed, before in Chapter 4 different multi-hop radio network deployment concepts and related routing concepts are introduced and discussed. The relay based concepts are divided in homogenous multi-hop concepts (Section 4.2) based on one radio interface mode only, heterogeneous deployment concepts (Section 4.3) based on multi-hop links that interconnect on more than one radio interface mode or two different radio interface technologies. The relay based concepts presented in Section 4.2 are accompanied by some promising simulation results, which highlight the potential of relay based concept for B3G network deployment in terms of traffic performance and spectral efficiency. In addition some other concepts are discussed in Section 4.4 where some multi-hop concepts are discussed that comprise WINNER and non WINNER air interface technologies. Further a methodology, Weighted Spectral Efficiency, will be introduced that allows a fair comparison of deployment concepts, either multi- or single hop.

The Annex contains benchmark results (Annex I) of well known system concepts, which have been simulated in the context of WINNER to be able to compare the new WINNER concepts against current radio systems, namely HSDPA, HiperLAN/2, IEEE 802.16 and IEEE 802.11. Further the Annex contains some more detailed information on the multi-hop concepts (Annex II)

1.1 WP3 Goals and Objectives

In order to fulfil the overall WINNER objectives, WP3 tackles the R&D activity of developing new deployment concepts with the following objectives:

- Definition, assessment and comparison of radio access network deployment concepts for each identified scenario based on both enhanced and new radio interfaces.
- Identification of agreed basic characteristics for each deployment concept in terms of physical and spectrum limitations, traffic performance, feasibility, complexity, ease of introduction and technologies on the link and system level.
- Identification of the best suited deployment concept for each of the identified scenarios based on the agreed basic characteristics and the scenario inherent requirements.
- Design and assessment of algorithms that allow coordination across base stations, access points and relay stations in order to coordinate the resource allocation and improve the mutual interference situation.
- Design and assessment of radio network protocols able to support the identified new deployment concepts and related technologies, especially multi-hop based systems, ad hoc networking with dynamic topologies and dynamic routing in either infrastructure based or infrastructure-less systems and coordination across base stations, access points and relay stations.

1.2 Terminology and Definitions

One **Access Point (AP)** coverage area can be divided into several cells. An AP might be geographically distributed. A **cell** is defined by the geographical coverage area of the broadcast channel. E.g., the cell as defined by 3GPP [3] is a radio network object that can be uniquely identified by a User Equipment (UE) from a (cell) identification that is broadcasted over a geographical area from one *UTRAN Access Point*.

Radio waves are received and transmitted by antennas that have certain antenna diagrams (antenna gain and phase as function of angle). A **beam** is radio waves received or transmitted by an antenna system with a certain antenna diagram. **Beamforming** is a function that determines the antenna diagram.

WINNER Connection Service (CS_W): The CS_W is the sum of network elements that are required to allow interworking between different Access Equipments (AE, see below), e.g. for inter-AE handover. Thereby the AEs can also be of different WINNER modes. The CS_W is also the interface to the backbone network, most likely the internet. *Optional (not in the scope of WINNER WP3):* The CS_W might also enable the direct interworking with other (non WINNER) AEs

WINNER Access Equipment for Mode X (AE_{WX}): The AE_{WX} as shown as enlargement in Figure 1-1 comprises always one AP and can be extended by one or several Relay Stations (RS) that can be either fixed (FRS) or movable/mobile (MRS). The AE_{WX} is assumed to be WINNER mode X specific. The AP-UT connection can involve several RS. Whether or not AP and RS will share common broadcast channels or not is subject to further study.

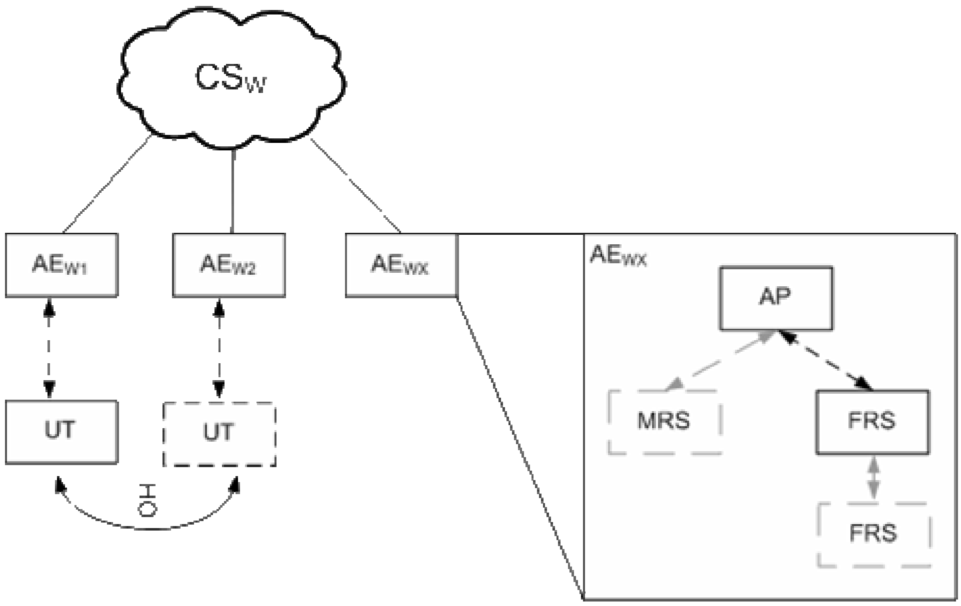


Figure 1-1: Proposed WINNER High-Level Access Network Architecture including relays

The following definitions have been extended from those given in D7.1 chapter 10 [1]:

RAN

Current 2G and 3G cellular radio system comprise the core network and the Radio Access Network (RAN). A RAN terminates all protocols involved in managing and maintaining radio connections towards user terminals. Example: in the case of UMTS the RAN is called UTRAN (Universal Terrestrial Radio Access Network) and is a conceptual term identifying that part of the network which consists of RNCs and Node Bs, see [3]. The RNC (Radio Network Controller) is a logical node in the RNS (Radio Network Subsystem) in charge of controlling the use and the integrity of the radio resources. An RNS offers the allocation and release of specific radio resources to establish means of connection in between an UE and the UTRAN. A Radio Network Subsystem contains one RNC and is responsible for the resources and transmission/reception in a set of cells.

Radio Link

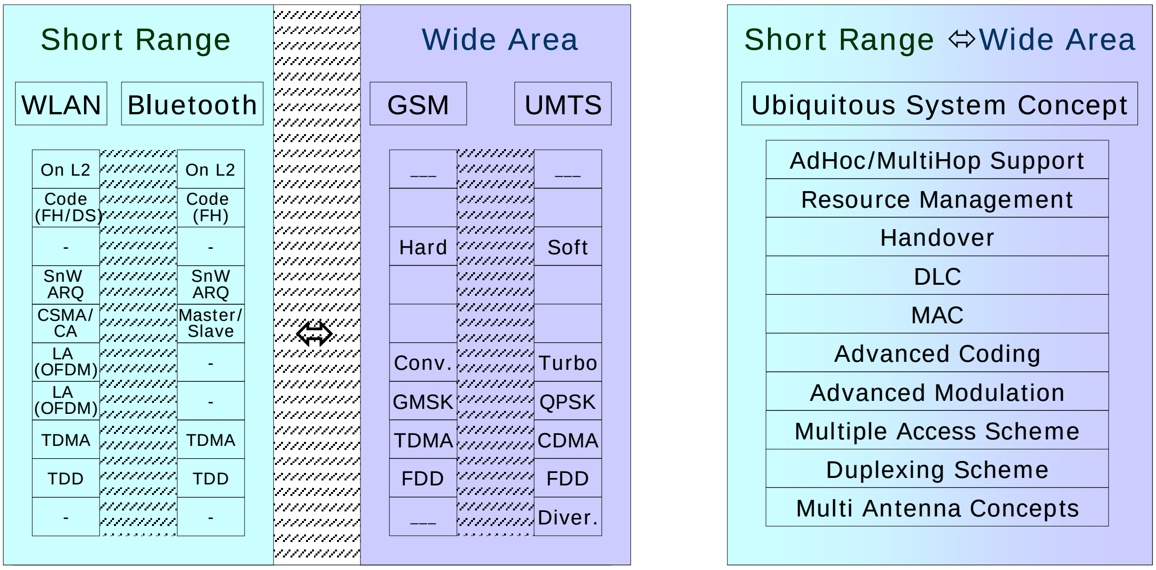
The radio link is a logical association between a single User Terminal (UT) and a single UTRAN access point. Its physical realization comprises one or more radio bearer transmissions. The radio bearer provides the services by the RLC (Radio Link Control) layer for transfer of user data between the UT and Serving RNC.

RAT

The Radio Access Technology is the air interface that is used to allow the link between end user equipment and Access Point or Base Station of the RAN. This includes also the links between RS and AP/BS. In current cellular systems, one RAT is usually associated to a RAN, e.g. UMTS RAN uses radio access technology CDMA. The RAT has influences on the functionalities and nodes (logical and physical) required in the associated RAN.

Mode (WINNER)

A goal of the WINNER project is to develop one RAT which can be adapted to a wide range of situations and environments, e.g. ranges, mobility, user densities. The adaptation of the RAT might require different parameterisations or use of different algorithms. Certain combinations of parameter or algorithm assignations or ranges of parameter or algorithm assignations may be referred to as Modes. The figure below shows the "one adaptable RAT" concept in contrast to different RATs. (The WINNER mode definition is further discussed in section 2.6).



Code:	Selection of different spreading-codes to separate user groups FH: Frequency Hopping, DS: Direct-Sequence
SnW:	Stop-and-Wait Automatic Repeat Request
LA (OFDM):	Link Adaptation for the OFDM-based systems

Figure 1-2 WINNER Common Radio Interface

Feeder system (WINNER)

The WINNER feeder system is the system used to feed the WINNER Access Points. The distinctive characteristic is that WINNER users can not connect to this network. The transmission technology used by the feeder system could be wireless or wireline and is irrelevant and transparent for the final user.

Radio access system (WINNER)

The WINNER radio access system is the wireless network that is used to connect the WINNER user terminal. The elements of this network are the WINNER access elements (see above, consisting of access point and the relay station), WINNER user terminals may function as mobile relays and thus become part of the radio access system. WINNER RAT and its modes are used by the WINNER radio access system.

Homogeneous Relay

The homogeneous relay station is a network element with relaying capabilities that is wirelessly connected to an AP, another RS and/or a UT and that uses the same radio technology (RAT_{modex}) for all its connections.

Heterogeneous Relay

A heterogeneous relay station is a network element that is wirelessly connected to another relay station or an access point (base station) by means of a given radio access technology, and serves to another relay station or to a User Terminal (UT) using a different radio access technology. The radio access technologies that a heterogeneous relay incorporates can be different modes of the same RAT (i.e. in the WINNER context), one WINNER RAT-mode and another (possibly legacy) RAT, or two (legacy-) RATs, where the latter is not inside the WINNER scope of research.

Logical and Physical Nodes

A logical node denotes an entity in the (radio access) network that terminates a certain set of protocols. A physical node denotes a physically existing device in the network that incorporates certain functionality, thereby representing one or possibly even more logical nodes.

2. The WINNER Radio Interface Architecture

2.1 Introduction

In addition to the conventional service delivery mode employed in present day cellular networks, it has been proposed in WINNER that mobile terminals could receive service through multiple intermediate nodes, called relay nodes. The coexistence of the two network paradigms will inevitably levy new requirements on the radio protocols as well as require new functionality on the radio interface. A wide range of questions needs to be answered relating to the proposed concept. For instance, what kind of functionality will be needed in the relay nodes, in what kind of scenarios is it beneficial to deploy them, and how do relays integrate into the WINNER Radio Network Architecture?

The benefits of relays are that they improve link budgets and may hence ease covering remote (or shadowed) areas. But there also exists issues that need to be considered, like in the case of rapidly moving users, which may, due to inherent delay issues, not be able to take advantage of the relay nodes. Furthermore, scenarios may exist where the deployment of relay nodes may be undesirable due to resulting complicated interference management issues. A similar kind of problem has been confronted in hierarchical cellular systems where the deployment of overlapping macro-cells and micro-cells is always a source for unpredictable interference and near-far effects.

Introduction of mobile relay nodes have also been proposed for WINNER. Making the relays mobile also introduces new challenges beyond those that are seen when dealing only with fixed relays. For instance, dealing with interference between mobile relays may be challenging especially in situations where the mobile relay population is “relatively speaking” dense. However, mobile relays, when acting as moving gateways, may also offer benefits by reducing the number of simultaneous handovers in the network (only the moving gateway is involved in the handover) and by allowing terminals to operate under more favourable link conditions.

It is also envisioned that a mobile device equipped with a WINNER radio should support - and be able to use - multiple modes for communication. Moreover, a mobile device should be capable of maintaining communications in more than one mode concurrently, even in cases where it is only equipped with a single transceiver. To accomplish this it is foreseen that a set of multiple access protocols, duplex techniques and radio link control protocols that are not necessarily shared between the modes of communication, need to be employed. In addition to the capability to communicate with a base station, either directly or via intermediate relay nodes, the capability of communicating directly with other WINNER mobile devices without any involvement of an operator should also be supported. In relation to this, an important key element in the proposed architecture is the generic link layer (GLL) that aims to generalize common link layer functions for different modes of the radio interface (e.g. handover, packet scheduling and retransmissions). As such, the GLL may be viewed as a toolbox of link layer functions that may be adapted to the characteristics of the different supported modes in WINNER. Moreover, GLL is envisioned to facilitate mode transfer on the data link layer without losing any data and it also provides a common (i.e. for all modes) interface towards upper layer functions/protocols.

Besides that the Modes Convergence Protocol (MCP) has been introduced in Section 2.7 to enable a well structured design of B3G protocols of layer 2 and 3 according to the various operating modes. The MCP supports seamless interworking between different WINNER RAT modes and supports the idea of generic user plane and control plane functions. The MCP supports the transfer of protocol status information during the handover between two modes. Therefore a cross layer optimisation as well as a cross plane management is foreseen.

The link layer functionality may also need to support spectrum sharing between WINNER networks and other systems.

Questions like where control and coordination entities should be located to accomplish this will be discussed in this section of the document.

The deployment and operation of WINNER networks needs to be easier than current cellular networks. Thus the operational aspects need to be taken into account in the design from the beginning. The WINNER architecture should support auto-configuration and auto-tuning. These requirements are especially important in relation to relay nodes, as the amount of relay nodes may exceed significantly the number of current cellular BTS sites.

In other words, the WINNER radio interface architecture should support a multitude of scenarios where completely different kinds of solutions or radio link configurations may be required. One possible approach to the problem is to use conventional (restrictive) deployment models that specifies different (classes or) types of network nodes and defines interfaces between them. Unfortunately, relying on such models often results in the need of large number of logical network nodes, each having a distinct set of capabilities. A side effect of this is that a large number of inter node interfaces need to be defined, standardized and maintained. Moreover, the architecture needs to allow flexible interworking (switching) in between the different modes currently being developed within WINNER. The remainder of the section seeks to define a flexible and logical radio system architecture that attempts to keep the number of distinct interfaces to a manageable number.

The rest of this chapter is organized as follows. Section 2.2 outlines the requirements on the architecture. Sections 2.3 and 2.4 discuss control plane and user plane functionalities and their objectives. Section 2.5 introduces a link layer concept that generalizes common link layer functions for all modes as well as provides a unified interface towards upper layers. Section 2.6 describes the functional layer architecture and explains how the different radio interface functions and protocols are related to each other. Section 2.8 proposes a logical node architecture and exemplifies the usage and deployment of logical network nodes for different operational scenarios. Section 2.9 in turn presents envisioned deployment scenarios and supporting physical nodes. Finally, section 2.10 draws some conclusions and discusses some topics for further studies.

2.2 Requirements on the Architecture

The purpose of this document is to outline a first proposal for the WINNER Architecture, including the overall functions and the inter-relations of logical nodes in this network, as well as a high-level protocol architecture. Important requirements for the WINNER architecture are:

1. The architecture shall facilitate provision of high bit-rate services (i.e. up to 1 Gbps) with low round-trip delay in line with WINNER requirements. Especially in the context of multi-hop operation this requires very fast connection establishment procedures, which will have to be enabled by both the physical layer, the layer 2 and (where involved) the layer 3 protocol. The envisioned high peak data rates of the air interface require fast access network transport (especially important in the case of wireless feeders) and the low latency constraints require fast switching.
2. The architecture shall allow for radio resources to be used as efficiently as possible. For the protocols this means that they have to be flexible to exploit the available spectrum in all its dimensions: time, space and frequency.
3. The architecture shall allow for scalability as well as optimizations for a given scenario, i.e. it shall allow flexibility in network deployment[1]. Drivers that facilitate this flexibility are ease of configuration (plug and play) and low-cost equipment
4. The architecture shall preserve layer transparency by providing to upper layers a unified interface for transparent utilization of the various accesses. However this interface towards upper layers needs to have the capability to inform these layers about the capabilities and options (in terms of QoS) that the current configuration of the lower layers provides.
5. The number of logical network nodes and interfaces shall be sustained as low as possible (supporting requirement number 3). Moreover, in the proposed architecture, the choice of radio interface capabilities or radio link modes can be based on the actual situation in the radio environment, the properties of the involved network nodes and the requested services (i.e. mode switching should be supported).
6. The Architecture needs to allow interworking in between different modes as well as with other (new and legacy) radio technologies. It should therefore integrate seamlessly into the Multi-Radio-Access (MRA) Architecture defined within WWI-Ambient Networks (AN)

To support this integration, the architecture has to support

a. the transfer of protocol status information of existing connections between the different modes. The potential applications for this are multi-mode terminals or multi-mode base stations performing inter mode handover.

b. information gathering about existence, status, capabilities of different modes for efficient decision making. This includes cross-mode and cross-system measurement procedures.

[1] For instance, it should enable full-scale networks with centralised functionality as well as stand alone networks (e.g. similar to a WLAN AP with decentralised functionality) to be deployed.

2.3 Control Plane Functionality

2.3.1 Introduction

Control-Plane Functionality (or Radio Resource Management (RRM)) is a general term for numerous algorithms and protocols used to govern the radio resources at different layers in the architecture. The overall goal is to utilize the given radio resources in an efficient manner. Examples of RRM functionalities are power control, cell selection (handover), admission control, load (congestion) control and channel allocation.

The different RRM algorithms are usually not regulated by any specification (i.e. they are vendor specific), but the specifications may provide some framework description to alleviate interoperability. The protocol part on the other hand includes specifications on messages exchanged between the different network nodes (and the frequency of these messages) and the associated network interfaces.

To aid these RRM functions *measurements* have to be performed. Sometimes, relevant information can be extracted directly at the transmitter side, but more often the measurements are performed in the receiver and reported back to the transmitter via some feedback link. When developing a RRM measurement strategy one must consider what quantities that should be measured, how frequent the measurements should be performed (or what should trigger such measurements) and how these measurements should be reported back to the transmitter. An example illustrating measurements performed for a RRM function could be cell selection measurements. These measurements are used for cell selection (association) and handover purposes. The measurements are generally based on some transmitted downlink pilots and comprise e.g. pilot strength, user positions and antenna directions.

The RRM measurements may be performed regularly, upon request or perhaps be triggered by some event. To achieve the best possible performance, the time period of the measurements should be chosen such that it may track the variations of the measured quantity. However, since some quantities (like a fading radio channel) undergo not only short-term but also long-term variations, lower measurement frequencies may still be useful. Moreover, the measurement frequency may be adaptive, adjusting the time period according to the variation of the measured quantity.

Similarly as for the measurements, the reporting of the measurements can be performed periodically, upon request or be event triggered. Naturally, the measurement reporting strategy should not be selected independent of the measurement strategy. We may distinguish between the case in which the measurement data itself is reported to the network and the case in which the measurement data is processed in the user terminal and only a RRM command is reported back to the network.

As previously described, it is perceived that the WINNER Air-Interface will encompass a number of modes, each of them targeted and optimized for a specific deployment scenario (and possibly also user scenario). Nevertheless, many of the RRM functions may be regarded as mode-agnostic and may hence be shared between the different WINNER modes (e.g. cell selection and admission control). The RRM functions may hence be divided into Generic RRM functions and Mode Specific RRM functions. However, it is very likely that there will exist an overlap between the Generic and Mode Specific RRM functions, where the Generic part will administrate the Mode Specific part. An example could be power control where the Generic part may specify an interval of admissible power allocations (for instance regulated by some maximum induced inter-cell interference), whereas the Mode Specific part performs the actual power allocation within these limits. The multiple modes on WINNER will also result in that new Generic RRM functions will need to be defined, e.g. mode discovery and mode selection[2] (see section 2.3.3.1 for a more detailed description). Even though some functions may be regarded as being generic, these functions will still rely on measurements performed by the mode-specific protocols (e.g. CQI reporting). As the underlying modes may rely on totally different principles (e.g. duplexing schemes) they may perform measurements in a different manner and have to be reported in such a way that they may be compared to one another.

2.3.2 RRM functionality integration for WINNER

In WINNER systems, different scenarios may employ different WINNER modes. These modes differ in terms of deployment concepts and/or radio interfaces. There are two basic cases for the co-existence of these WINNER modes: case A and case B, as shown in Figure 2-1. Note that the co-existence actually

[2] It should here be noted that cell selection only operates on one single mode whereas mode selection is only active when two modes coverage areas are overlapping – however they may be tightly coupled (e.g. making use of the same channel quality measurements)

refers to spatially dispersed user populations that use certain modes. When the number of modes in the systems is more than two, the actual co-existence situation could be a combination of case A and case B.

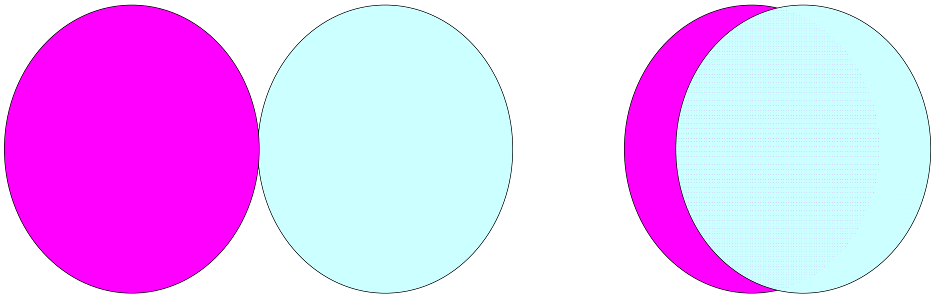


Figure 2-1 Basic cases for the co-existence of WINNER modes

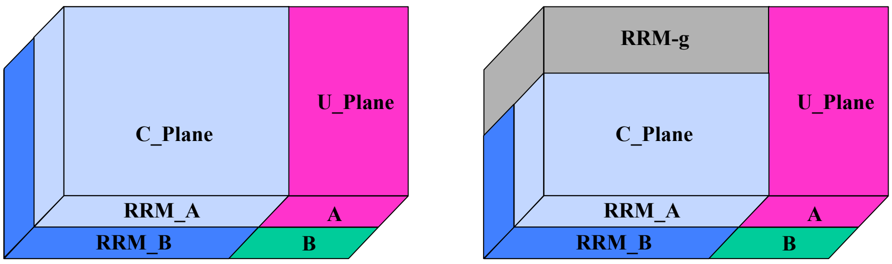
In case A, the two depicted WINNER modes are allocated orthogonal resources (e.g. in time or frequency) and hence the influence on one mode from the other is only marginal. These orthogonal resources may be allocated to the different modes either on a permanent basis or they may be adapted to the current situation in the two modes (e.g. one mode may be allocated more resources than originally planned for this mode by using resources originally planned for the other mode). On the contrary, in case B, the different WINNER modes are allocated non-orthogonal resources, and hence, the actions taken for one mode will (highly) affect the radio link quality in the other mode (and vice versa).

For case A, the RRM functions may be developed individually (though with some module assigning orthogonal resources to the two modes in case of adaptable resources). Nevertheless, as previously mentioned it is envisioned that several functionalities may be identified and developed as mode-independent (i.e. these functions may be generalized without hampering the performance of the underlying modes) and by doing so several benefits may be encountered e.g.:

- Cooperation of individual modes (e.g. the case above where resources may be re-allocated to a different mode).
- Mode selection (see section 2.3.3.1) based on e.g., load, service, and current radio conditions for the modes.
- Concurrent admission control (even if the two modes may not have resources to allow a new flow to be accepted perhaps it is possible to multiplex the flow over both modes and hence accept the flow).
- Easy integration of future modes into the WINNER context.
- …

Moreover, this will also lower the standardization efforts needed since fewer functions need to be standardized.

Figure 2-2 illustrates the RRM integration for case A, depicting both the case where the RRM functionalities have been developed individually (Figure 2-2.a) and the case where it is assumed that some functionality may be generalized for all underlying modes (Figure 2-2.b). Here C-Plane and U-Plane is short for control-plane functionality and user-plane functionality respectively and the generic RRM module is denoted as RRM-g.



a) b)

Figure 2-2 RRM Integration for case A

As described above, in case B the two WINNER modes are allocated non-orthogonal resources and hence what is performed for one mode will influence the performance of the other mode. Here two different sub-cases may be distinguished. In the first case the RRM functions have been developed under the assumption that the developed mode will be allocated resources that are non-orthogonal to all other modes as well as legacy and future radio access technologies that may exist in the same frequency spectrum. In this case, one would have to re-design the control-plane considering all modes (and legacy and new RATs) in conjunction and thereby develop a new generic control plane working for all involved modes. This case is illustrated in Figure 2-3.a. In the second case the RRM functions have been developed with the aim of being able to coexist with other modes that are able to allocate the same resources (an example of a technology employing this strategy is IEEE 802.11 that first senses the medium before allocating any resources, however, any of the technologies developed to be used in any unlicensed spectrum may also be used as examples). The RRM protocol stack for this latter case would look exactly the same as for case A were it has been assumed that some important parts have been generalized and is once more depicted in Figure 2-3.b.

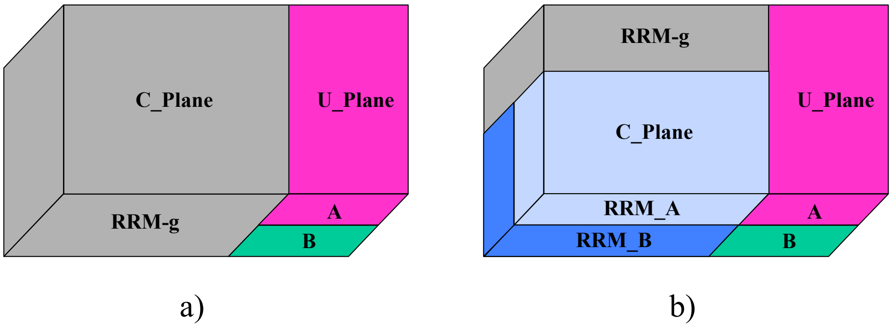


Figure 2-3 RRM Integration for case B

The question of what RRM functions that may be generalized for all modes is an issue that highly depends on the characteristics of WINNER modes and requires further studies. Nevertheless, in the next section some key RRM functions that may perhaps be generalized are presented. In the rest of this document it will be assumed that the different modes share the resources according to case A were some functionality has been generalized.

2.3.3 Potential key generic RRM functions and their objectives

2.3.3.1 Mode Selection

Generic RRM has a responsibility in the selection of one or several modes to serve a particular session (this also includes the allocation of the required radio resources). An important aspect is that Generic RRM may assign more than one mode to a session, allowing the Generic Link layer (GLL) (see section 2.5) a certain degree of freedom for packet scheduling through any of the configured modes. Modes may also be released from a session (e.g. in response to congestion) and if so, this decision is communicated to GLL and/or other relevant entities.

There are several mode selection principle options. These include the architecture, i.e. what node takes the mode selection decision, the time scale of the mode selection, and the mode selection decision variable.

Architecture

The mode selection can either be controlled by a node in the fixed part of the network, an access point, a relay node or by the user terminal. If one operator runs the network, this operator likely prefers network controlled mode selection to be in control of its users. If the network is run by different operators, terminal based mode selection is perhaps more likely. The different architecture alternatives may be expected to differ in what decision variables are available.

Time Scale

The mode selection can be made on different time scales. These include per call or session, per packet, or triggered by non-traffic related events like changes in radio conditions. More frequent mode selection leads to larger potential capacity and quality gains, but also increases signalling load. In some cases mode selection will require user interaction or be dictated by an application.

Decision Variables

Mode selection decision variables include those listed below. A combination of these may also be used.

- Service – the mode, which is most efficient for the service in question, is selected.
- Signal strength – the mode from which the strongest signal is received is selected. This is a simple version of the more sophisticated selection parameter radio resource cost discussed below.
- Radio resource consumption – the mode for which the least relative amount of radio resources (e.g. power, codes, timeslots etc.) is required to support the users is selected.
- System load – the mode with the least relative load is selected. This does not necessarily improve capacity, but improves user quality at loads below the capacity limit, and avoids reallocations as the load approaches the capacity limit.
- Packet size – the mode which is most efficient for the packet size in question is selected.
- Price – the mode for which the price for supporting the user is selected. This is especially interesting in case the different modes are run by different operators, and the decision is taken by the end-user or the terminal.

2.3.3.2 Cell selection and handover

Handover is responsible for handling the mobility of the user when he or she is moving from the coverage area of one cell to another.

In WINNER systems, handover can be broken down into two main groups of functions: inter-mode handover, and intra-mode handover. Inter-mode handover here refers to the case where some connection is handed over from one mode to another mode (between different cells or possibly within the same cell depending on how the cells are composed), e.g., handing over from a cooperative relay mode to a store-and-forward relay mode[3] could be classified as an inter-mode handover. Intra-mode handover on the other hand refers to the more classical case were a connection is handed over from one cell to another cell while still using the same mode.

Both types of handover could be further broken down into further sub-types, e.g., soft handover and hard handover. In soft handover, terminals or RNs could maintain more than one simultaneous connection to the network. While in hard handover, the old connection must be broken before the new connection is established.

Furthermore, cell selection and handover supports mode selection (and vice versa) by enabling some connection to be handed over to a new mode.

Architecture:

The main control functionality of handover is envisioned to be sitting in some central network node (see section 2.8 for a more thorough description of envisioned logical nodes and possible location of RRM functionality), but terminals and RNs should conduct some necessary measurements to assist the decision-making.

Time Scale:

The handover procedures should be invoked in a relatively fast manner and should at least be as frequent as mode selection to support its operation.

Decision Variables:

- Signal strength, such as averaged downlink pilot power etc.
- Terminal and/or RN capabilities, such as the supported modes etc.
- Terminal and/or RN locations
- Service in conjunction with mobility

[3] For a thorough description of these two possibilities for forwarding in a relay network the interested reader is referred to section 4.2.

2.3.3.3 Admission control

Admission control is responsible for controlling the load of the system so that the available capacity may be exploited without compromising the system stability. Admission control may further be broken down into two different parts, here denoted session admission control and handover admission control. The session admission control function is responsible for whether or not to admit a new session into the system as a whole. The new session is admitted if the service requirements (QoS, security etc.) of the new session as well as already established sessions may be fulfilled (N.B. even though one mode may not have the necessary resources available, a new session may be admitted if it may use more than one mode to fulfil its requirements). Session admission control is also invoked if the requirements for the session are altered. Handover admission control is invoked whenever a session is moved from one cell or- one mode - to another to decide whether the session may be admitted in the new cell/mode. If designed properly admission control will result in a moderate increase in call blocking probability but a substantial decrease in call dropping probability.

Admission control plays a very important role in circuit-based systems to prevent some extremely “bad” users to enter the system and create intolerable interferences, or to block some users when the system is heavily loaded. However, in packet-based systems, the necessity of admission control may be questioned due to the existence of packet scheduling. This is so since the scheduler can make the decision on who may transmit and who should wait and as a result, the stability of the system could be well maintained. Nevertheless, it may still be necessary to control the admission of new sessions into the system in order to protect services with strict QoS requirements, e.g. real-time services which are sensitive to packet delay.

The admission control should prevent new users to be served when the systems reach their hardware limits (such as number of channels per cell) by means of blocking these users directly (Erlang B) or put these users into a waiting list and serve them later when resources are once more available (Erlang C).

Normally the system capacity is however limited by interference. Hence, before the admission control makes the decision to admit or reject a new user, it should first measure/estimate the current system interference/load and/or the amount of available resources, e.g. transmitter power, meanwhile, predict the interference/load, e.g. power consumption, increment of this new user, then summarize these two loads and compare with the maximum load threshold and/or the maximum amount of available resources e.g. transmitter power. If it is under the threshold, admit this new user otherwise block it.

In WINNER, the admission control could be even more complex since it is envisioned that WINNER should support multiple modes. Hence, it may happen that the admission control scheme can be different for different modes. Nevertheless, it is also a task of WINNER to define *a minimum set of harmonized modes* and hence this may not be a problem in the end (however this is currently not known since there are currently no modes available). Furthermore, regardless of the (mode-specific) capacity limitation factors, care has to be taken when defining the set of (generic) primitives that should be used by the admission control algorithm so as to enable comparability in between different modes.

Architecture:

The main control functionality of admission control is envisioned to be sitting in some central network node (see section 2.8 for a more thorough description of envisioned logical nodes and possible location of RRM functionality) but terminals, RNs and APs should conduct some necessary measurements to assist the decision-making.

Time Scale:

The session admission control is invoked whenever a new session enters the system and whenever the QoS requirements are changed for an already admitted session. Handover admission control is invoked whenever an admitted session makes a handover to a different cell or mode.

Decision Variables:

- AP or RN load status
- AP or RN power headroom
- Terminal capabilities, such as supported modes etc.
- Terminal locations

2.3.3.4 Control of interval of admissible power allocations

Power control is an important RRM function used to control the terminal power consumption (hence save their battery), and thereby control the experienced interference level. One specific challenge for power control in WINNER systems is how to allocate the power to different relay nodes along the route from access point to user terminal (and vice versa) to guarantee the QoS requirements and distribute the interference properly.[4]

This form of power allocations is envisioned to be part of the mode-specific part. Nevertheless, it is foreseen that the generic part may specify an interval of admissible power allocations (for instance regulated by some maximum induced inter-cell interference) as a means to promote coordination between access points, whereas the mode specific part performs the actual power allocation within these limits.

Architecture:

The main control functionality of power control is envisioned to be sitting in some central network node (see section 2.8 for a more thorough description of envisioned logical nodes and possible location of RRM functionality), but terminals and RNs should conduct some necessary measurements to assist the decision-making.

Time Scale:

This kind of power control should be invoked in a relatively slow manner

Decision Variables:

- Induced interference
- Available transmitter power

The following control plane functions should be executed close to the user plane termination points and it is therefore envisioned that they will be governed by GLL (see section 2.5).

2.3.3.5 Traffic policing

The goal of traffic policing is to protect the network against any kind of persistent congestion and traffic overflow situations. This could be the result of misbehaving, i.e., non-rate-adaptive applications, or of Denial-of-Service attacks.

Architecture:

Traffic policing is envisioned to be residing close to the generic data link layer buffers situated within GLL at RLC-g (see section 2.5).

Time Scale:

Constantly active

Decision Variables:

- Requested capacity by some flow (user)

2.3.3.6 Congestion control and buffer management

Congestion control is responsible for returning the system back to the target load quickly and controllably when the system is overloaded. Since rate-adaptive applications have congestion control “built-in” it is envisioned that we do not have to design much explicit support for congestion control into the WINNER system. Instead, we can solely rely on flow-based queuing and Active Queue Management, AQM (see for instance [11])[5], to be implemented in the WINNER system (further outlined below).

AQM is a proven mechanism to efficiently maintain queues. AQM can either be operated with “packet drop” as the implicit congestion signal or with the Explicit Congestion Notification (ECN), where instead of dropping a packet the packet is marked with “congestion experienced”. That mark is then echoed by the receiving rate-adaptive application (e.g., TCP). ECN is starting to be deployed today. The key features of AQM is to react (drop or mark packet) before the queue has become full in order to leave buffer space to absorb transient packet bursts, and to queue as many packets as needed to maximize end-to-end throughput, but as little as possible to minimize end-to-end delays. When performing these actions the quality-of-service requirements (extracted from application/user requirements, e.g. delay boundaries and user priorities) need to be considered.

---

[4] Two different schemes may be envisioned here: centralized (i.e. the access point is responsible for the power allocations) and distributed (every node along the route is responsible for the power allocation on a hop-by-hop basis)

[5] N.B. this function is most often implemented at the IP-layer.

Architecture:

AQM is envisioned to be residing close to the generic data link layer buffers situated within GLL at RLC-g (see section 2.5).

Time Scale:

Constantly active

Decision Variables:

- AP or RN load status
- AP or RN power headroom
- Service class
- User priorities
- Age of PDU

2.3.3.7 Packet scheduling and flow classification

Packet scheduling may be considered as management between flows of the same user and/or between different users (taking fairness between different flows as well as service classes into consideration) and determines the order in which PDUs will request resources from the *resource allocation function*. In downlink (uplink) operation the resource allocation function determines which flow (user), out of the ones that have requested resources, that should get access to the radio channel and what resources that should be used for this.[6] The means to classify different flows into different service classes (possibly application and/or user dependent) will henceforth be referred to as *flow classification*.

Packet scheduling between the different service classes is what determines the level of QoS provided to any specific service class. Scheduling within one service class determines how flows of the same service class are treated with respect to each other, e.g., fairness based on number of bytes or packets sent, fairness based on amount of consumed radio resources, etc. It is believed that one packet scheduler may perform both these tasks.

Out of these functions it is believed that packet scheduling and flow classification may be generalized whereas resource allocation is believed to be mode-specific.

Time Scale:

Constantly active

Architecture:

The packet scheduling and flow classification functionality is envisioned to be residing close to the generic data link layer buffers situated within GLL at RLC-g (see section 2.5), in the user terminal and possibly also in any intermediate relay node.

Decision Variables:

- Service class
- User priorities
- Age of PDU
- Fairness

Figure 2-4 gives an example of how the different functionalities described in this sub-section (as well as some other functions) could be implemented in the downlink. Here, and as outlined above the packet scheduler and the resource allocation are kept as separate entities. The resource allocation entity is controlled from the packet scheduler through requests from the packet scheduler.

[6] N.B. this form of resource allocation is often referred to as packet scheduling but to enable a distinction (and avoid confusion) between the two functionalities we have introduced the definitions above.

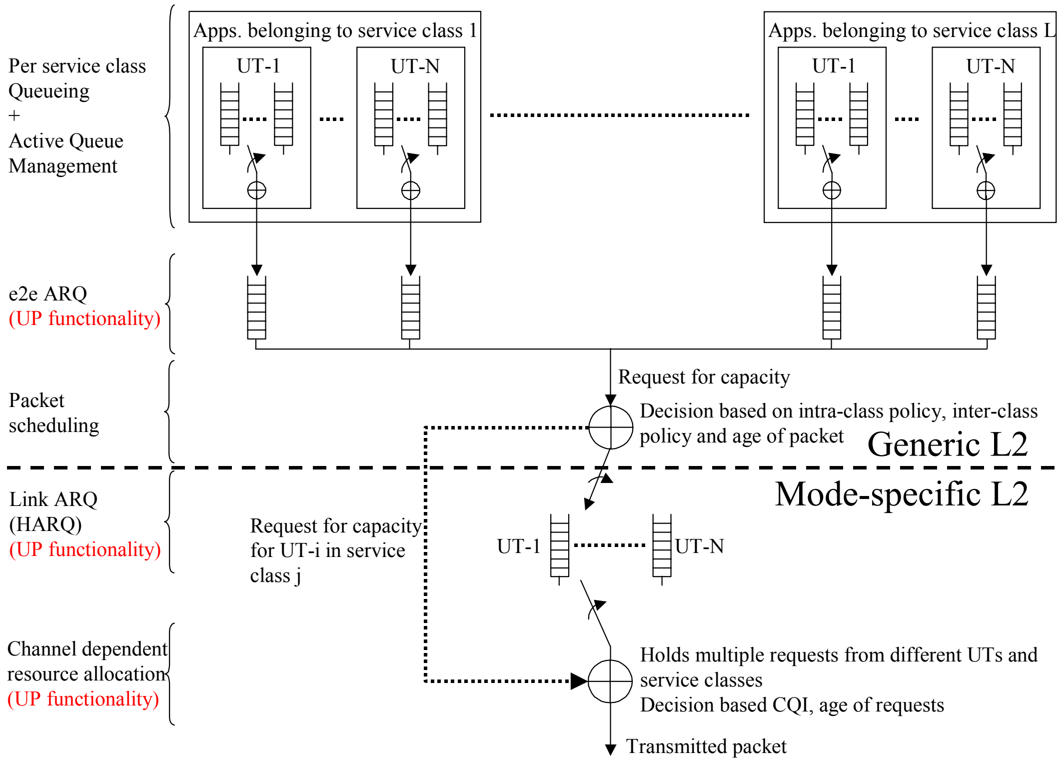


Figure 2-4 Example of scheduling, priority handling and resource allocation split between generic and mode specific functions.

2.3.3.8 Routing

Routing is a crucial RRM function for multi-hop networks to find the best routes to forward packets on. Routing schemes could be basically classified into two groups: proactive routing, and reactive routing. In the first type of routing, network nodes continuously evaluate the routes within the network, so that when a packet needs to be forwarded, the route is already known and can be immediately used. Reactive protocols, on the other hand, invoke a route determination procedure on demand only. Thus, when a route is needed, some sort of global search procedure is employed.[7]

Architecture:

Two main scenarios may be envisioned for routing functionality: either the main control functionality of routing resides in the access points only (centralized routing) or it resides in all network nodes on the route between some access point and some user terminal (for one connection) that also participates in the route determination (distributed routing). NB that, in the former case, the terminals and relay nodes should conduct necessary measurements to assist the access point.

Time Scale:

The time scale of when the routing protocol should be invoked is highly dependent of employed routing protocol and is therefore only applicable when discussing one specific routing protocol.

Decision Variables:

- Load status of nearby RNs
- Available resources at RNs
- Signal strength of nearby RNs, such as the averaged downlink pilot power etc.
- Distance or pathloss between the terminal of interest and nearby RNs

[7] N.B. in a two hop cellular system, the routing is pretty much determined by the cell selection in the terminal.

2.4 User Plane Functionality

2.4.1 Introduction

User plane functionalities are usually used to refer to a group of protocols and algorithms that facilitate higher layer packet transfer through the radio interface. Important requirements for user plane functionalities are reliable (and secure) end-to-end information transfer over the radio access system. Examples of user plane functions are data link layer functionalities such as retransmission protocol and medium access control schemes.

Similar to the case for the control plane functions, it is perceived that many of the user-plane functions may be regarded as mode agnostic may hence be generalized and thereby shared between the different WINNER modes (e.g. error recovery, compression of higher layer protocol headers and ciphering). Hence (as in the previous sub-section) the user plane functions may be divided into generic and mode specific user plane functionalities and thus similar kind of benefits (as for the case of control plane functions) may be encountered.

The next sub-section will outline a few potential key generic user plane functionalities.

2.4.1.1 Potential key generic user-plane functions and their objectives

2.4.1.1.1 Header compression

Header compression is a functionality used to compress higher layer protocol fields to enable high resource efficiency.

Architecture:

The header compression functionality is envisioned to be placed at the top of the data link layer.

Decision Variables:

- May depend on higher layer protocols

2.4.1.1.2 Segmentation and reassembly

Usually the data link layer segments the incoming higher layer packets into (smaller) data link layer protocol data units (PDUs) that are (in some sense) of suitable for radio transmission. These PDUs are subsequently assembled at the receiving side into higher layer packets.

Architecture:

The segmentation and reassembly functionality is envisioned to be placed within GLL in RLC-g (see section 2.5).

The reason for including this functionality within GLL is to be able to perform buffer context transfer when switching from one mode to another.[8] Thereby one is able to ensure seamless mobility between different WINNER modes.

Nevertheless, there are also reasons for why segmentation and reassembly should (in addition?) be placed below GLL. One such reason could be if it is deemed to be important to be able to dynamically adapt the packets size to the momentary channel characteristics.

Another option is to not use segmentation and reassembly at all, i.e. a data link layer PDU is equivalent to a network layer PDU (e.g. IP packet), but this is a subject for further studies.

Decision Variables:

- Frame structure
- Mode capabilities
- Packer error rate

2.4.1.1.3 Retransmission protocol

An important issue that needs to be taken into account is that the information transfer over the radio interface should be reliable and therefore retransmission protocols are important.

Even though the typical reason for radio interface transmission errors are experienced interference and receiver noise, also some user mobility related issues may give rise for data loss. It is indeed possible that

[8] This does not necessarily mean that the buffer context is moved from one node to another.

some data units may be lost whenever the user terminal is handed-off from one node to another. One such scenario occurs when the user terminal is handed off — not only between different access points as in conventional cellular systems — but also between different relay stations and between access points and relay stations and combinations thereof. In that case, some data units may be lost since the transmitter and receiver buffer context cannot be necessarily moved back and forth (over the air) from one transmitting relay node to another relay node (or access point).

Moreover, if the purpose of relay nodes is to help coverage problems, they are likely dispersed far away from the access points. It may therefore happen that most of the hand-offs will be performed between different relay nodes. In addition, the density of relay nodes could be higher than that of access points and therefore the hand-off rate between relay nodes may become higher than the handoff rate between distant access points.

Since the GLL is envisioned to provide a common interface (for all modes) towards the upper layer functions and protocols, GLL level retransmission functionalities could help the above-mentioned problem and ensure reliable information transfer over the radio interface, i.e. from the external network to the user terminal.

Finally, it should be noted that the retransmission functionality should not be mandatory (e.g. may depend on service) and hence multiple retransmission protocol modes should be supported (acknowledged, unacknowledged or transparent mode)

Architecture:

The retransmission protocol is envisioned to be placed within GLL in RLC-g (see section 2.5).

Decision variables:

- Status reports
- Polling
- Sequence numbers

2.4.1.1.4 In sequence delivery

Since error events and subsequent retransmissions may cause PDUs to arrive at the destination out of order. This may also happen when PDUs (belonging to one flow) follow different paths (perhaps even using different modes) from source to destination. Consequently reordering mechanisms are needed when reassembling the higher layer packets.

Architecture:

The in sequence delivery functionality is envisioned to be placed within GLL in RLC-g (see section 2.5).

Decision Variables:

Sequence number check

2.4.1.1.5 Link security

Link security is important to be able to ensure confidentiality and integrity protection of user data (and signalling information) over the data link layer (and may include security key management).

Architecture:

The link security functionality is envisioned to be placed within GLL in either RLC-g or MAC-g (see section 2.5).

Decision Variables:

- May perhaps depend on service type

2.4.1.1.6 Flow control

Flow control refers to a scheme that is used by a node in order to regulate the rate at which the data units enter its buffer. More specifically in a relaying scenario it is generally the case that the propagation conditions differ between different links. Either the link between access point and relay node provides more capacity (e.g. due to line-of-sight propagation, better hardware, etc.) or the link between relay node and terminal provides more capacity (better propagation, less user multiplexing).

The first case may be problematic from a relay buffer point of view since more data might be received by the relay than it is able to forward to the terminal. In this case flow control is required to protect the relay node from buffer overflow.

The flow control mechanism could be integrated with the window mechanism of the retransmission protocol.

Architecture:

The flow control functionality is envisioned to be placed within GLL in RLC-g (see section 2.5).

Decision Variables:

- Channel state information
- Acknowledged PDUs
- Explicit signalling

2.5 Generic Link Layer

As mentioned in the introduction, an important key element in the proposed architecture is the generic link layer (GLL) that aims to generalize common link layer functions for different modes of the radio interface [1]. As such, the GLL may be viewed as a toolbox of link layer functions that may be adapted to the characteristics of the different supported modes in WINNER. Moreover, GLL provides a common (i.e. for all modes) interface towards upper layer functions/protocols.

It is envisaged that GLL may be divided into following two sub-layers *generic radio link control* (RLC-g) and *generic medium access control* (MAC-g). The lower sub-layer (denoted as MAC-g in Figure 2-5) is responsible for prioritization and scheduling between user data flows to/from the same user terminal whereas the upper sub-layer (denoted as RLC-g in Figure 2-5) performs the common link layer functions, e.g.:

- queuing of higher layer data,
- compression of higher layer protocol headers,
- segmentation of higher layer packets into radio blocks,
- in-order delivery of higher layer packets,
- error recovery by means of retransmission functionality and
- Quality of Service monitoring for radio link connections.

Moreover, the RLC-g layer provides a unified interface towards the upper layers, hiding the heterogeneity of the envisioned WINNER modes from upper layer protocols and should a connection be switched between two different WINNER modes, the RLC-g would be responsible for the context transfer between the protocol stacks of the two WINNER modes. More on this will be given in Section 2.7

As already mentioned, there are certain radio link layer functions that remain separate and that are placed in the configuration specific sub-layer (denoted as MAC-r in the Figure 2-5), e.g.

- channel-dependent scheduling of radio resources,
- other functions such as additional retransmission functionality, and
- segmentation into radio condition dependent transmission units.

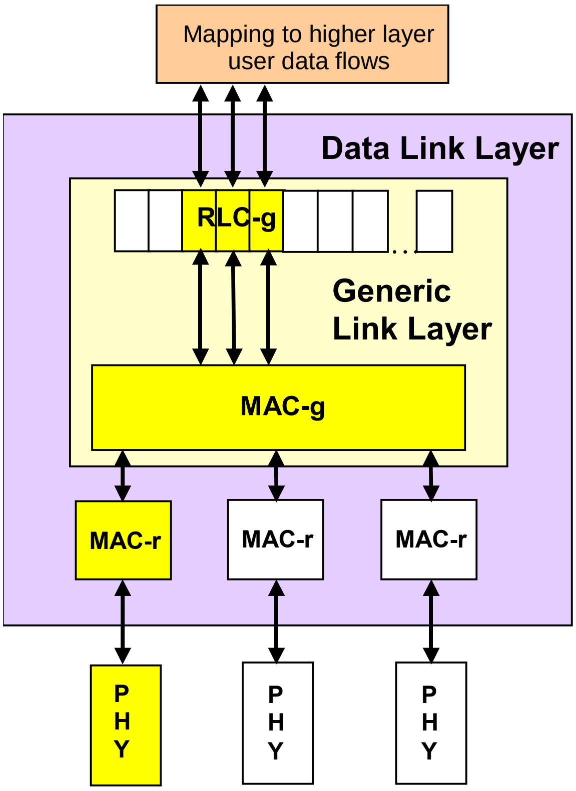


Figure 2-5 GLL protocol stack

As discussed in Section 2.3, there is a possibility for the RRM to assign multiple modes to one specific traffic flow. This may be exploited by GLL to provide multi-mode macro-diversity for the transmission (and possibly also reception) of packets. An example of this involves dynamic scheduling of user packets across multiple modes, whereby instantaneous quality of each radio link determines the utilized accesses. In order to be able to perform the necessary calculations for this, GLL requires up-to-date channel state information (presented in some abstract/generic form by the modes to GLL) from all involved link/modes. As a consequence of this, retransmission may be carried out via a different mode than the one used for the initial transmission. More sophisticated multi-mode macro-diversity techniques may involve simultaneous transmission over multiple modes with the options of either selection or soft-combining at the receiver.

2.5.1 Relay network issues

As described above, GLL acts as a bridge between the different WINNER modes and this may be utilized in a (heterogeneous) relay network by performing forwarding decisions (i.e. routing) at the GLL-level (this is useful for the cases were relaying is performed at layer 2). Moreover, multi-mode macro-diversity relaying (i.e. multiple hops over multiple different WINNER modes) may also be envisioned. An example of this is the case were the AP (RANG) has the possibility to either forward some packets directly (single-hop) to some UTs using mode 1 or to forward the packet over a two-hop path using mode 1 over the first hop and mode 2 over the second hop. To enable this possibility the AP (RANG) needs to be aware of the possible paths that the UT may be reached by and the quality over each link in this path. This topic is further pursued in section 4.3.4.3.

2.6 Functional Layer Architecture

The control plane functions implement system-wide control of the total radio resources and the user plane functions ensure end-to-end control of user data through the WINNER radio access system.

The behaviour of the control plane and user plane functions are defined by the RRM algorithms (being centralized, decentralized or partly decentralized), and the RRM decisions are distributed to the different nodes and the user terminals by the Radio Resource Control (RRC) protocol.

The RRC protocol is a network protocol in the WINNER radio access system and as such should be supported by all nodes implementing the RRM functions in this network. However, certain commands are not applicable for all nodes supporting only a specific WINNER mode. Common configuration commands could be scheduling policy between and inside modes (e.g. fairness criteria between services and users), and specific commands could be spectrum allocation and interference control. Thus, the RRC protocol needs to consist of a generic part RRC-g that is supported by all nodes implementing the RRM functions in the WINNER radio access system and one specific part RRC-r for each WINNER mode supported by the nodes in that WINNER mode. RRC configures the radio link layer, but it also uses the radio link layer to transport configuration commands to remote nodes, as shown in Figure 2-6.

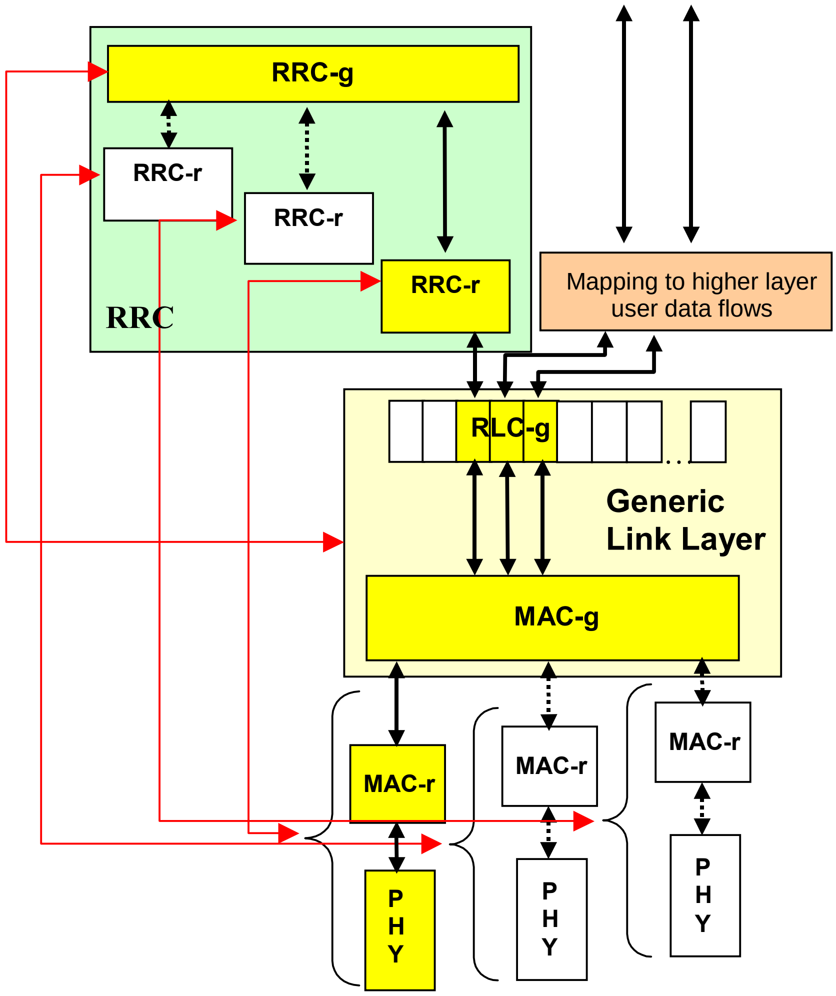


Figure 2-6 Radio interface protocol architecture

The radio link layer defines services to the upper layer both in the control and in the user plane, and consists of a Medium Access Control (MAC) sub-layer and a Radio Link Control (RLC) sub-layer. A convergence protocol sub-layer may also be included in the user plane if it is needed to support a certain user data flow. A goal of the WINNER system design is to support a wide range of deployment techniques, but at the same time keep the complexity low. For this reason, the standard link control functions positioned in the RLC sub-layer should ideally be common to all WINNER modes (RLC-g) and there should then not be a need for a specific RLC for the different WINNER modes (mode specific Hybrid ARQ schemes are seen as a part of the mode specific MAC sub-layer (MAC-r)).

The RRM functions configure and reconfigure the GLL based on e.g. radio resource availability, user terminal capability and admission control. A configuration of the GLL may include room for the GLL to make its own decisions of mapping flows onto WINNER modes based on e.g. information on instantaneous mode capacity and capability received directly from the MAC-r entities in a Radio access system node or in a user terminal (originating in physical layer (PHY)). Hence, the MAC-g sub-layer performs inter-mode scheduling between the modes available for each user data flow based on a generic scheduling metric common to all the modes.

Within a WINNER mode, the MAC-r performs resource allocation of user data flows onto the different transport channels provided by the mode specific PHY layer. MAC-r also communicates status changes in the mode affecting RRM to GLL.

Now, it is possible to refine the WINNER mode definition (in section 1.2) as follows. A WINNER mode is characterized by its MAC-r and RRC-r pair where

- PHY defines a set of resource (in time, frequency etc.) that may be allocated,
- MAC-r controls the usage of a single PHY instance,
- RRC-r controls the set of radio resources available to the associated MAC-r (e.g. power control).

Internally, GLL consists of a number of RLC-g entities, which are connected to a number of MAC-r entities according to a mapping function. Each GLL entity maintains a specific control data channel or user data flow (call or session) as shown in Figure 2-7 and Figure 2-8 respectively. The mapping function also assists in the context transfer during mode switch, and its behaviour is configured by RRM. A MAC-r entity allocates the resources that are available within a specific WINNER mode to the MAC-g entities to which it is connected.

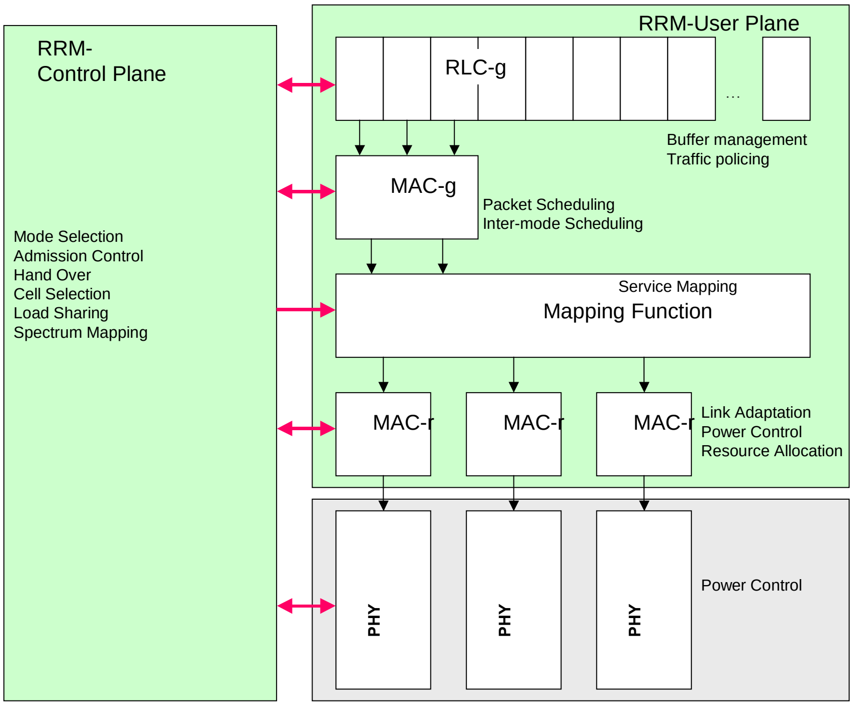


Figure 2-7 Mapping of Control Plane Functionality onto Protocols

In the control plane a reliable datagram service, typically with certain delay requirements, is maintained by each RLC-g entity. In the user plane, user data flows with quite different QoS requirements are to be maintained by each RLC-g entity. In addition to scheduling between user data flows (calls/sessions) to/from different users, MAC-g performs inter-mode scheduling for the different user data flows (maintained by different RLC-g entities) associated with the same user terminal.

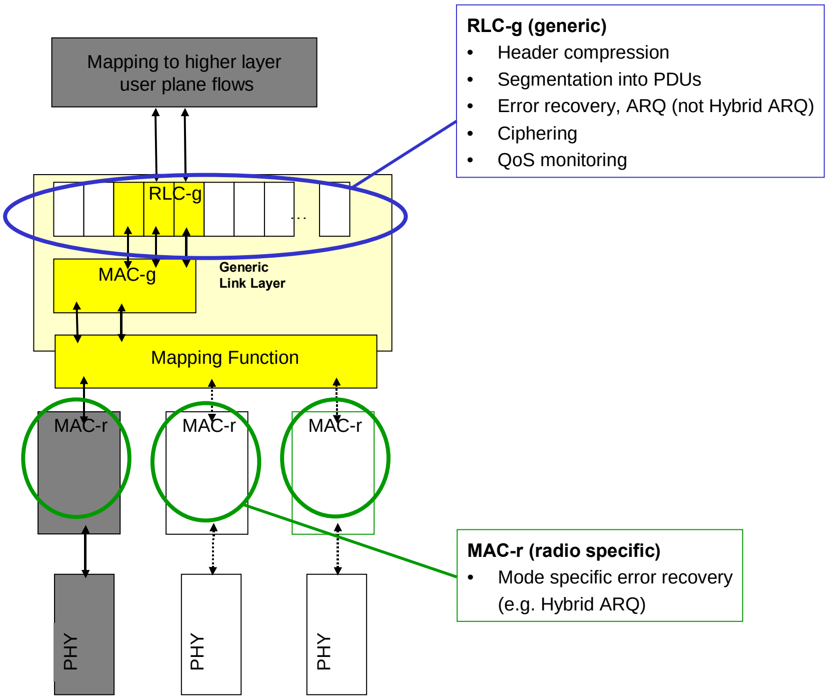


Figure 2-8 Mapping of User Plane Functionality onto Protocols

2.7 Modes Convergence Protocol (MCP)

To support the seamless interworking between different modes, it is necessary to transfer the status of existing traffic flows between the modes (see Figure 2-9). The potential applications for this are multi-mode terminals or multi-mode access points performing inter mode handover.

Prerequisite for the transfer of this status information is a partitioning of the functions inside the involved layers into mode-specific (-R) and mode-independent (-G) functions, as proposed in the previous sections. As shown above and already suggested by [7][8], not only the Link Layer Protocols but also those involved in Radio Resource Management (RRM) will have to be partitioned in this fashion. Figure 2-9 exemplarily illustrates a context transfer on the MAC layer, following the partitioning of functionalities as suggested in 2.6.

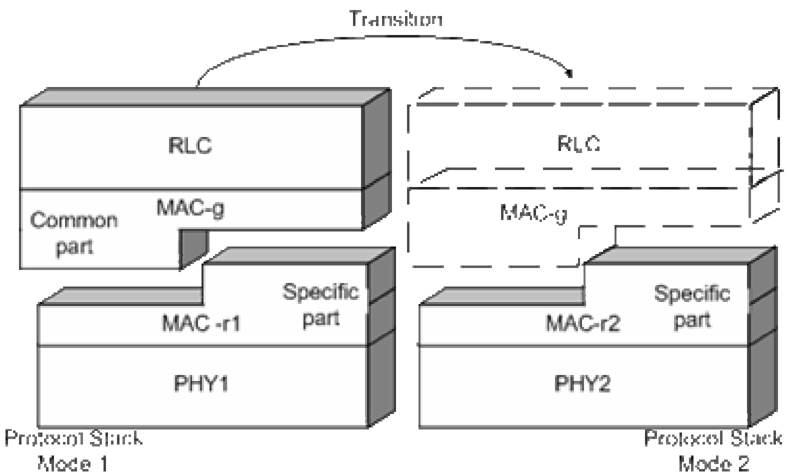


Figure 2-9: Context transfer (e.g. on the MAC Layer) between different WINNER modes

2.7.1 Cross-Plane Optimisation

The ability of a user-plane protocol to re-use status information and protocol data after switching to another mode requires that the functionality of this protocol be extended into the control plane, even though it serves pure user-plane purposes.

As an Example we use here the Status of an ARQ Protocol located in the Link Layer.

The status of e.g. ARQ-bitmaps or other internal data should be preserved when switching to another mode. Some c-Plane functions will have to gather this information and re-apply it when the new modes' specific parts of the EC protocol are in place.

Organising this process efficiently requires a cross-plane (perhaps cross-layer) optimisation, which means that a re-arrangement of functionalities and access to User Plane status data by the control plane has to be done by means of Control-Plane functions common to the different modes (illustrated see Figure 2-10). Effectively, these common functions would have to rely on a well-defined interface towards the mode-specific functions in their layer.

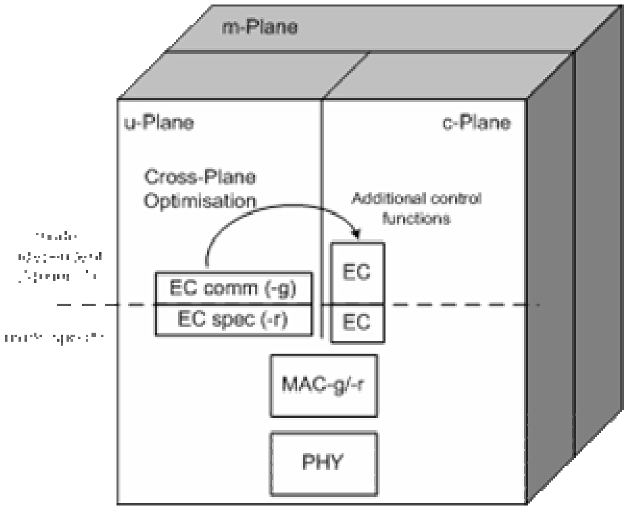


Figure 2-10: Cross-Plane Optimisation and functional partitioning

2.7.2 Cross-Stack Mgmt. with the Modes Convergence Protocol

Under the assumption that the protocol layers are structured and partitioned in the way described in the last section, they would allow to easily switch between different modes.

The transition between the two modes can be visualized as a "virtual extension" of the Management Plane, to achieve a cross-stack management, as shown in

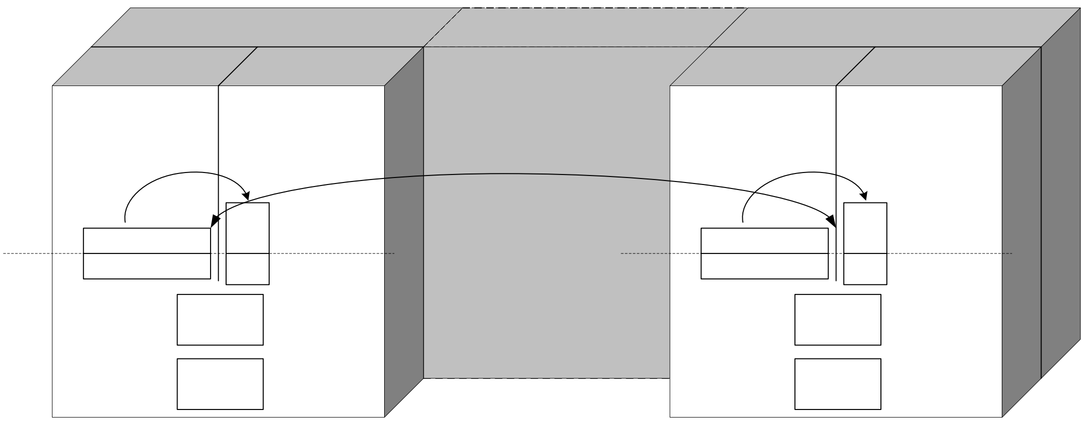


Figure 2-11.

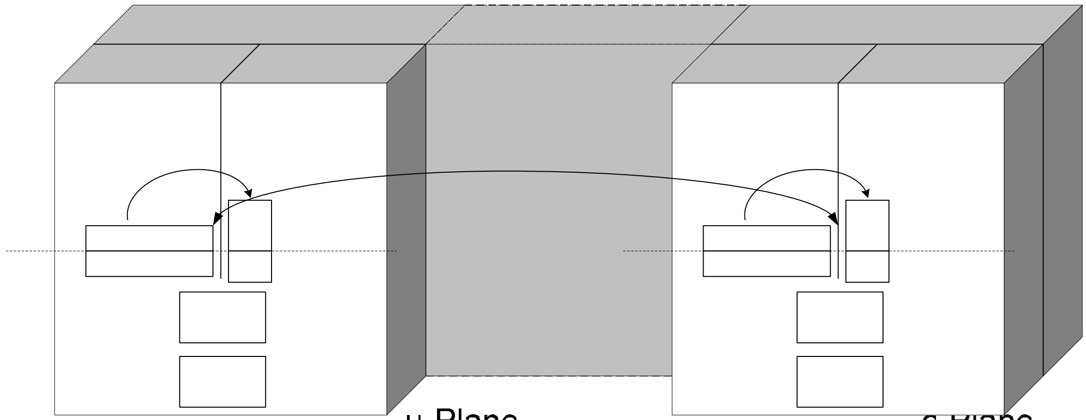


Figure 2-11: Cross-Stack Management

This extension is achieved by a Layer-Specific Modes Convergence Protocol (LSMCP), which exists for each Layer. In the context of the u-Plane and c-Plane functionalities described in Sections 2.3 and 2.4, such Modes Convergence Protocols could exist for the MAC Layer (MAC-MCP), the RLC Layer (RLC-MCP) and the RRC Layer (RRC-MCP) and possibly even for their respective sub-layers.

Figure 2-12 shows as an example how the common functions of the EC (part of RLC) protocol achieve horizontal convergence between the specific functions in mode 1 and 2 by means of the RLC-Modes Convergence Protocol (RLC-MCP).

The information flow is as follows:

1. Protocol Status Data is collected by Control-Plane functions common to all modes
2. The Transfer of Protocol Status Information is performed by the RLC-MCP
3. The status information can be re-used in the new mode.

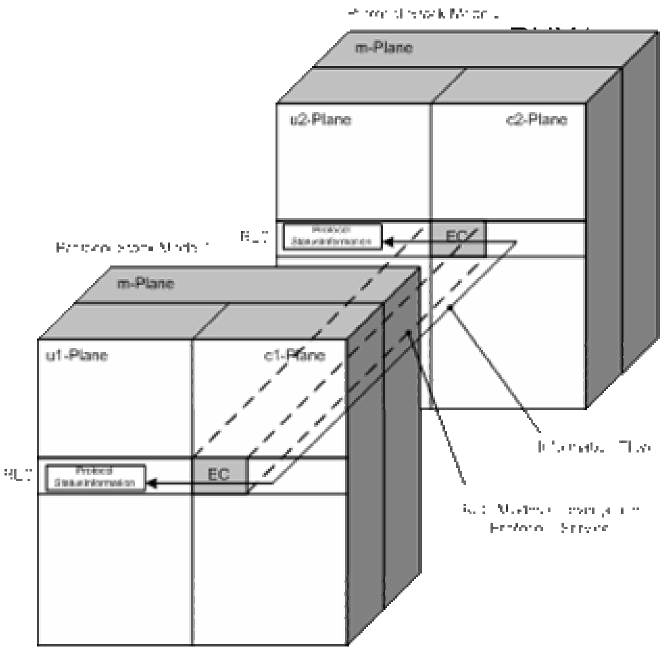


Figure 2-12: The LLC-Modes Convergence Service and Protocol

2.7.3 The Modes Convergence Protocol in the context of G-LL and G-RRM

The view that is usually taken in standardisation is that "generic" functions are not fully functional and have to be enriched by specifying the implementation parts in more detail. Contrary to that, the GLL results from the identification of a common set of Link Layer-functions of different modes, which are then marked with a "-g". As stated several times before, this identification of common functions also has to be performed in other layers. The Layer-specific Modes Convergence Protocols for these Layers have to rely on the "-g"-functions of these layers, i.e. the identified set of common (mode independent) functions.

This makes the existence of Modes Convergence protocols for the sub-layers of the Link Layer (LLC) like Radio Link Control and Medium Access Control (MAC) a prerequisite to implement a Generic Link Layer. The same applies to the implementation of Generic Radio Resource Management with a Radio Resource Control (RRC) Protocol that is divided into common (-G) and specific (-R) parts. The common parts have to rely on a RRC-MCP to exchange status information.

The proposed Modes Convergence Protocols support horizontal convergence of different protocol variants within one protocol layer.

2.8 Logical Node Architecture Model

2.8.1 Introduction and motivation

The main goal of the logical node architecture model presented in this chapter is to assist in grouping functions, between which there may be a need for defining open interfaces. In particular, the logical node architecture needs to support all envisioned deployment scenarios for WINNER (as well as not yet foreseen deployment scenarios) without introducing too many logical nodes and/or interfaces. Note that the list of reference logical nodes presented here is preliminary, and logical nodes may have to be added, removed or combined during the development of the WINNER architecture. Even though the logical node architecture includes protocol termination points, they do not suggest a certain physical placement of functional entities. The physical distribution of protocol termination points is decided when mapping these logical nodes into physical nodes, see section 2.8.3 for examples.

Figure 2-13 proposes a preliminary logical node architecture model. Here, dashed lines indicate control relations whereas solid lines indicate user plane data transport. N.B. the Access Router (AR) is outside the primary focus of WINNER WP3 but is included to be exhaustive.

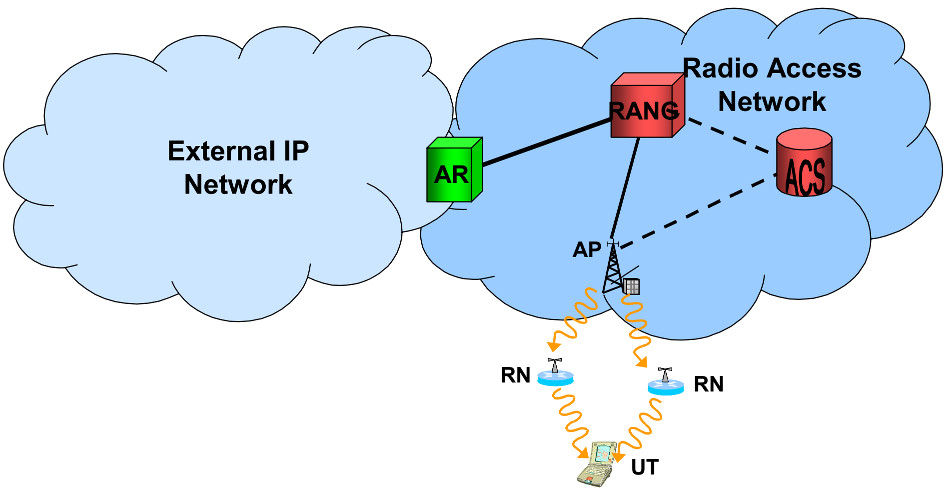


Figure 2-13 Logical nodes

User Terminal Logical Node (UT_{LN}) is a logical node comprising all functionality necessary for it to communicate directly with another UT or the network.

Access Point Logical Node (AP_{LN}) is a logical node terminating the transport network layer protocols on the network side as well as mode specific radio protocols on the user terminal side.

Radio Access Network Gateway Logical Node ($RANG_{LN}$) is a logical network node terminating the data link layer. It terminates generic user plane protocols.

Access Router Logical Node (AR_{LN}) is a logical IP layer node that performs the tasks attributed to an Access Router as defined in relevant IETF specifications.

Relay Node Logical Node (RN_{LN}) is a logical network node with relaying capabilities that is wirelessly connected to an AP_{LN}, UT_{LN} or another RN_{LN}. Hence, one major difference to an AP_{LN} is that it does not terminate the transport network layer protocols. In many cases this classification is not sufficient and the RN_{LN} may hence need to be further partitioned, e.g. depending on whether it is mobile or not (i.e. classified as a Fixed Relay Node (FRN_{LN}) or Mobile Relay Node (MRN_{LN})) or on what layer it is performing forwarding on (e.g. classified as a relay node with layer 3 routing capabilities ($RN_{3,LN}$)). The number of necessary logical relay nodes is not currently known and is left for future work. In the remainder of this section the more detailed partitioning will be used only in those cases where it is deemed necessary.

Access Control Server Logical Node (ACS_{LN}) is a logical network node that controls the access to the radio interface resources. It terminates generic control plane protocols.

Consequently, the radio interface is, on the network side, terminated in the ACS_{LN} and in the $RANG_{LN}$.

The subsections below further explain the roles, functions and interfaces of the different nodes.

Moreover, as stated above it is left for further work to determine if additional logical nodes are necessary (e.g. Moving Gateway Logical Node (MG_{LN}) for the case of moving networks, see section 2.8.5 for more on this) or if any of the above-mentioned logical nodes are redundant (e.g. the RN_{LN} may for certain cases be defined as an AP_{LN} or a combination of AP_{LN} and $RANG_{LN}$, see section 2.8.5 for more on this). As definition of the WINNER system and related protocols progress refinement of the above definitions is likely. The aim of the above definitions is to afford the nodes a distinct identity.

2.8.2 Logical nodes and their relations

2.8.2.1 User terminal logical node

The user terminal logical node comprises all functionality necessary for an end user to communicate with either another UT_{LN} or a network. Moreover, the AP is the node closest to the core (backbone) network that a UT may be (directly) connected to.

The user terminal logical node may be connected to one or several RN_{LN} and/or to one or several AP_{LN}. In addition the UT will be the protocol peer of the $RANG_{LN}$[9], the ACS_{LN} and the AR_{LN}.

2.8.2.2 Access point logical node

The AP_{LN} is envisioned to terminate and handle the physical layer functions as well as the mode specific link layer functionality. Additionally the AP may act as a proxy between the UT_{LN} and the ACS_{LN} for certain protocol transactions. The AP_{LN} may be connected to one or several $RANG_{LN}$ and one ACS_{LN}.

2.8.2.3 Radio access network gateway logical node

The $RANG_{LN}$ terminates the link layer on the network side. It acts as the link layer of the AR towards the UT_{LN}s. One $RANG_{LN}$ can connect to a UT_{LN} via one or several AP_{LN}. As there is a need to be able to change the serving $RANG_{LN}$ of a UT_{LN} by relocating the layer-2-context information, an interface between $RANG_{LN}$s is required. Moreover, the $RANG_{LN}$ is connected to an ACS_{LN}, and to an AR_{LN}.

2.8.2.4 Access router logical node

The Access Router acts as a normal Access Router per relevant IP layer specifications.

2.8.2.5 Access control server logical node

The ACS_{LN} controls one or several AP_{LN} and coordinates RRM with corresponding functions in the UT_{LN}. Moreover, it may also coordinate resources between neighbouring AP_{LN}s.

2.8.2.6 Relay node logical node

The task of an RN_{LN} is to perform relaying between AP_{LN}, other RN_{LN} and UT_{LN}. For some scenarios the RN_{LN} may be equivalent to a combined UT_{LN}, AP_{LN}, ACS_{LN} and $RANG_{LN}$. In these scenarios the RN_{LN} on one side would connect as a UT_{LN} to the AP_{LN}, ACS_{LN} and $RANG_{LN}$, whereas the other side acts as a combination of AP_{LN}, ACS_{LN} and $RANG_{LN}$ towards another RN_{LN} or UT_{LN}. It is for further study if this applies to all relevant scenarios and if a valid consequence is that the RN_{LN} can be removed from the reference architecture.

The RN_{LN} may be connected to one or several RN_{LN} and/or to one or several AP_{LN}, one or several $RANG_{LN}$, one or several ACS_{LN} and one or several AR_{LN}.

[9] One user flow is always associated with one $RANG_{LN}$, however one $RANG_{LN}$ may hold several flows and hence, one UT_{LN} may be associated with several $RANG_{LN}$.

2.8.2.7 Interfaces

The nodes: UT_{LN}, AP_{LN}, ACS_{LN} and $RANG_{LN}$ result in the following logical interfaces which are of special interest from an architecture perspective:

$UT_{LN} \leftrightarrow ACS_{LN}$

$UT_{LN} \leftrightarrow RANG_{LN}$

$UT_{LN} \leftrightarrow RN_{LN}$

$AP_{LN} \leftrightarrow ACS_{LN}$

$AP_{LN} \leftrightarrow RANG_{LN}$

$AP_{LN} \leftrightarrow RN_{LN}$

$RANG_{LN} \leftrightarrow RANG_{LN}$

$RANG_{LN} \leftrightarrow AR_{LN}$

$ACS_{LN} \leftrightarrow RANG_{LN}$

$ACS_{LN} \leftrightarrow ACS_{LN}$

2.8.3 Physical implementations using logical nodes

In Figure 2-14, several possible realizations that are based on the previously defined logical network nodes are exemplified. The purpose of these examples is to explain how logical nodes can be combined into different physical implementations in order to cope with different operational scenarios envisioned within WINNER. In addition, they also demonstrate how completely different solutions that are aimed for different scenarios are used without introducing an excessive amount of nodes and interfaces.

The logical network nodes can be implemented separately but combinations of some or all of them are also possible. It can be desirable to separate AP_{LN} from the ACS_{LN} and $RANG_{LN}$ in scenarios where the spectrum is the limiting factor and the user mobility is high. In such scenarios the spectrum efficiency calls for centralized ACS_{LN} and the high mobility profits from centralized $RANG_{LN}$. On the other hand, in scenarios with low mobility and vast access to spectrum, it may be preferred to co-locate ACS_{LN} and $RANG_{LN}$ with an AP_{LN}.

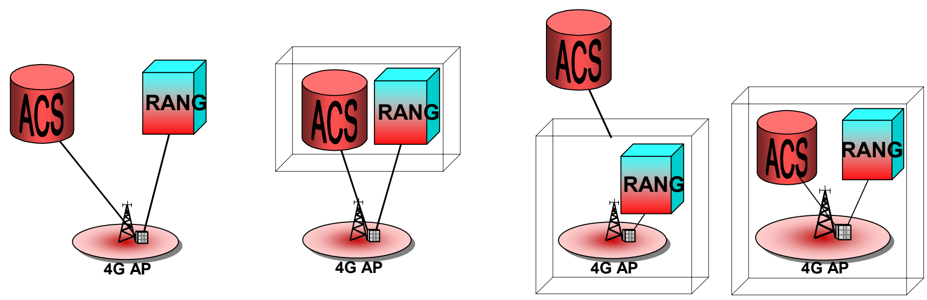


Figure 2-14 Illustration of possible physical node realizations using logical nodes

- In the first example (in Figure 2-14), AP_{LN}, ACS_{LN} and $RANG_{LN}$ are implemented in separate physical nodes. The main benefit is that the functionality is centralized and therefore buffer management and radio resource allocation is efficient. On the other hand, the GLL retransmission protocol round trip time can be negatively influenced by the transport network delays. By locating the user plane functions into the $RANG_{LN}$ and the control plane functions into the ACS, we achieve independent scalability, dimensioning and evolution of the user and control planes.
- In the second example, the ACS_{LN} and $RANG_{LN}$ are placed in the same physical node, which allows more efficient user and control plane information exchange than in the previous example.
- In the third example, $RANG_{LN}$ and AP_{LN} are located in the same physical node whereas the ACS_{LN} is sustained in some central node in the radio access network. The user plane is thus moved closer to the physical layer which can be beneficial with respect to the GLL retransmission protocol round trip time. A drawback is that the buffer management at high user mobility is awkward. In turn the control plane is centralized and therefore the radio resource management can be efficient. However the delay requirements on the transport network can be high.

- In the last example, all logical nodes are implemented in the same physical node. Since both the control and user planes are moved close to the radio interface the buffer management at high user mobility can be difficult and the radio resource management may become inefficient. However the GLL retransmission protocol round trip time is independent of the transport network layer delays.

It is assumed that the AR_{LN} is located in a central network node. In the two first examples (where the $RANG_{LN}$ and AP_{LN} are located in different physical nodes) AR_{LN} can be co-located in the same physical node as the ASG. In the two last examples (where the $RANG_{LN}$ and AP_{LN} are co-located) AR_{LN} and $RANG_{LN}$ are thus located in different physical nodes.

2.8.4 Protocol terminations in logical reference nodes

In this section, some examples on protocol termination in the logical nodes of the model previously described are shown for single-hop and multi-hop scenarios. The protocol termination examples are results of mappings between the functional layer protocol architecture presented in section 2.6 to the logical node model. The list of examples is not claimed to be complete, nor is it claimed that all examples are equally probable. Moreover, the examples are presented on a very general level and do not show any details of how user plane and control plane are internally sub-divided. Further refinements, and possibly also revisions, will have to be made during the development of user plane and control plane functionality and their mappings.

All figures in the following sub-section focus on the radio interface protocol termination points. For completeness a Transport Network Layer (TNL) is indicated in all figures representing the protocols used to carry the radio interface protocols between infrastructure nodes, i.e. the protocols of the transport network. To connect the AR_{LN} and the $RANG_{LN}$ a layer 2 tunnel (L2T) is used, which can be based on any appropriate (set of) protocol(s). However, the actual protocols in the TNL as well as L2T are out of the scope of WINNER WP3.

2.8.4.1 User Plane

2.8.4.1.1 Single hop case

In this example a UT_{LN} is directly connected to an AP_{LN}, i.e. there is no intermediate RN_{LN}. On the infrastructure side the $RANG_{LN}$ terminates the GLL, whereas the mode specific L2 parts is terminated in the AP_{LN} (however, parts of the mode specific L2 parts may also be terminated in the $RANG_{LN}$ as previously mentioned)

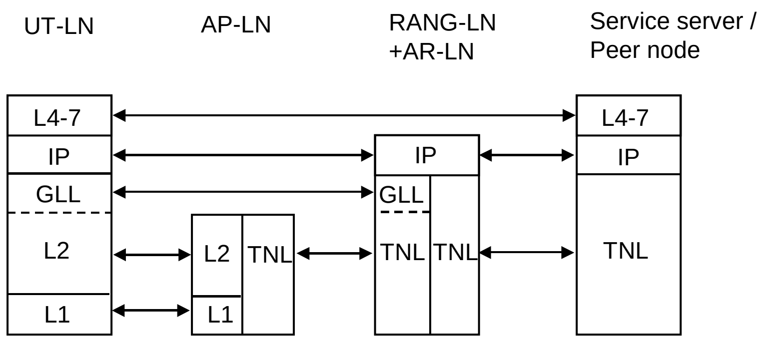


Figure 2-15 Single-hop case; User plane protocol terminations

2.8.4.1.2 Multi-hop case

The following figures depict the cases were a RN_{LN} is included in the communication path. Here, forwarding may be performed either on L1, L2 (mode specific forwarding), GLL (generic forwarding) or L3 between UT_{LN}, RN_{LN} and $RANG_{LN}$ (N.B. since forwarding is performed at different layers they may actually be different logical nodes). This results in GLL peer-to-peer relations between UT_{LN}, RN_{LN} and $RANG_{LN}$.

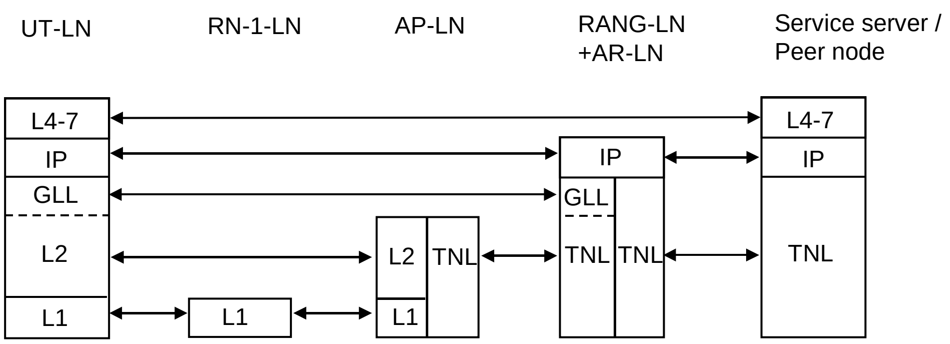


Figure 2-16 L1 level relaying; User plane protocol terminations

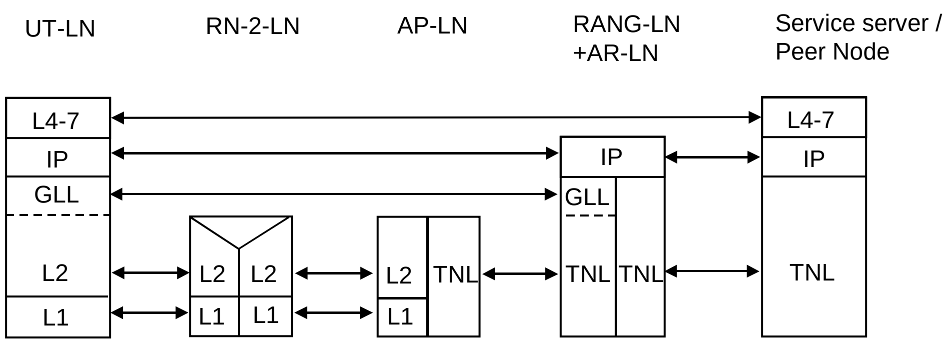


Figure 2-17 L2 level relaying; User plane protocol terminations

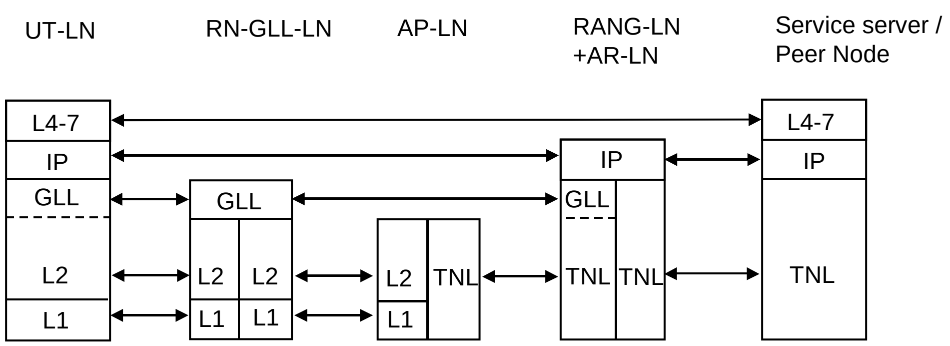


Figure 2-18 GLL level forwarding; User plane protocol terminations

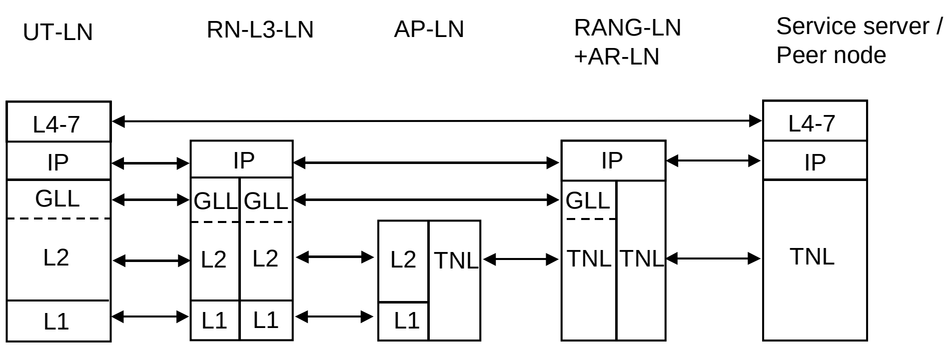


Figure 2-19 L3/IP level relaying; User plane protocol terminations

2.8.4.2 Control plane

The control plane examples shown below focus on the relation between the UT_{LN} and the infrastructure nodes and do not show all relations between the infrastructure nodes themselves (such as between ACS_{LN} and AP_{LN} as well as between ACS_{LN} and $RANG_{LN}$). For various reasons (explained above) it is envisioned that one goal for the WINNER air-interface is to terminate RRM functionality as close to the air interface as possible.

2.8.4.2.1 Single-hop case

The L2, GLL and L3 RRM functionality is terminated in AP_{LN}, $RANG_{LN}$ and ACS_{LN} respectively. The need for separated protocols regarding L2, GLL and L3 RRM functionalities need to be worked out further. Therefore the figure below should be viewed as just one possible example. Here, it is also assumed that the L3 RRM signalling is carried by GLL between the UT_{LN} and ACS_{LN} and hence the signalling has to pass through the $RANG_{LN}$.

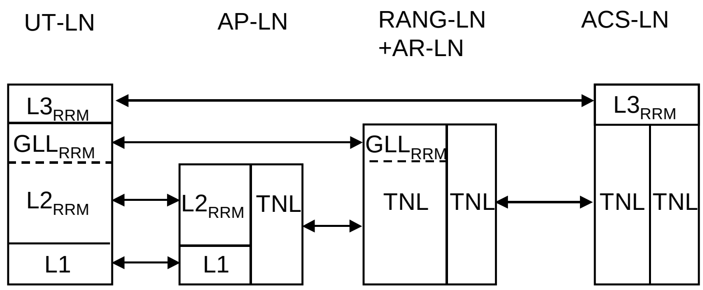


Figure 2-20 Single-hop case; Control plane protocol terminations

2.8.4.2.2 Multi-hop case

The L2, GLL and L3 RRM functionality is terminated in AP_{LN} or RN_{LN}, $RANG_{LN}$ and ACS_{LN} respectively. As for the single-hop case, the need for separate protocols regarding L2, GLL and L3 RRM functionality need to be worked out further, and hence the figure below should merely be viewed as one possible example.

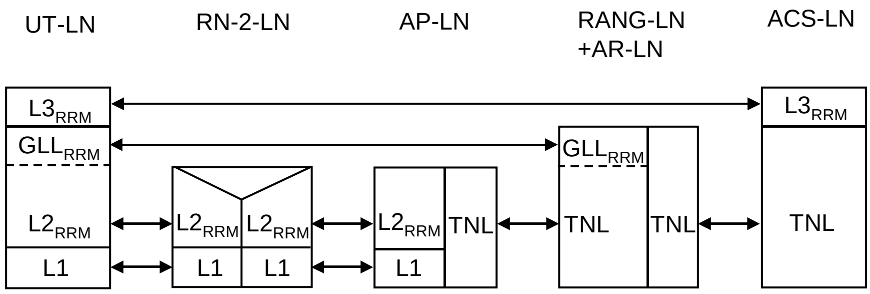


Figure 2-21 Multi-hop case; Control plane protocol terminations

2.8.5 Reduction of number of logical nodes in relay and moving networks

This subsection will outline a few examples of how nodes in relay networks as well as moving networks may be composed of the previously introduced logical nodes without introducing any logical relay nodes or logical moving network nodes.

2.8.5.1 Relay networks

As previously stated, even though the RN_{LN} is defined as a separate logical node it can, in some implementations, also be considered as a combination of other logical network nodes. Two such examples (taken from section 2.8.4.1.2) are shown in the figures below. In Figure 2-22 the relay node may be seen as a combination of an AP_{LN}, $RANG_{LN}$ and AR_{LN} on the terminal side whereas it may be seen as a UT_{LN} on the network side (henceforth referred to as option 1). From a user plane perspective option 1 corresponds to having several RANs in cascade. That is, going from left to the right in the figure the UT_{LN} connects to the leftmost RN and considers that as a RAN. The RN in its turn connects to the next RN (if any) and sees it as a new RAN. This continues until one reaches the AP_{LN} of the stationary RAN, which connects to the $RANG_{LN}$ using any kind of Transport Network Layer (TNL), e.g. IP or ATM.

There are some potential issues with this kind of architecture, since layer 2 contexts such as buffer management and security procedures (like ciphering) are handled on a per hop basis. If the route through the multi-hop radio network is changed, e.g. a RN is added or deleted, the layer 2 context has to be transferred to the new node to avoid loss of data and guarantee a ciphered connection. In addition the UT has to trust the intermediate RNs since they handle un-ciphered data if not end-to-end security solutions are used.

The transfer of layer 2 context can be avoided by using option 2 (see Figure 2-23) where the relay network may be seen as an ordinary TNL. In this case, the first RN seen from the UT_{LN} acts on one side as an AP_{LN} for the user terminal and on the other side it acts as an UT_{LN}. The RN connects to the next RN (if any), which is seen as RAN on one side and a UT_{LN} on the other side (as in option 1) and this continues until the AP_{LN} of the stationary RAN is reached. The AP_{LN} connects to the $RANG_{LN}$ of the stationary RAN, which terminates the remaining parts of layer 2 (including GLL) of the UT_{LN}. Consequently, security and reliability is maintained between the $RANG_{LN}$ and UT_{LN} across the multi-hop path avoiding the difficulties described for option 1.

Only the user plane in the relay network has been addressed so far. In a similar way, the control plane can be handled in two ways, either each RN contains an ACS_{LN} or a set of RNs are controlled by one ACS_{LN}. Which solution is the best depends on the requirements on radio resource efficiency and the deployment scenario.

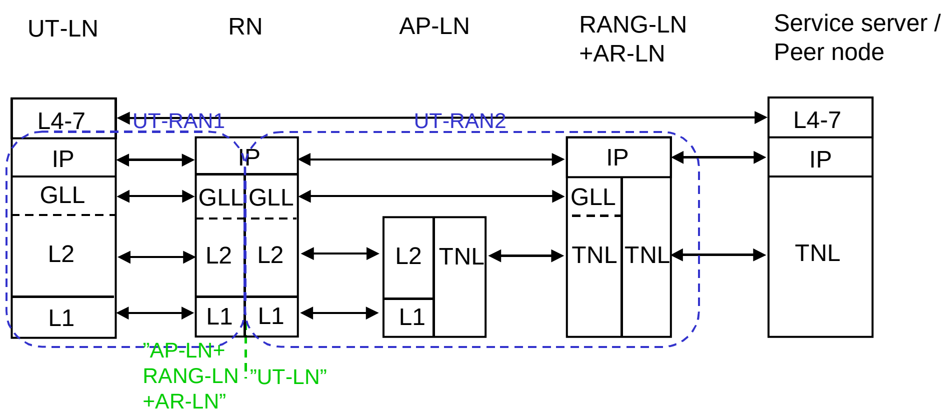


Figure 2-22 User plane protocol stack option 1 for a relay network

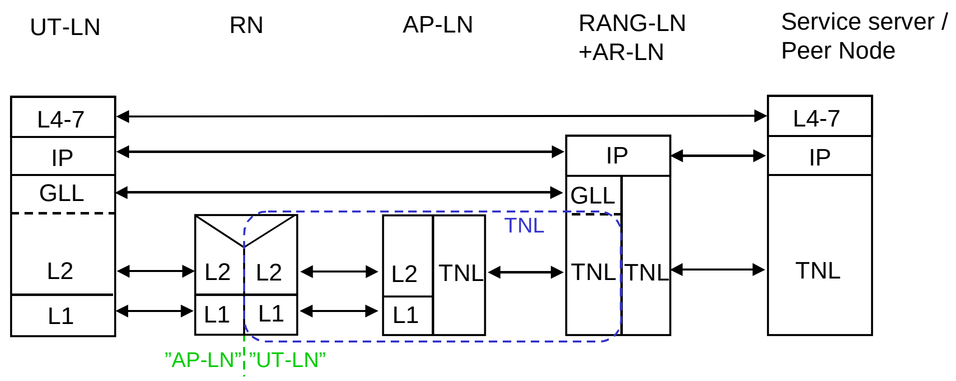


Figure 2-23 User plane protocol stack option 2 for a relay network

2.8.5.2 Moving networks

This sub-section will give one example of how the previously described logical nodes may be combined to compose a moving network. In this example the moving network is seen as a moving combination of AP_{LN}, $RANG_{LN}$ and ACS_{LN} (for simplicity in the following referred to as moving $RANG_{LN}$) and is illustrated in Figure 2-24. A moving $RANG_{LN}$ allows the possibility to have multiple AP_{LN} (and hence physical APs) within the moving network. The problem with the move of layer 2 context is not as severe as for the relay network, since only a single hop is involved.

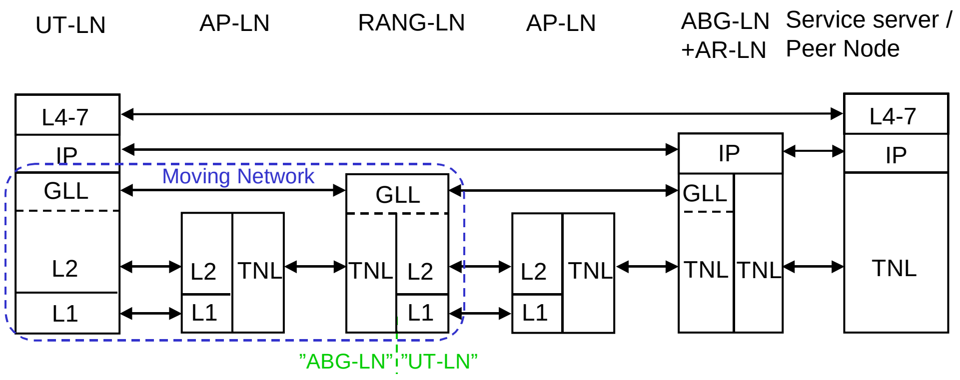


Figure 2-24 User plane protocol stack for a moving BG

2.9 Envisioned Physical Nodes

2.9.1 Envisioned deployment scenarios and supporting physical nodes

2.9.1.1 Network architecture description

The example architecture depicted in Figure 2-25 comprises several different network elements called physical nodes throughout the following. In general two different modes of network operation can be distinguished.

The first operational mode relies on a certain network infrastructure similar to today's cellular mobile communications systems. It shall therefore be named as infrastructure-based mode. Elements of this infrastructure are the Fixed Routing Relay Nodes (FRRNs) depicted on the right hand side of Figure 2-25.

Transmissions between FRRNs and FRRN and Access Point (AP) may take place in different transmission resources. Because visibility (line-of-sight, LoS) between FRRNs can be guaranteed, it is possible to utilize a higher frequency band than that used for the other transmissions. While transmission in the same frequency band may be useful during initial network start, multi-hop transmission may be shifted to a different band if greater network capacity is required. In this concept it is assumed that a homogeneous air-interface is used for both type of communication (fixed to fixed, mobile to mobile) because of the generally cheaper implementation.

The second mode is characterized by its ad-hoc behaviour. This mode does not rely on any network infrastructure and shall therefore be named as infrastructure-less mode (depicted on the left hand side of Figure 2-25). The role of forwarding data packets is taken over by mobile devices that are called Mobile Routing Relay Nodes (MRRN) in the sequel, which could be subscriber devices as User Terminal (UT) are. In addition UTs are able to directly communicate with one another without intervention of any intermediate node. Because fixed network elements are not mandatory in this mode, connectivity can not be guaranteed and will be a problem if no fixed elements are available or cannot be reached since routing capable terminals are not available. Therefore, a combination of infrastructure and infrastructure less mode is likely and shown in Figure 2-25. In this architecture the infrastructure is dynamically extended via MRRNs. This flexible mode of operation allows to extend the coverage and even to increase the spectral efficiency on demand. With increasing user densities it becomes possible that this architecture scales towards higher overall system throughput and larger coverage by means of mobile multi-hop extension. By this the MRRNs become part of the extended infrastructure-based architecture. However, it is worth to mention a number of problems that need to be addressed if the pure infrastructure-less mode is used. Since user equipment is used for forwarding data and valuable battery power is consumed by other users, a decent beneficial system for users allowing their equipment to act as relay needs to be introduced. Furthermore security and authorization issues need to be addressed.

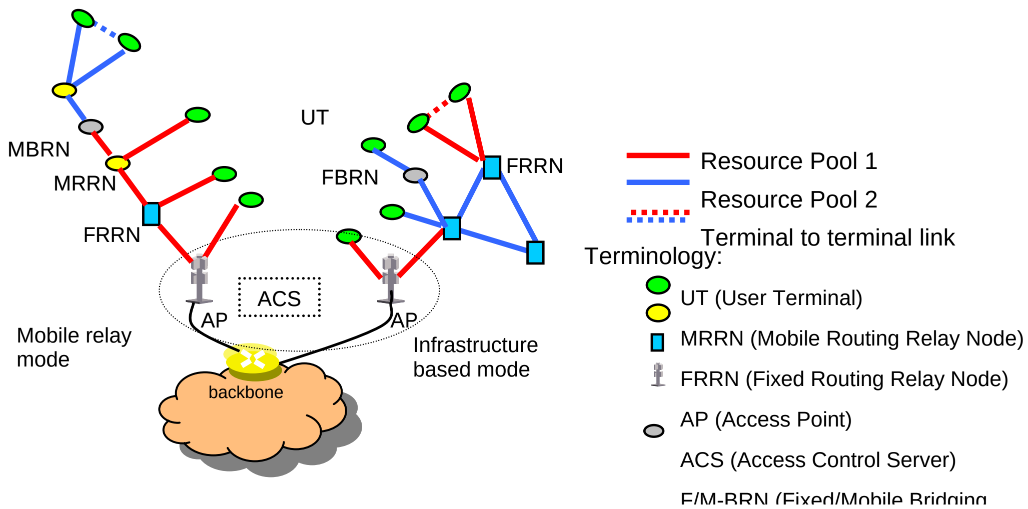


Figure 2-25 Proposed architecture

2.9.1.2 Fixed Network Elements

Four fixed network elements have been defined for this architecture proposal. With respect to the definition of the logical nodes given in the section above it is possible to map the logical node functions to the physical nodes in the following manner:

AP_{LN} = AP

ACS_{LN} = ACS

RN_{LN} = FRRN

RN_{LN} = FBRN

2.9.1.2.1 Access Point (AP)

The AP provides access to the fixed line part of the Radio Access Network (RAN). On the air interface side it coordinates the access to the transmission medium (i.e. time, frequency, code or space) of UTs in a centralized manner. This function is similar to the Central Controller (CC) functionality in Hiperlan/2. In order not to interfere with other APs or F/MRRNs, which use the same physical transmission resource, the AP receives RRM information from the Access Point Controller that indicates the transmission resources available for the AP's CC.

An additional AP functionality might be Admission Control (AC), which controls overall network load by rejecting resource requests in case of excessively high loads.

2.9.1.2.2 Access Control Server (ACS)

The ACS is in charge of synchronising APs attached to it. By this synchronisation it can be ensured that different portions of the RAN can be assigned orthogonal transmission resources. The level of synchronisation needed (i.e. frame-, slot, symbol, chip-level synchronisation) depends on the physical layer transmission scheme and is to be defined.

Furthermore, important parts of the RRM functionalities reside in the ACS. The ACS RRM ensures that AP's, FRRNs and MRRNs with overlapping coverage areas are assigned orthogonal resources. I.e., the RRM ensures and controls the spatial re-use of the transmission resource throughout the network deployment area.

2.9.1.2.3 Fixed Multi-hop Node (FRRN)

A FRRN can be regarded as AP without connection to the backbone NW. It controls the access to the medium of attached UTs. The FRRN might explore simultaneously different resource pools in case of concurrent transmission to FRRNs/APs and UTs.

The FRRN is likely an operator-owned device and part of the fixed network infrastructure. Since FRRN will presumably be the dominant elements in the NW infrastructure self-configuration and self-organization possibilities are likely to be required for these elements.

FRRN antennas are located at dedicated positions, e.g. at lamp posts or on top of traffic lights allowing to "illuminate" specific areas such as street canyons and places. FRRN are assumed to be located within their mutual coverage areas.

2.9.1.2.4 Fixed Bridging Node (FBRN)

A FBRN does not contain CC and scheduling functions.

A FBRN is used to enable a lean coverage solution without additional resource partitioning in the infrastructure based mode by forwarding data packets from a AP/FRRN to UTs and vice versa. It can be regarded as FRRN without CC functionality where scheduling is provided by the corresponding AP/FRRN.

2.9.1.3 Mobile Network Elements

Three mobile network elements have been defined for this architecture proposal. With respect to the definition of the logical nodes given in the section above it is possible to map the logical node functions to the physical nodes in the following manner:

AP_{LN} = MAP

RT_{LN} = MRRN

RT_{LN} = MBRN

UT_{LN} = UT

2.9.1.3.1 Mobile Access Point (MAP)

The MAP provides access to the fixed network part of the Radio Access Network (RAN). On the air interface side it coordinates the access to the transmission medium (i.e. time, frequency, code or space) of UTs in a centralized manner, similar to the Central Controller (CC) functionality in Hiperlan/2. In order not to interfere with other (M) APs or F/MRRNs, which use the same physical transmission resource, the AP receives RRM information from the Access Point Controller that indicates the transmission resources available for the AP's CC.

An additional AP functionality might be Admission Control (AC), which controls overall network load by rejecting resource requests in case of excessively high loads.

It needs to be noted that the MAP concept needs further studies concerning

- differentiation of MAP and movable (heterogeneous) relays
- differentiation between MAP and AP in general
- use cases
- …

2.9.1.3.2 Mobile Multi-hop Node (MRRN)

A MRRN can be regarded as AP without connection to the backbone NW. It controls the access to the medium of attached UTs. Depending on the MRRN capabilities if may or may not explore simultaneously different resource pools in case of concurrent transmission to FRRNs, APs, other MRRN and UTs.

The MRRN is likely a subscriber-owned device. Self-configuration and self-organization possibilities are required because of the attributes of these elements.

It cannot be assumed that MRRN will always have the connection to an access point.

2.9.1.3.3 Mobile Bridging Node (MBRN)

In the infrastructure less mode a MBRN ensures the connectivity between MRRNs in cases where mutually exclusive resources are used (e.g. frequency bands). In these cases a MBRN acts as store and forward device.

A MBRN may in addition determine the level of synchronisation between MRRNs and propagates corresponding synchronisation control messages to nodes further away from the respective AP/FRRN in the infrastructure based mode.

2.9.1.3.4 User Terminal (UT)

UTs are subscriber owned devices which terminate the multi-hop connections. Their access to the transmission medium is controlled either by the AP's, the MRRN's or the FRRN's CC functionality which they are attached to.

2.10 Conclusion and Further Work

In this chapter, a preliminary proposal for the WINNER radio interface architecture is presented. Logical protocol architecture is proposed for the WINNER radio interface and both user plane and control plane functionalities are discussed. The usage and deployment of logical network nodes is exemplified with several physical implementations that are aimed for different operational scenarios.

One important feature is that the proposed architecture can make use of multiple radio interface operating modes. Consequently, different sets of multiple access protocols, duplex techniques and radio link control protocols can be used for different operational scenarios. Hence, the proposed protocol architecture can support a large variety of completely different deployment scenarios such as wide-range and short-range communication as well as integration of conventional cellular system infrastructures with multi-hopping and relaying. These radio interface modes may be viewed as co-designed radio interface technologies optimized for different deployment and user scenarios that can interact seamlessly.

Moreover, the concept of generic link layer (GLL) is adopted to enable cooperation of individual modes and easy integration of future modes. The main principle is to generalize and undertake all common link layer functions for different modes of the radio interface and also provide a common (i.e. for all modes) interface towards upper layer functions/protocols. The GLL may be viewed as a toolbox of link layer functions that may be adapted to the characteristics of the different supported modes.

The radio interface functions are further divided into user plane and control plane functions for the sake of scalability. By locating the user plane and control plane functions into different network nodes, independent scalability, dimensioning and evolution of the user and control planes is achieved.

The concept of the Mode convergence Protocol (MCP) has been proposed to enable a well structured design of B3G protocols of layer 2 and 3 according to the various operating modes based on the same split of generic and specific user and control plane functionalities. This MCP allows by means of cross layer optimisation and cross plane management a context transfer for an efficient handover between two radio modes. The MCP allows e.g. the transfer of status information of the ARQ mechanism.

A flexible logical node architecture model is also proposed to reduce the inter node interface definition, standardization and maintenance effort. It is envisioned that the logical architecture model can result in a lower number of distinct interfaces than conventional (restrictive) deployment models. It is desirable that the node architecture can support all envisioned deployment scenarios for WINNER without introducing too many logical nodes and/or interfaces.

The future work is highly dependent on the development of the WINNER radio interface and its different operating modes. As already mentioned, the architecture proposal is still preliminary and therefore some effort should be concentrated to clarify the radio interface specificities, its architecture, its functional elements and functional distribution.

The details of the RRM function integration in WINNER are a subject for future work. In a near future, it should be addressed the integration of the different radio access technologies in collaboration with WINNER WP4. Each of the RATs has its own specificities and works using specific protocols. The integration process may show the need of the development of new protocols understandable by all the technologies. All the control plane information exchanged between the different entities should be normalized regarding the technology used. Each entity should maintain its RRM components, but it should add new functionalities to take care of the Common RRM. Also, a study about the interferences between RATs should be achieved to see in what situation different radio technologies might or might not live together in the same area. Therefore, the different RATs using the proposed common protocols should be evaluated, in order to take more in depth conclusions.

Despite the fact that relay functions could bring some extra performance, in the operator point of view, there is the risk that QoS control problems arise in communication involving these moving relays. This happens because the operator cannot control the movement of the relays neither the number of moving relays in a given area. Therefore, the operator cannot control the QoS in communications where relay stations are used.

An extra effort is needed in order to carry out the seamless interworking of the different radio access networks. Then, at the end, customers may finally have the benefit of using their services, which should run with the required performance, within several environments made of several radio technologies.

3. Enhanced Conventional (single-hop) Deployment Concepts

3.1 Introduction

Although the main focus of WINNER WP3 is set on relay based deployment concepts it is important to investigate the potentials of single-hop deployment in the context of WINNER as well. Therefore as first step the current radio systems in the area of wide range and short range have been discussed.

This section will provide a description of some enhancement to current systems like the existing 3G (HSDPA), WLAN IEEE 802.11, then well known WLAN HiperLan/2 and the emerging WMAN 802.16a, that will in a first preliminary, assessment be compared against the developed benchmarks with respect to the identified scenarios in D7.2. A description of the discussed systems and the related benchmarks can be found in the Annex I of this document.

3.2 Possible Enhancements to HSDPA

3GPP release 5 identifies several phases of HSDPA evolution. The latest phase will see the introduction of a new air interface to HSPDA to increase the average bit rate:

- OFDM physical layer in combination with higher modulation schemes and array processing.
- MAC-hs/OFDM with fast scheduling to optimize performance by selecting dedicated sets of sub carriers for each mobile according to the quality of the air interface.
- Multi-standard MAC as a control entity to realize fast switching between OFDMA and CDMA channels.

HSDPA provides a significant cell capacity gain for packet data traffic in WCDMA and is thus an important part of the continuous 3G evolution. Since the HSDPA concept offers improved code efficiency and dynamic range in user data rates, it can utilize improvements in detector performance foreseen in the future. Hence, it may be viewed as an enabler for more advanced communication techniques, including equalizers, multi-user or multi-code interference cancellation, as well as advanced MIMO techniques. The HSDPA concept can be introduced gradually in the network with incremental introduction of advanced packet scheduling and link enhancement strategies. The performance and cost/complexity issues of further improvements will be considered within future 3GPP standardization framework to further evolve the WCDMA concept.

3.3 Possible Enhancements to IEEE 802.11

In this section, the requirements and potential technical solutions to improve the 802.11 system are briefly reviewed. Several proposals are also examined and assessed in terms of criteria including throughput, power-consumption, complexity and range.

The use of several antennas at the transmitter, at the receiver or at both sides is known to significantly increase the theoretically achievable capacity of a wireless system (see for instance [14]), and this, of course, also in the framework of OFDM-based schemes [15]. Indeed, in wireless communication systems, multiple antennas provide spatial diversity, which can mitigate the effects of multipath-induced fading and alternatively, multipath can also be exploited to increase link capacity with multiple-input multiple-output (MIMO) or spatial multiplexing techniques.

Currently, in 802.11a, selection diversity can be applied at the Access Point by selecting between two antennas the one that experiences the best Signal to Noise Ratio. This very simple scheme improves the coverage but still relies on a single Radio-Frequency (RF) module and a switch for antenna selection.

With the addition of a complete RF stage per available antenna, more complex digital signal processing can be envisaged in order to fully exploit the multi-antenna capabilities in terms of diversity or multiplexing gains. With the objective of increasing the peak bit-rate, spatial diversity can be used in order to increase the robustness of the transmission of higher order modulations that are costly in terms of RF requirements, whereas spatial multiplexing enables to transmit simultaneously several parallel data flows with lower modulation order, but with additional inter-antenna interference. The latter can be reduced by interference cancellation techniques such as BLAST.

For instance, a two transmit antenna (MT=2) wireless link with two or more receive antennas (MR≥2) can be designed to either maximize spatial diversity (e.g. via the Alamouti Space-Time Block Coding technique [16]) or to perform spatial multiplexing (e.g. via [14]). However, for a symmetric antenna configuration, when MT=MR=2, it appears that spatial multiplexing gain is very difficult to exploit, whereas Space-Time Block Coding techniques combined with coherent receive processing can achieve significant improvement of the Bit Error Rate performance with low computational complexity[10]. This gain can translate into a larger range and therefore a larger throughput at a given distance. It can also be traded-off against a reduction of the transmit power, and hence an increase of battery life for portable devices, or in order to support higher order modulations to increase the bit-rate. However, with MT>2 and MR>1, transmit diversity and spatial multiplexing gains can be obtained simultaneously, but there is a fundamental trade-off in the level of transmit diversity and multiplexing gains that can be achieved in any transmission scheme. Thus, the performance of specific transmission and reception strategies for achieving both diversity and multiplexing gains is an interesting area of research which needs to be carefully considered.

A more conventional means to achieve a higher bit rate is to increase the bandwidth. The technique that consists in transmitting on several channels in parallel is called channel bundling. Several 802.11a chips currently on the market implement a proprietary channel bundling, although regulatory bodies disapprove this. The complexity of the receiver is naturally increased, but transmitting on two adjacent channels can be less complex than transmitting two separate streams on antennas with spatial multiplexing, because it is not necessary to duplicate all the RF components. However, several issues remain to be clarified, if such a scheme is considered in a future high throughput standard:

- Several implementations are possible, which result in different throughput performance. For instance, the bundling of two channels makes it possible to transmit either a single packet per transmission opportunity in half the normal duration, or two packets during the normal duration. The first scheme is less complex but exhibits a lower throughput since the weight of the overhead is increased, as explained in the previous section. The second scheme, on the contrary, would increase the PHY bit rate without decreasing the MAC efficiency.
- In the WINNER short-range scenarios (e.g. indoor and hot spots), it is likely that more than one channel will be unoccupied. However, as soon as the spectrum gets saturated, channel bundling will cause problems. Therefore, just like Dynamic Frequency Selection was adopted (at least in Europe so far) to have a fair and coherent spectrum sharing, some mechanisms can be studied to enable a fair and efficient use of channel bundling at 5 GHz.
- It should be kept in mind that channel bundling, contrary to multiple antennas, does not provide higher system capacity since it does not improve the spectral efficiency.

Turbo-codes were adopted in third generation cellular systems and allow approaching the channel capacity in additive white Gaussian noise (AWGN) channels. In multi-path environments such as those encountered in WLANs, the application of turbo-codes to an OFDM PHY typically brings a 2 to 3 dB gain [17].

Finally the use of higher order modulations has also been proposed. Introducing a 256 QAM mode would theoretically allow the peak bit rate of 802.11a to increase from 54 Mbit/s to 72 Mbit/s. But the Signal to Noise Ratio required by such a modulation is also much higher. This degradation of the SNR could be compensated by a gain brought by for instance Space-Time Block Coding. However, implementation and quantization requirements for such a modulation scheme will have an impact on the device complexity and have to be carefully assessed.

In order to evaluate the various technical solutions presented, we performed throughput versus range simulations in a typical indoor environment, with the packet size of 1500 bytes and transmit power of 200mW. We focused on the impact of space-time block coding and two-channel bundling, as well as the improvement provided by 802.11e. Moreover we take also into account the Hiperlan/2 (see section 6.2.3) system as it is well known to present a significantly better MAC effectiveness with respect to 802.11 in virtue of its centralized TDMA access scheme. The results are plotted on Figure 3-1. The throughput achievable with plain 802.11a is plotted as a reference and hardly reaches 30 Mbit/s. With the first channel bundling strategy mentioned in the previous paragraph, the reduction of the useful transmission time increases the weight of the MAC overhead, making 802.11e less efficient than HiperLAN/2. With the second and most efficient channel bundling scheme, in which two packets are transmitted in parallel, not only the PHY layer bit rate but also the MAC throughput are doubled. In both cases however, we assume that the two channels share the same Power Amplifier, therefore the maximum transmitted power per channel is half that of plain 802.11a. At distances close to the effective cell range (here around 50 meters), the signal-to-noise ratio per channel is 3 dB lower, which balances the gain of transmitting on

[10] In European project FITNESS (http://www.ist-fitness.org/) the Bit Error Rate performance of the 802.11a system was improved by up to 8.7 dB in the best case, compared to a single antenna system with the same total transmitted power.

two channels in parallel. The implementation of space-time block coding (here a 2x2 Alamouti coding) in conjunction with channel bundling drastically increases the range: at 40 meters the throughput is still higher than the maximum of 802.11a (30 Mbit/s). Finally, the highest efficiency of 802.11e and HiperLAN/2 allows reaching about 90 Mbit/s (the maximum PHY layer bit rate being 108 Mbit/s).

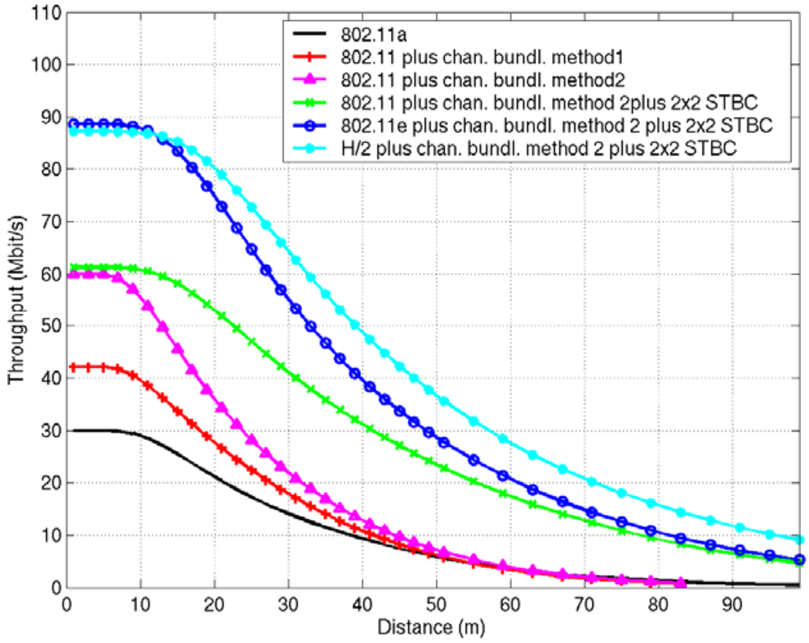


Figure 3-1: Impact of channel bundling and space-time coding on throughput vs. distance performance

Other enhancements targeted for both the physical and the MAC layers, are currently under definition within IEEE 802.11n. As far as the MAC layer is concerned, many of the initial proposals seem to move toward a refinement of the mechanisms that have already been defined for the 802.11e rather than introducing a completely innovative design. This choice is mainly leaded by the fact that one of the requirements imposed for the definition of a new standard, has been the backward compatibility with the original 802.11 MAC. Hence, propositions such as large packet aggregation and segmentation or reduction of IFS have been presented and will be subsequently assessed. On the other hand, the WINNER system has not to be compliant to such a constraint and a different approach will be investigated too.

3.4 Possible Enhancements to HiperLAN/2

Potential Enhancements of HiperLAN/2-like Wireless LAN-systems and deployments are possible through the application of advanced signal processing and the exploitation of macro- and micro-diversity much in the same way as it is already described in 3.3 for 802.11 type WLAN systems, since the proposed enhancements are only related to the PHY and therefore to a large extent protocol-independent. As witnessed in the context of 802.11 market introduction, first-step enhancement can be expected to be limited to using different modems as a basis for identical MAC protocols (e.g. 802.11b → 802.11g evolution). In the case of HiperLAN/2 an application of the same, frame-based MAC protocol on top of a 100Mhz channel is envisaged, which can be expected to provide a trunking gain in performance as compared to 5 20Mhz Channels.

The next step, which is especially important for HiperLAN/2-like systems being deployed as fixed infrastructure (as opposed to the Home Environment which in most cases features an ad hoc structure), is to exploit spatial diversity through the use of directive antennas and through intelligent deployment concepts. Simulation results (Figure 3-2) show that, depending on the propagation conditions, very tight re-use factors (of up to 1) can be achieved, e.g. in urban scenarios which feature a high degree of shadowing.

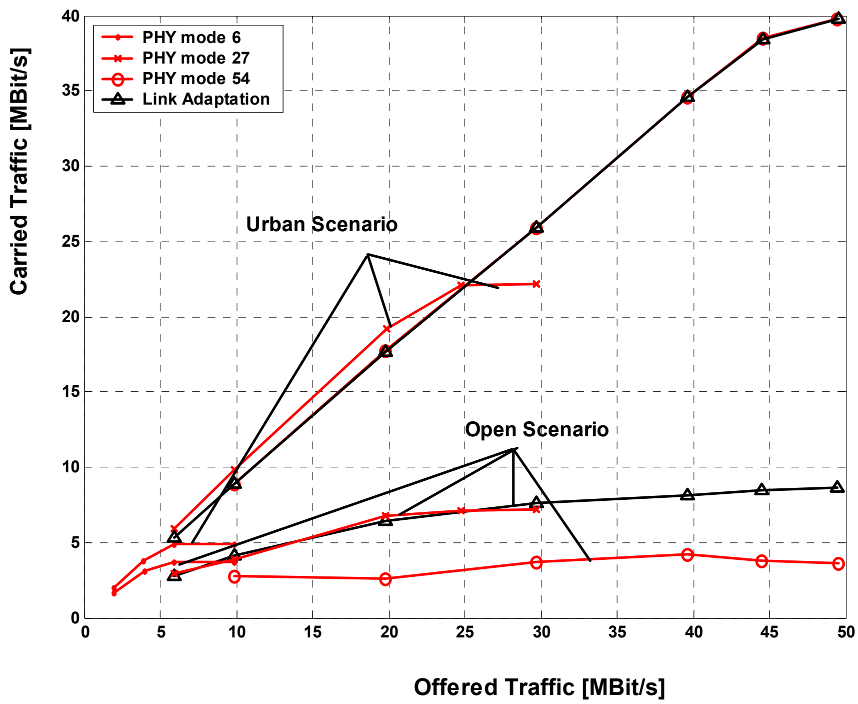


Figure 3-2: Performance difference (Aggregate carried traffic per cell vs. aggregate offered traffic for different fixed PHY-modes and Link Adaptation) between high-shadowing (urban) and open scenarios of similar geometry (Manhattan Grid Placement, 100m Block Size, with and without Buildings, frequency reuse: 1, omni antennas at AP and UT)

The next potential enhancement step, which directly arises from the exploitation of spatial diversity, is to enhance the medium access protocol in order to support said spatial diversity. A related proposal has been made by the IST-STRIKE project for HiperMAN systems [33]. The enhanced protocol supports Multiple Transmit Multiple Receive (MTMR) schemes by way of Smart Antennas. This way, a Space Division Multiple Access (SDMA) component can be introduced into the standard HiperLAN/2 frame. An exemplary MAC frame is illustrated in Figure 3-3: After a broadcast phase announcing the structure of the next MAC-Frame, which is transmitted by way of an omni directional antenna, the individual transmissions towards terminals are transmitted using the beamforming capability of the AP antenna.

Simulation results [34] have shown that for a scheme as depicted in Figure 3-3, the achievable aggregate throughput scales almost linearly with the maximum number of parallel transmissions.

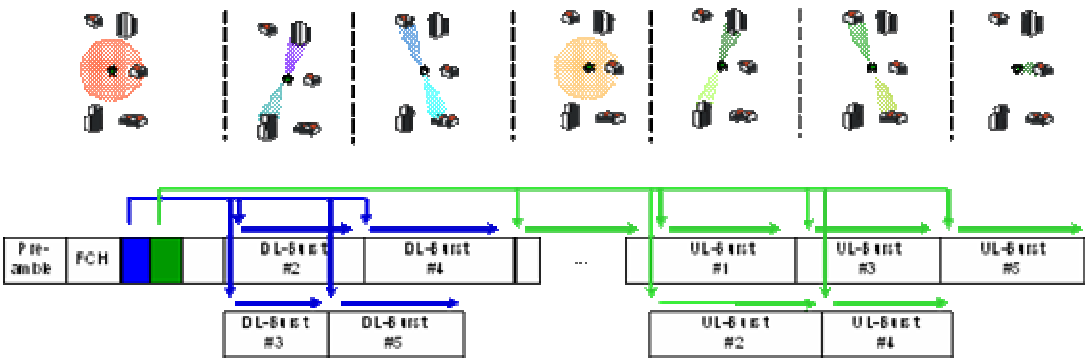


Figure 3-3: SDMA support for TDD MAC-frame-based Systems (e.g. HiperlAN/2)

3.5 Possible Enhancements EEE 802.16a

As stated above the IEEE802.16a standard foresees and allows the usage of several advanced PHY layer features either mandatory (e.g. adaptivity of burst profiles) or optional (e.g. Adaptive Antenna systems, Space Time Codes, …). The introduction of such features does not impact the system in deep but allows always the backward compatibility with the mandatory features stated in the specifications.

Between other, even if it is an optional feature, the adaptive antenna systems support is accurately specified in the standard. The usage of such feature brings the well known advantages that spans from coverage extension, interference reduction, and advantages over multipath. Furthermore, the introduction of AAS allows a considerable capacity increase through SDMA (Spatial Division Multiple Access), which reuses bandwidth by multiplexing signals based on their spatial signatures. The introduction of the SDMA approach impacts heavily on the MAC layer architecture which has to collaborate deeply with PHY layer in order to fully exploit the spatial dimension. Multiuser diversity, provided by independent fading channels across different users, can be exploited through scheduling which should be performed according to the knowledge of the channel state information (CSI) available at the scheduler Radio Resource Management (RRM) strategies should then include concepts where spectrum optimization takes into account physical-layer features. Scheduling (that can be defined as a Channel Aware Scheduling) can be conveniently combined with linear pre-coding (e.g., beamforming) in multi-antenna systems and in this context can be studied as a specific problem of power allocation or matrix modulation design.

Additional remarks about the evolution perspectives from a network viewpoint are summarized here below.

- A first decisive step of the foreseen evolution will consist of the miniaturization of the SS and its consequent integration into laptops. The scenario of WiFi and WiMax enabled portable PC's is considered particularly attractive, and it is realistic in a medium term.
  This fact, by itself, without any change in the WiMax network (i.e., since 802.16a) will allow PORTABILITY, i.e., the freedom of accessing the network from any covered place, and of working in (quasi)-static conditions (realistic from the viewpoint of data applications).

- The second major step will be obviously the evolution to 802.16e, which, from a conceptual viewpoint, will allow exchange of information between neighbours Base Station, and will define the protocols necessary to support mobility. More specifically, this will certainly allow NOMADICITY, i.e., the possibility of changing the point of attachment to the network without having to re-initiate a working session.
  The real capability of FULL MOBILITY support (really seamless handover and support to relatively high speeds) is in principle pursued by 802.16e. It should be however checked against the capabilities of the overall network (access + core) to support these functionalities. An overall network view, in the scenario of full mobility support, is still under study.

- It should be noted that the last scenario will be significant in case of support of real time services. This will require the implementation of packet-based technologies such as VoIP which is in principle possible. In this longer term scenario, WiMax may be seen as a real competitor of 3G and its evolutions. However, several critical aspects will have to be solved before achieving this step: e.g., the real functionality of VoIP in an IP-based mobility support; the kind of terminals to be adopted in this case, etc.

The evolution roughly outlined before is seen from the WiMax perspective only.

It will be obviously possible (and exactly consistent with the WINNER approach) that the 802.16 family will become one of the "contributors" of the B3G network, in a "convergent" evolutionary perspective.

4. New (multi-hop) Deployment concepts

Broadband radio interface technologies with high multiplexing bit rates are characterised by an unevenly distributed QoS characteristics depending on the distance between the terminal and the BSs and very limited range due to their high sensitivity against interference, attenuation and shadowing and the high spectrum bands in the scope of WINNER (3-5GHz)and the limited transmission power. One of the most important factors for high data rate communication is the existence of low path loss. In today's cellular systems obstacles such as buildings and hilly terrain introduce high path loss to be overcome, which affects communication robustness and throughput adversely. An improvement would be possible by using advanced antennas and setting up a high number of pico cellular BSs which will result in high deployment costs. A more cost efficient and innovative solution is to trade capacity against range where the capacity of a BS is partly utilised by a number of relay stations acting as wireless BSs in some desirable configuration to divide the path into segments, and then communicate around or over any obscuring object. Compared to pico cellular BSs the relay stations do not need a wired network connection. The lack of a direct backbone connection makes the relay stations not only cheaper, but also more flexible in positioning as they don't have to rely on a fixed network access, but only on power supply. The relays might be either fixed or mobile and serve to cover otherwise shadowed areas and to enlarge the limited coverage range of a BS. This solution appears very attractive since it is deemed unlikely that the high traffic capacity of a broadband BS will be used up by the user terminals roaming in its cell.

Further the offered capacity distribution of a single AP is highly unfair against the UT, i.e., requested capacity distribution as shown in Figure 4-1 (a). As shown in the figure, only users next to the AP can profit from the high multiplexing bit rates, but the user density increases with increasing distance due to the increasing area. Figure 4-1 (b) shows the optimal solution, which is, at least today, not feasible, but it is further shown in Figure 4-1 (c) that the employment of relay stations helps to approach a more optimal deployment of capacity.

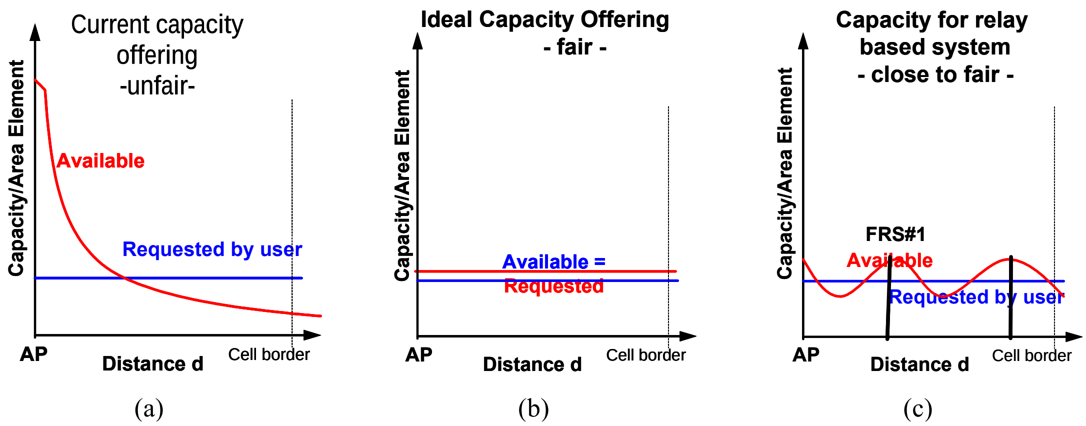


Figure 4-1: Capacity offering per area element, with and without relays

But although the fixed relay station (FRS) is cheaper to deploy due to the not needed fixed backbone network access the FRS cost some additional radio resources and also hardware cost. These two things have to be taken into account if someone wants to judge about the economical efficiency of relay based deployment. Therefore a methodology called Weighted Spectral Efficiency is introduced in the following section (Sec. 4.1) that allows a fair measure for deployment cost without the final knowledge about the exact hardware cost. Following the introduction to Weighted Spectral Efficiency new multi hop deployment concepts will be described in partly evaluated. In Section 4.2 concepts based on homogeneous relays will be shown. After that the approach of heterogeneous relays will be discussed in Section 4.3 and finally some "other" concepts will be introduced in Section 4.4 where next to the access network also some parts of the transport network will be taken for heterogeneous multi-hop link.

4.1 Weighted Spectral Efficiency for a fair comparison of Deployment concepts

In the scope of spectrum estimation methodologies [1][39], a measure usually taken for the comparison of radio access technologies is the spectral efficiency. However, the authors feel that a comparison of the sheer technical capabilities of a system is lacking fairness, because different technical approaches are also associated with different system deployment costs. This concept proposes a "Deployment Cost Factor", which is a measure that takes into account normalised deployment costs of a certain radio access technology and the coverage area that can be achieved by its infrastructure elements. In conjunction with the area spectral efficiency, this measure leads to a weighted spectral efficiency that allows a fair comparison between radio access technologies.

As an example to describe our methodology, we calculate the "Deployment Cost Factor" for a wireless broadband system based on a conventional single-hop air interface and another wireless broadband system based on Fixed Relay Stations.

4.1.1 Comparison Methodology

This section gives an introduction to the notations used for the methodology. Table 4-1 lists the symbols used in the methodology. Their use is outlined in the subsections of this section.

Table 4-1: Definition of used symbols

Symbol	Value	Unit
η	Area Spectral Efficiency	bps / (Hz m^2)
Acov	Coverage area of a cell	m^2
A	Normalised coverage area of a cell	
R	Ratio of coverage areas of cell types to be compared	
CAPHW	Cost of AP Hardware	currency (e.g. EUR)
CCONN	Cost of AP Backbone Conn.	currency (e.g. EUR)
CGAIN	Cost of Gain Ant.+Hardware	currency (e.g. EUR)
C	normalised cost measure	
ζ	Deployment Cost Factor	
E	Weighted Spectral Efficiency	bps / (Hz m^2) or Bps / (Hz EUR)
lb, ls	Size of buildings and width of streets	m

4.1.1.1 Normalised Infrastructure Costs for one cell

Since the aim of this methodology is to regard the costs of different deployment concepts, the first step of the methodology is to make these costs comparable. We suggest to obtain the normalised costs *c* by normalising the deployment costs with respect to the costs of a common infrastructure element. In our example, we use the costs of the Access Point (AP) hardware C_{APHW} for this.

4.1.1.1.1 Single Hop

We assume that deploying a conventional (single-hop) system, the costs for one cell mainly comprise the costs for the Access Point hardware C_{APHW} and the costs for the backbone connection C_{CONN}. The normalized cell costs c_{SH} then become:

$$c_{SH} = \frac{C_{APHW} + C_{CONN}}{C_{APHW}} = 1 + \frac{C_{CONN}}{C_{APHW}} =: \frac{C_{Cell,SH}}{C_{APHW}}$$

In Figure 4-14, we show 3 variants of possible deployment concepts (taken from [1]), the scenario geometry can be obtained from Figure 4-13. The 3 variants result in different cell sizes as shown in the following.

Cell area A_{cov} for case:

$$A_{\mathrm{cov},a} = 4(l_b \cdot l_s) + 2l_s^2 = 25800\ m^2$$

$$\overline{A}_{\mathrm{cov},b} = (4l_b + 4l_s)l_s + \frac{3}{2}l_s^2 = 26250\ m^2 \text{ (average cell area)}$$

$$A_{\mathrm{cov},c} = 4\left(l_b l_s + \frac{l_s^2}{2} + l_b l_s \frac{l_s^2}{4} \right) + l_s^2 = 51600\ m^2$$

Note that in case b) we obtain 2 types of cells with different coverage areas, therefore we use the average cell area $\overline{A}_{\mathrm{cov},b}$.

4.1.1.1.2 Multi Hop

In the case of a wireless broadband system based on fixed relays, we consider as a basic element a cell consisting of one Access Points and 4 Fixed relay stations that are equipped with directional antennas as proposed in [40] -[43].

In that case, the costs for one cell would comprise:

- 5 times the AP hardware, because we consider the hardware costs for Relay stations and APs to be comparable
- 4 times the additional hardware costs C_{GAIN} to equip the Relay Stations with gain antennas
- Costs for 1 backbone connection.

Again, we normalise these costs with respect to the AP hardware, as suggested above and obtain the Relative cell costs:

$$c_{MH} = \frac{5C_{APHW} + C_{CONN} + 4C_{GAIN}}{C_{APHW}} = 5 + \frac{C_{CONN}}{C_{APHW}} + 4\frac{C_{GAIN}}{C_{APHW}} =: \frac{C_{Cell,MH}}{C_{APHW}}$$

As can be calculated from Figure 4-16, the covered cell area in this case results to:

$$A_{\mathrm{cov},2hop} = 18(l_b \cdot l_s) + 9 l_s^2 = 116100\ m^2$$

4.1.1.1.3 Normalised coverage area a

The cell sizes of the systems to be compared have to be normalised to an (arbitrary) cell size A_0 (e.g. the smallest among all systems to compare). The result shall be the normalised coverage area a :

$$a = \frac{A_{\mathrm{cov}}}{A_0}$$

4.1.1.2 Deployment Cost Factor ζ

The combination of the normalised infrastructure costs and the normalised coverage area of a cell leads to the Deployment Cost Factor ζ proposed. It allows to answer the following question:

"How big is the service area that can be covered with a fixed investment?"

either in a qualitative way (shown below, left) or alternatively, if we do not normalise the costs and coverage areas, in a quantitative way (show below, right)

$$\zeta = \frac{a}{c},$$ alternatively: $$\zeta = \frac{A_{\mathrm{cov}}}{C_{Cell}}\ [\mathrm{m^2 / EUR}]$$

4.1.1.3 Weighted Spectral Efficiency Ε

The Deployment Cost Factor facilitates to weigh the area spectral efficiency of different proposed system concepts and take their deployment costs into account by comparing the Weighted Spectral Efficiency instead of the Area Spectral Efficiency. Consequently, the Weighted Spectral Efficiency provides a means for fair comparison of entirely different network deployment concepts.

Following the proposal from 4.1.1.2, the Weighted Spectral Efficiency can be calculated either by using the normalised Deployment Cost Factor (shown below, left) to directly compare different system concepts or use the absolute values (shown below, right). The latter option results in a figure that directly relates the achievable system capacity to the infrastructure investments that have to be made.

$$\mathrm{E} = \eta \cdot \zeta\ [\mathrm{Bps / (Hz\ m^2)}],$$ alternatively: $$\mathrm{E} = \eta \cdot \zeta\ [\mathrm{Bps / (Hz\ EUR)}]$$

4.1.2 Exemplary use of weighted spectral efficiency method for single-hop multi-hop Comparisons

In the following, we will give an example how the methodology can be used. As indicated in the last section, we compare a conventional single-hop deployment based on a ETSI BRAN HiperLAN2 system with a multi-hop deployment concept based on Fixed Relay Stations (FRS). For the example, we compare deployment variant (c) from Figure 4-14 with multi-hop variant (a) from Figure 4-17. Table 4-3 shows simulation results for the area spectral efficiency of the different concepts.

The basic building block of the two-hop deployment concept, which consists of an AP and 4 FRSs is shown in Figure 4-13, while Figure 4-15 shows the cellular deployment of these building blocks for various cluster sizes.

The actual process of comparing 2 system approaches will be presented in this section.

When comparing the Weighted Spectral Efficiencies of e.g. Single and Multi-Hop Systems as described, we want to investigate if the following is true:

$$\frac{\mathrm{E}_{MH}}{\mathrm{E}_{SH}} = \frac{\eta_{MH} \cdot \zeta_{MH}}{\eta_{SH} \cdot \zeta_{SH}} > 1 \quad \Leftrightarrow \quad \frac{A_{\mathrm{cov},MH} \cdot A_0 \cdot c_{SH}}{A_0 \cdot A_{\mathrm{cov},SH} \cdot c_{MH}} =: r\frac{c_{SH}}{c_{MH}} = \frac{\zeta_{MH}}{\zeta_{SH}} > \frac{\eta_{SH}}{\eta_{MH}}$$

We use this notation because the direct comparison of the weighted spectral efficiencies is difficult to visualize. Here, we are comparing the Deployment cost factors (which can vary depending on the Connection and the Gain Antenna Costs) with a fixed relation of the area spectral efficiencies, which we can obtain from the rightmost column of Table 4-3.

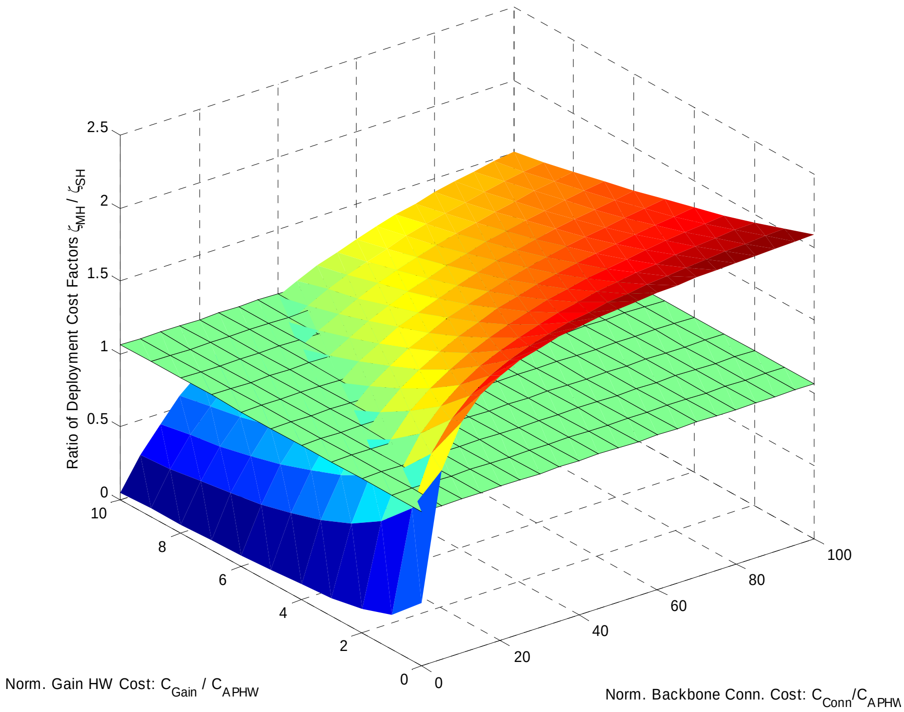


Figure 4-2: Comparing the Cost efficiency with the spectral efficiency

4.1.2.1 Example

To provide an example, we compare single-hop (index SH) deployment (c) versus multi-hop (index MH) variant (a) (clustersize N=2). The plot in Figure 4-2 shows as the horizontal plane the ratio of spectral efficiencies and the coloured grid is the ratio of the deployment cost factors as a function of the normalised backbone connection costs (x-axis) and the normalised costs of the gain antenna equipment (y-axis).

Below are the parameters used to generate this plot:

Ratio of spectral efficiencies:

$$\frac{\eta_{SH}}{\eta_{MH}} = \frac{2{,}37}{2{,}31} = 1{,}026$$

Ratio of Cell Areas:

$$r = \frac{A_{\mathrm{cov},MH}}{A_{\mathrm{cov},SH}} = \frac{116}{53{,}4} = 2{,}172$$

The way to read this plot is the following: Where the coloured grid lies above the “reference plane”, the Multi-Hop system has advantages over the single-hop system. On the x-axis, we see that this is the case for rising backbone connection costs (or falling AP hardware costs). Looking at the y-axis we see that rising costs for gain antenna equipment reduce the advantage of the multi-hop system.

4.1.3 Conclusions on Weighted Spectral Efficiency

With Weighted Spectral Efficiency a methodology was presented, which allows for fair comparison of the spectral efficiency of candidate radio network deployment concepts, which takes into account the deployment costs resulting from a certain deployment concept. A so-called “Deployment Cost Factor” is proposed to achieve this goal. An exemplary application of the methodology proves that it can be used for the comparison of entirely different radio network deployment concepts, namely a conventional single-hop wireless broadband system and a new deployment concept for future wireless broadband systems based on Fixed Relay Stations.

4.2 Homogeneous Relays

4.2.1 Introduction

This section is devoted to a detailed discussion of such multi-hop and relaying systems in the context of homogeneous relays and is organized as follows. Paragraph 4.2.2 outlines basic relaying concepts for initial multi-hop investigation. In paragraph 4.2.3 a description of relay channel capacity is presented. Paragraph 4.2.4 is devoted to the new first muti-hop concepts based on enhancing existing concepts, while paragraph 4.2.5 addresses models and algorithms for routing/forwarding in multi-hop wireless networks contexts. Finally, cooperative relaying concepts are described (4.2.6).

4.2.2 Basic Concept

4.2.2.1 Reference Radio Interfaces for initial multi-hop investigations

It has been assumed by WP3 that enhancements to existing Radio Interfaces will be studied as first step of WP3 work. To limit the number of candidate systems a list of Radio Interfaces that will be used for the basis of WP3.2 investigations is given in Table 4-2. In order to concentrate the activities in T3.2 Scenario 2 and Scenario 3 have been selected as the initial focus.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
HSDPA		**X**	**X**	**X**	
IEEE 802.11x	**X**	**X**			
IEEE 802.16x		**X**	**X**		**X**
HiperLAN/2	X	X			

Table 4-2: Radio Interfaces for initial multi-hop investigations

4.2.2.2 Reference multi-hop topologies

4.2.2.2.1 Business scenarios

Although the future evolution of mobile communication is hard to predict it will clearly depend significantly on external influences, e.g. local regulation or general availability of fixed network (NW) access opportunities. Two different paradigms of handling spectrum may be observed in the current wireless communication landscape and these are licensed and license exempt spectrum. In licensed frequency bands (most often used in traditional operator driven mobile communication networks) the resource (i.e. spectrum) is exclusively owned and controlled by one operator and unauthorized use is prohibited by law. The resource is leased to the operator for money or for certain conditions (e.g. to build a NW which reaches a certain percentage of the population). As the use of the resource will generally not be totally free of cost a natural requirement is to use it efficiently. That means in other words that the deployed system will likely use a spectral efficient design. Another observation for this kind of deployment is the general use of network engineering i.e. the planning of the network elements because it enables greater spectral efficiencies and ensures that coverage is available in the desired area.

Contrary to this, for license exempt frequency bands (as the name suggests) no restrictions on who is allowed to access the band exist. That means on the other hand that the availability of the resource cannot be guaranteed and strong interference may exist and hence prohibit transmission. There is still the trend of efficient transmission observed based on the motivation of the individual user to obtain the maximum link capacity, however costs restrictions are more relevant in this case. It should be noted that the overall system capacity is in general not in the mind of the customer.

Besides the two scenarios a third possibility exists in which the resource is exclusively assigned to a number of operators, e.g. divided in the time domain. No deployment does currently exist for this scenario.

When it comes to the trends for a next generation system, using multiple (radio) hops (henceforth referred to as a multi-hop network) can be seen as on option to handle the small cell sizes that are likely to appear when moving to the envisaged data rates. In general, multi-hop systems may comprise two different type of elements, fixed and mobile relays. If the two different approaches, i.e. the one based on fixed and the other based on mobile relays, are mapped to the two network deployments mentioned above it is likely that in a scenario where the whole resource is exclusively assigned to one or more operators a fixed multi-hop topology will be employed. This is mainly because the operator(s) has(have) the general ability to engineer and design the network according to the current demands and can control and guarantee the grade-of-service (GoS) required.

When considering the hot-spot/hot area deployment, the actual multi-hop deployment is not very easy to foresee. In an operator oriented scenario where multiple operators share the medium it is still likely that fixed multi-hop infrastructure elements will be used because a reliable coverage can be provided in this case. Mobile relays could be used in this scenario for coverage extension if an incentive is provided to the users that relay data. This assumption can additionally be justified by looking at the average coverage sizes of all operators. As no restrictions apply (like they do today) to enter the market more operators will try to serve customers. It is very likely that most of them will not provide a complete coverage due to cost reasons (especially the smaller they are) but rather cover a distinct area. Mobile relaying can be seen as one possibility to increase these coverage areas without generating additional costs for the operators.

In a Wireless LAN oriented deployment where most of the Access Points are provided by individuals owning a flat-rate high speed internet connection, and like to share it with everybody else, mobile relaying will become the more important deployment case. The whole mobile relay scenario could be based on sharing the transmission capacities with everybody else because the bandwidth is available anyway. Users might be concerned about the power needed to relay data. However, users that relay data might come in a situation where they might benefit from relaying capabilities of others stations and, hence, participate in a cooperative network that is based on a win-win situation. Therefore, no additional payments between the users would be required. If transmission bandwidth is still valuable operators may act as broker to enable payment between a multitude of partners in this scenario. Other operators may exist providing a general relay network where the APs are provided by users. A combination with the scenarios mentioned above (e.g. in airports, where no individuals live) is likely.

4.2.2.2.2 Fixed Relays Networks (FRNs) as wireless mesh networks

Introduction

After having deeply investigated the importance and benefits that will derive from the introduction of multi-hop concepts and relay nodes in the new system architecture, an analysis about how to set relays from a topological point of view is required. The simplest way to set relays is to have just one relay for

BTS or AP, or a cascade (line) of relays. Actually, it is likely that relays will have a more complex role and it is almost natural to presume the future existence of relay networks.

More specifically, the concept of "relay nodes" can be associated to the concept of "mesh network" that is emerging in several contexts, related to wireless communications.

Hence this section propose a reference application scenario called “fixed relay network”, as wireless mesh network

Fixed Relay Networks as wireless mesh networks

A wide variety of research and development efforts are geared toward understanding and addressing the technical challenges of wireless multi-hop mesh networks (or briefly wireless meshes) that are far from typical actual point to point or point to multi-point radio systems. Such kinds of network are addressed to be one of the most important wireless technologies for the next decades; wireless meshes are beginning to have an impact on realistic evolutionary scenarios and their importance will probably increase over time. Wireless mesh networks could be defined as a multi-hop system where links are established via radio transmissions, in which devices assist each other in transmitting packets through the network. Wireless meshes may be classified through the following conceptual taxonomy:

- ***Self-organizing or static:*** in self-organizing manner, wireless mesh may adapt its connections according to channel conditions or devices failures by simply modifying the transmission range (or even through mobility), by performing a sort of "topology management" without any external intervention; whereas in static manner connections are preset and a pre-configured mesh topology is deployed. In both cases there is the possibility to add new nodes on the fly.

- ***Infrastructure, client or hybrid:*** a wireless mesh network may play different roles; it can act as an infrastructure collecting and moving data or voice from and to personal/mobile terminals that are the clients (it is likely that wireless meshes working in this way would be static). Anyway it is also possible to have self-organizing mesh networks formed directly by clients on the fly as envisioned by MANETs; or alternatively it is even possible to construct hybrid cellular/mesh networks where a wireless mesh may be connected to one or more BS improving capacity.

- ***Fixed or mobile:*** since no wired connections are required, mobility can be in principle ensured in such networks, if a MANET-like concept is assumed. Obviously mobility brings a series of new problem to face. Alternatively, a wireless mesh may be constituted by nodes in fixed position, (e.g. if the mesh acts as an infrastructure) keeping anyway also in this case a high degree of flexibility as several different paths may be found to connect to given nodes.

- ***Temporary or permanent:*** one of the main advantages that may derive from wireless meshes is the possibility to deploy a network quickly and at low cost where it is needed, so that a wireless mesh may be deployed temporarily to answer particular exigencies or permanently to serve a given area.

Relays, as envisioned in project WINNER, seem to be particularly suitable to create such wireless multi-hop mesh networks especially for meshes acting as fixed infrastructure (both permanent and temporary). In fact relay stations are potentially low cost, easy to deploy and to move, easy to configure and sufficiently versatile to be inserted both in stand-alone wireless meshes and in hybrid wireless meshes. Thus, it is introduced here the concept of a wireless multi-hop mesh network formed by relays that act as a permanent or temporary infrastructure, that are connected, at least in WINNER proposal, to a fixed network access (e.g., AP or BS) and meant to serve client terminals. Relays are able to manage both traffic that is internal to the mesh island (i.e., it is generated and terminated in the mesh island without the need of AP or BS intervention), and traffic directed outside the mesh island. We will refer to such wireless multi-hop mesh network formed by relays as **Fixed Relay Networks (FRNs).**

Wireless mesh networks and hence FRNs as we have defined them, may be used in several scenarios ranging from wireless Metropolitan Area Network to hot-spot wireless access, even to home private wireless Local Area Network. In a further scenario, a wireless multi-hop mesh network may be created among IEEE 802.11 hot spots, to optimize the connections towards one fixed AP (see e.g., Figure 4-4).

It should be noted that the concept of wireless mesh networking is not only matter for long-term research. For example, it is explicitly considered (and studied, although preliminarily) in the context of IEEE 802.16.

In the IEEE 802.16a mesh mode, data transmission can directly occur among Subscriber Stations (SSs), since the protocol allows to set up data connections among neighbours with multi-hop communication.

In this scenario, SS themselves play the role of "relays", allowing an optimization, or an easier extension, of the last mile access realized through the 802.16 technology. If the "mesh" functionality of a SS is seen as a possible upgrade of the basic SS, a gradual evolutionary path from the first deployment of a wireless access network (without mesh) to an enhanced wireless access network (meshed) can be envisaged.

The Figure below shows a possible scenario for wireless meshes and FRNs.

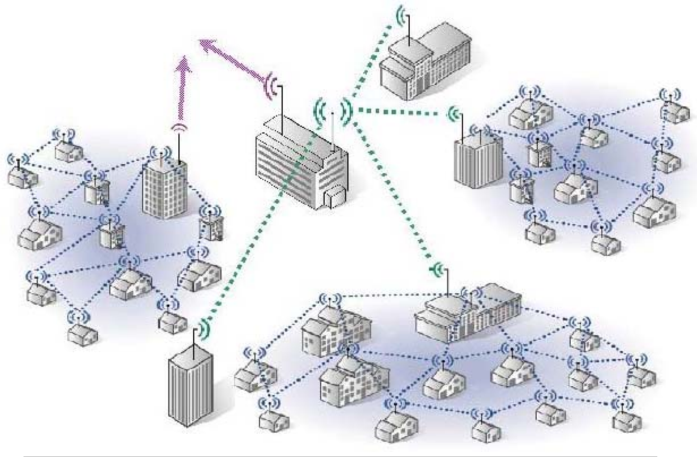


Figure 4-3: Wireless multi-hop mesh networks (Relay as user node)

The figure above shows a wireless multi-hop mesh networks (blue links) connected to a wireless backbone (green links). Some rather realistic scenarios (from a technical viewpoint) can be imagined. For example, in the above figure each mesh element (relay) might be an end user equipment that people can buy and install (e.g. on the rooftop of their own house, similarly to the installation of antenna TV, parabola). It could be forseeen that power supply is available even if the owner is not using the devices, so it is possible that individual user equipment may be used for relaying external traffic. In such case, a bonus might be envisioned for the owner, e.g. free traffic. It is a sort of self organizing "user oriented" mesh network without any mobile operator deployment before. It should be however noted that the related "business scenario" may be critical and its identification not yet defined.

Another possible scenario for mesh networks development is shown in the following figure, where each mesh network element (relay) is a Point-to-MultiPoint (PMP) node. Relays are added so they can collect the traffic originated by terminals and transmit to the access point or to another relay. Hence two different types of data traffic flows could be foreseen: mesh traffic (red links) managed by the access point and PMP traffic (blue links) managed by each relative PMP node. This is a sort of "mobile operator oriented" or "distributed feeder" mesh network, because the mobile operators can set up the FRNs in the needed positions.

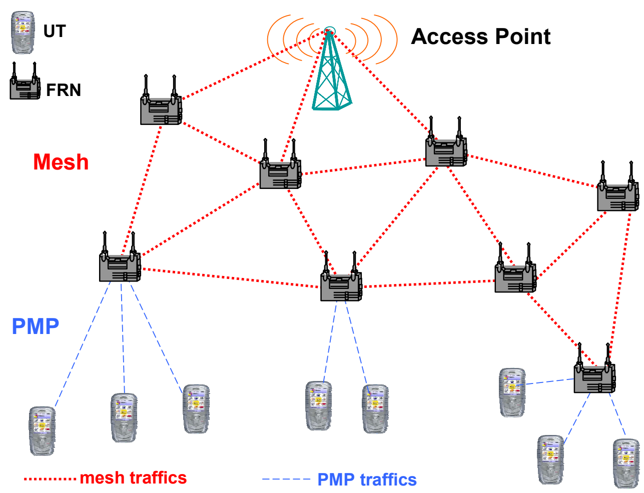


Figure 4-4: Wireless multi-hop mesh networks (Relay as PMP node)

In both the scenarios outlines above, a first "layer" of relays surrounds the access point and can directly communicate with it. A second and further layer of relays can then be added to connect the terminals not already connected to the first layer, forming a tree-shaped topology. Relay stations are supposed to be fixed or movable, where movable means that they can be easily moved from one site to another (differently, e.g., from actual cellular BTS that once deployed is clearly very hard to move), preserving even the possibility of keeping the relays active during their re-positioning. In general, FRNs have a defined topology; topology changes are possible (e.g., due to node failures), but probably not frequent. Therefore, the distribution of network state information is not much more complex than in wired networks, and even a centralized control of route selection can be adopted. In fact it is assumed that the AP knows the topology and global network state. Having fixed position, relay stations are expected not to suffer from power supply problems since they can be easily fed through electric distribution network, or through very long life rechargeable battery, or even through solar panels.

The interest in wireless multi-hop mesh network and in FRN may be justified by several potential advantages achievable thanks to this new architecture concepts. Mesh topology allows to have several possible paths between a given pair of nodes and thus provide a better redundancy with respect to usable routes. Redundancy implies several benefits: firstly, traffic balance may be performed when there is a localized high interference: one or more connections may be routed on alternate path; or, if one connection requires a large amount of bandwidth, it is possible to dynamically reroute traffic along new routes so to avoid congestion. In current single-hop network it is not possible to achieve the dynamical adaptation to interferences or overburdened nodes. Route redundancy also provides a high level of fault tolerance; in fact if a node of the mesh goes down or a wireless link becomes too noisy, traffic is simply rerouted dynamically. Therefore, the loss of one or more nodes doesn't necessarily affect the network's operation. A great advantage of mesh topology is that redundancy and hence reliability may be achieved simply over-designing the network adding some extra nodes. It is also possible to imagine a situation where no human intervention is necessary, but everything may be done through a proper routing algorithm, so that FRN may be considered as auto-balancing and self-healing.

Further, a Fixed Relay Network can perform such functions as load balancing between BTS/AP to which it is connected, in order to avoid overload or congestion or bottleneck phenomenon; the deployment of CAC (Call Admission Control) functionalities can be envisioned since users access to a core network through relays network.

An advantage of fixed relays, with respect to mobile ones, is the possibility of adopting simpler centralized routing algorithms instead of complex completely distributed routing algorithms

Finally, wireless mesh multi-hop network and FRN clearly offer a much higher degree of scalability than the actual point to point or point to multi-point systems, being able to handle hundreds of nodes that simply can be added to the mesh; the addition of new nodes to the mesh implies also bandwidth scaling since overall network capacity is increased as well as total available bandwidth.

In conclusion, wireless meshes potentially represent a very flexible, powerful and adaptable solution for deploying networks across a wide typology of geography and application scenarios.

4.2.3 Relay Channel Capacity

4.2.3.1 Macrodiversity versus microdiversity system capacity when considering the receiver RFFE model

4.2.3.1.1 Introduction: ad hoc networks

Wireless networking constitutes an important component of future information technology applications. Recently, the use of multiple antennas at wireless transmitters and receivers has been identified as an enabling technique for high-rate multi-media transmissions over wireless channels. Although the point-to-point Multiple Transmitter Multiple Receiver (MTMR) system channel is relatively well-understood, the general area of multi-user MTMR system communications, namely, *ad hoc* networks or multi-hop wireless networks, is still at an infant stage and poses a rich set of challenges to the research community.

These networks consist of a group of nodes that communicate with each other over a wireless channel without any centralized control. Examples of such networks are in coordinating an emergency rescue operation, networking mobile users of portable-yet-powerful computing devices (laptops, PDAs, smart phones) on a campus using, for instance, IEEE 802.11 wireless local-area network (LAN) technology, sensor networks, automated transportation systems, Bluetooth, and home RF. Ad hoc wireless networks differ from conventional cellular networks in that all links are wireless and there is no centralized control. As every node may not be in direct communication range of every other node, nodes in ad hoc networks cooperate in routing each other's data packets. Thus, lack of centralized control and possible node mobility give rise to a number of challenging design and performance evaluation issues. An important performance analysis issue is to determine what the traffic-carrying capacity of such multi-hop wireless networks is.

4.2.3.1.2 Capacity of channel networks

Depending of the nature of the network and the assumptions made, channel network capacity is known only in special cases. For example, the theory of flow in network has satisfying answers in domains like circuit theory and the flow of water in pipes. However, the theory of information flow in networks does not have the same simple answers as the theory of water flow in pipes. Although an upper bound on the rate of information flow across any cut-set was computed by Cover and Thomas in [44], these bounds are not achievable in general. However, it is gratifying that some problems like the relay channel admit a simple interpretation of the capacity upper bound.

The following subsections develop the concept of water networks and of upper-bound capacity. The cut-set upper bound is also recalled and a weaken bound is also given. Then, a brief summary of the main results found for the ad hoc network, under simplifying assumptions, is provided.

Water networks

Consider a network of water pipes. It could be seen as a set of nodes or a set of source/destination pairs, with a main node: the source of water, and the final destination: the sink, where all the water has to go. Given the structure and the parameters of the network, what is the maximum water flow that can be accommodated by the network?

This question was studied by Ford and Fulkerson [45]. Their basic example is illustrated in Figure 4-5. The nodes labeled s and d are the source and the sink, respectively, and the nodes labeled x and y are intermediate nodes. The labels on the edges are the corresponding flow capacities. A *cut* separates the network into two parts, one containing the source and the other containing the sink. The *flow* across a cut is just the *sum* of the capacities of the links cut by the cut. One such cut is the dashed line in Figure 4-5. Since the sum of the capacities of the edges crossing the cut from the source to the sink is 3, the cut flow is 3. It has been shown in [45] that the maximum flow from the source to the sink (the "max-flow") is equal to the minimum flow, minimized over all the cuts (the "min-cut"):

max-flow = min-cut

often referred to as the *max-flow min-cut* theorem.

It is the note that the dashed line in Figure 4-5 is a min-cut, this is could be verified quickly. Thus the capacity in the example is 3.

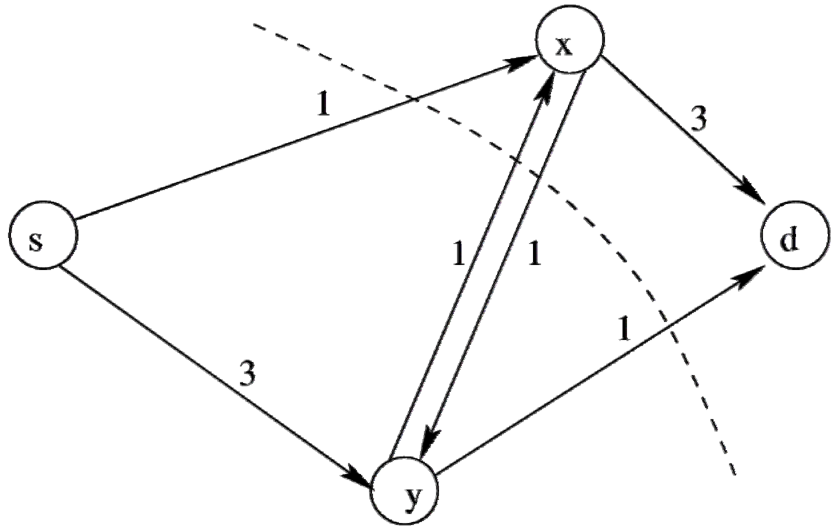


Figure 4-5 - An example of the max-flow min-cut theorem

Let consider the case where multiple sources have information to send to different destinations, which is the case of postal delivery networks. The problem now is no more that of the water network, and the solution presented above can not be applied to the latter network. Recently, an algebraic solution has been proposed by Koetter and Medard [46]. "Algebraic" here means that the authors define a number of multi-variable polynomials whose coefficients "inherit" from the network topology, and they show that the zeros of those polynomials constitute the answer to certain capacity questions.

Differing from the water network, the wireless network has the particularity that every node receives the signal from every other node. In the following, the relevant prior art on the investigation of ad hoc network capacity is quoted. In each study, simplifying assumptions are made.

Global ad hoc capacity under simplifying assumptions

The title of the subsection could be reformulated as a question: "how much information can be carried over a wireless network?"

The first attempt to address this issue was made in [47], where the authors introduced the *transport capacity* of a network, which is the sum of products of bits and the distances over which they are carried, expressed in *bit-meter*, i.e. the network transports one bit-meter when one bit has been transported a distance of one meter towards its destination. Their main results found under a certain model of communication, are that the transport capacity of a network of n nodes located in a region of unit area is $O(\sqrt{n})$ bit-meters/s. As a consequence, this implies that the per-user capacity scales as $O(\sqrt{n}/n)= O(1/\sqrt{n})$, *i.e., the achievable throughput per user in this point-to-point coding model tends to zero as user density increases.*

Then, in [48], Xie and Kumar have computed sharp information scaling laws under some conditions. In addition, they have established the optimality of multi-hop operation in some situations, and a strategy of multi-stage relaying with interference in some others. Then in [49], Gupta and Kumar have proposed an information-theoretic construction scheme for obtaining an achievable rate region in communication networks, which are of arbitrary size and topology and communication over a general discrete memory less vector channel. Moreover, using the proposed scheme, inner bounds for the multi-cast and all-cast capacities have been derived. They have also shown that the proposed scheme achieve the transport capacity of $\Theta(n)$ bit-meters/s in a specific wireless network of n nodes located in a unit area region.

For general wireless ad hoc networks, without making simplifying assumptions, an upper-bound has been computed by Cover and Thomas in [44], which will be quoted in the following paragraph.

For general networks and arbitrary coding, capacity is not known. A general upper bound on the capacity was given in theorem 14.10.1 in [44]. Let recall it. Consider a network of n nodes, the k^{th} node receives the signal Y_k and sends the signal X_k. The channel is represented by the channel transition function $p(y^{(1)},\ldots,y^{(n)}|x^{(1)},\ldots,x^{(n)})$, which is the conditional probability mass function of the outputs given the inputs. This probability function captures the effects of the noise and the interference in the network.

The channel is assumed memory less, i.e. the outputs at any given time instant depend only on the current inputs and are independent of the past inputs. The rate sent by the i^{th} node to the k^{th} one is denoted by $R^{(ij)}$. A cut will divide the nodes into two subsets denoted by S and S^c. Using the shorthand $X^{(S)} = \{X_k\}_{\{k \text{ in } S\}}$, the theorem 14.14.1 can be stated as follows:

Theorem 1:

If the information rates $R^{(ij)}$ are achievable, then there exists some joint probability distribution $p(y^{(1)},\ldots,y^{(n)}|x^{(1)},\ldots,x^{(n)})$, such that

$$\sum_{i\in S,\, j\in S^C} R^{(ij)} \le I\left(X^{(S)};Y^{(S)}\middle|X^{(S^C)}\right)$$

for all subset S in {1,2,… ,n}. Thus the total rate of flow of information across cut-sets is bounded by the mutual information.

It is to note that this upper bound can be interpreted as the min-cut max-flow theorem, where the set S characterizes the cut.

As Gastpar has shown in [50], it is possible to weaken the cut-set bound, which leads to a bound that is easier to compute.

***Corollary 1 (weak cut-set bound)*:**

For any subset S from the network,

$$\sum_{i\in S, j\in S^C} R^{(ij)} \leq \max_{p_{X^{(S)}}} I\left(X^{(S)};Y^{(S)}\middle|X^{(S^C)}\right)$$

Using this formulation, an upper-bound on the communication network could be easily computed: choosing a subset of the network, the mutual information across the cut could be maximized, and the sum of rates across the cut should be smaller.

The cut-set bound could be applied to the investigation of the broadcast channels, the multiple-access channels and the relay channels. In the following we will state the main results for the latter channel, specifying the simplified conditions made in each case of study.

4.2.3.1.3 Relay channel capacity

The Gaussian relay capacity upper-bound

The general Gaussian relay channel is composed of three nodes: the source node, the destination node and the relay node, which is assisting the data transmission from the source to the destination. Using the cut-set bound presented above, the general Gaussian relay channel is upper bounded by [51]:

$$C \leq \max_{p(x_1,x_2)} \min\left\{I\left(X_1;Y,Y_1\middle|X_2\right), I\left(X_1,X_2;Y\right)\right\}$$

where the first term in the min(.,.) can be treated as the sum rate from the source to the relay and the destination, corresponding to a Broadcast Channel (BC) part, i.e. the 1x2 macro-diverse SIMO system; and the second term can be viewed as the sum rate from the source and the relay to the destination, corresponding to a Multiple-Access Channel (MAC): the 2x1 macro-diverse MISO system.

Indeed, this equation has an interesting max-flow min-cut interpretation, as illustrated in Figure 4-6. Roughly speaking, the rate of the information flow transmitted on the relay channel is constricted by the bottle-neck corresponding to either the first cut (BC) or the second one (MAC).

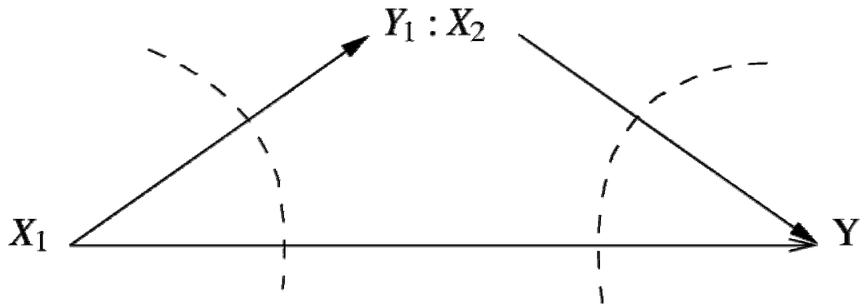


Figure 4-6 - Relay channel: max-flow min-cut

In the following, we give a summary of the main results established for the relay channel capacity.

The relay channel under simplifying assumptions

In the context of network information theory, the relay channel was first introduced by Van der Meulen [52]. Meulen gave capacity inner bounds in a time-sharing approach. However the capacity was established only for relatively degenerated channels. Then Cover and El Gamal [51] have found the capacity for the degraded relay channel. Then, in [53], Shein and Gallager have computed upper and lower bounds to capacity for the physically degraded Gaussian parallel relay channel. The parallel relay channel consists of a source node communicating with a destination and two relays assisting the data transmission. They have shown that the upper and the lower bounds coincide, resulting in reliably achievable rates, when considering one of the following scenarios. The first scenario is when a natural staggered block coding scheme is used, i.e. the source transmits in the first half symbol period, and in the second half symbol period, the relays decode the received observations and then transmits identical corresponding code words (with high probability). The second scenario is when the relays view independent observations of the channel input. Then each relay acts as a simple transponder, amplifying both signal and noise. Recently, Gastpar [54] has investigated the *relay traffic pattern* composed of one source/destination pair, and a certain number of nodes assist this transmission.

In a first step the source transmits the data information to the relays, and the destination, and then in the second step the relays amplify and forward the received signal to the destination node. He has shown code construction leading to achievable rates and derived upper and lower bounds from the max-flow min-cut theorem. He has also shown that the lower and the upper bounds meet asymptotically as the number of nodes in the network goes to infinity. He proved that the capacity of the wireless network with *n* nodes under the relay traffic pattern behaves like *log(n)* bits/second.

In [55], Xie and Kumar establish an achievable rate formula for the multiple-level relay channel, which is composed of *n* nodes: a source node, the destination node and the other nodes assist the transmission, as in the relay traffic pattern with the different way of processing information through the relays. Assume each node *i* (in *{1, 2,…, n-1})* sends $x_i(t)$ at time *t*, and each node *k* (in *{2,3,…,n}*) receives $y_k(t)$ at time *t*, then one-step time delay is assumed at every relay node to account for the signal processing time, so that for all *i* (in *{1, 2,…, n-1})*,

$$x_i(t) = f_{i.t}(y_i(t-1), y_i(t-2), \ldots), \forall t$$

where $f_{i,t}$ can be any causal function. Then they have shown that for a class of degraded channels, the found achievable rate is the exact capacity.

Note that all the relay channel rates, established in special cases, are proved to be achievable using information-theoretic coding strategies, for example, we quote the *block Markov superposition encoding*, *regular block Markov encoding*, and the *windowed decoding*. These coding strategies have been clearly explained and compared in [56]. These coding strategies are too complex for easy and practical implementation, as they apply to large sequences satisfying some mathematical assumptions. Even if such theory is found, it is too theoretic for a practical implementation. But the theory will be able to tell communication designers how close they are to the optimality and perhaps suggest some means of improving the communication rates.

Now, if we look how to make use of the advantages of the relay channel in practice, many solutions have been proposed. The strategies of encoding and processing the data through the relay channel are presented using realistic solutions. In [57], the author has established the relay capacity upper and lower bounds, when the relay uses the TDD mode: the relay is in receiving mode a fraction *a* of the time period, and in the transmit mode during the remaining time *(1- a)*. The transmitter on the other hand transmits with different powers in the two different relay modes. Applying the proposed scheme to a 4 terminal network, Host-Madsen has shown a gain of *8,9dB* in terms of outage capacity.

Then, Laneman in [58], developed energy-efficient transmission protocol for the relay channel network: the signal forwarded by the relay is taken into consideration by the decoder if its power is higher than a fixed threshold, else only the direct transmitted signal is taken into consideration. He has shown that the proposed protocol offers diversity gains over single- and multi-hop transmission.

4.2.4 First Concepts based on enhanced existing concept

4.2.4.1 Description of first relaying concepts for hot spot/area scenarios and their functional elements

4.2.4.1.1 Important features for fixed relays below roof top in hot spot/area scenarios

There are some important features the relay based system should fulfill in order to be successful in the envisaged hot spot/area scenario. Theses features are a guideline for the development of the relaying concept and technology.

4.2.4.1.2 The MAC-frame based relaying concept [65]

Future broadband radio interface technologies and the related high multiplexing bit rates will dramatically increase the traffic capacity of a single Access Point (AP), so that it is deemed very unlikely that this traffic capacity will be entirely used up by the user terminals roaming in an APs service area. This observation will be stressed by the fact that future broadband radio interfaces will be characterised by a very limited range due to the very high operating frequencies (5 GHz) expected. Furthermore, future broadband radio systems will suffer from a high signal attenuation due to obstacles, leading either to an excessive amount of APs or to a high probability that substantial parts of the service area are shadowed from its AP. By means of traffic performance evaluation, in this section establishes that a system based on fixed mounted relay stations is well suited to overcome the problems mentioned. The section is organised as follows. The introduction explains the advantages of relaying, presents fundamentals on how the proposed relaying concepts works in general and finally explains how to "misuse" existing standards to enable relaying in the time domain for wireless broadband systems based on a periodic Medium Access Control (MAC) frame, as used in IEEE802.11e, 802.15.3, 802.16a and HIPERLAN2 (H2). The latter system is taken to exemplify a detailed solution. Section 4.2.4.1.3 answers the question under what circumstances a relay based 2-hop transmission should be preferred to a 1-hop transmission between Mobile Terminal (MT) and AP. Section 4.2.4.1.5 presents the simulation environment, important parameters and the deployment scenarios used to obtain the performance results, which are given in Sections 4.2.4.1.6 and 4.2.4.1.7. Conclusions are drawn in Section 4.2.4.1.8.

We will focus on the time domain relaying. To illustrate the capabilities and properties of relaying in the time domain, results of a model based analysis of the throughput over distance of a MT from the AP and of the achievable capacity for the scenarios shown in Figure 4-15 are presented.

Based on the relation shown in Figure 4-11 for an 802.11a modem and an analytical calculation of the $C/(I+N)$ expected at certain distances from the AP and/or FRS, we obtain a relation between the packet error rate (PER) and distance of the MT from the AP/FRS. Assuming an ideal Selective Reject-Automatic Repeat Request (SREJ-ARQ) protocol, we have calculated the resulting relation between throughput and distance from the AP/FRS, see the solid curve in Figure 4-7 (left). We assume further that the FRSs have directive transmit/receive antennas to communicate with the AP and an omnidirectional antenna to communicate with MTs. Gain antennas at the FRS result in an improved throughput-distance relation between AP and FRS, as is visible from the dotted curve in Figure 4-7 (left). The throughput of a MT that is served by the FRS (dashed curve in Figure 4-7, left) in general obeys the same throughput-distance relationship that is also valid for MTs served by the AP directly. The dash-dotted curve finally denotes the maximum achievable two-hop throughput of a MT served by the FRS. It is clearly visible that a considerable extension of the radio coverage range can be achieved through the use of the relay station with a 16 dBi gain antenna assumed. Figure 4-7 (right) gives the capacity of the AP sub-cell (horizontal line) and compares that with the FRS sub-cell capacity for the case that the whole AP capacity of the 2-Hop-Cell is made available only to one single FRS with varying FRS receive antenna gain. “Capacity” denotes the achievable aggregate cell throughput under the assumption of uniformly distributed MTs generating a constant bitrate type load [67]. The capacity of the AP (this case is equivalent to the AP operated as a conventional BS) amounts to 22.51 Mbits. The capacity that can be made available at the FRS, i.e., when the whole capacity of the AP is transferred to the area that is covered by one of the FRSs, amounts, depending on the FRS receive antenna gain, to values between 2.7 Mbits for 0 dBi gain and 15.87 Mbits for 30 dBi gain. The gap between the two curves in Figure 4-7 (right) denotes the capacity that has to be invested into the extension of the coverage range by means of relaying.

Characteristics of the Relaying Concept

The properties of our relay concept and the benefits that can be expected are as follows:

Radio Coverage can be improved in scenarios with high shadowing (e.g. bad urban or indoor scenarios). This allows to significantly increase the Quality of Service (QoS) of users in areas heavily shadowed from an AP. **The extension of the radio range** of an AP by means of Fixed Relay Stations (FRS) allows to operate much larger cells with broadband radio coverage than with a conventional one-hop system. The FRS concept provides the **possibility of installing temporary coverage** in areas where permanent coverage is not needed (e.g. construction sites, conference-/meeting-rooms) or where a **fast initial network roll-out** has to be performed. The wireless connection of the FRS to the fixed network **substantially reduces infrastructure costs**, which in most cases are the dominant part of the roll-out and operations costs. FRS only need mains supply. In cases where no mains is available, relays could rely on solar power supply. A **standard-conformant integration of the relays into any MAC frame based system** would allow for a **stepwise enhancement of the coverage region** of an already installed system. Investments in new APs can be saved and any hardware product complying to a wireless MAC frame based standard is possible to be used without modifications. The proposed relay concept can be recursively used to extend the radio coverage range of a single AP by multi-hop links.

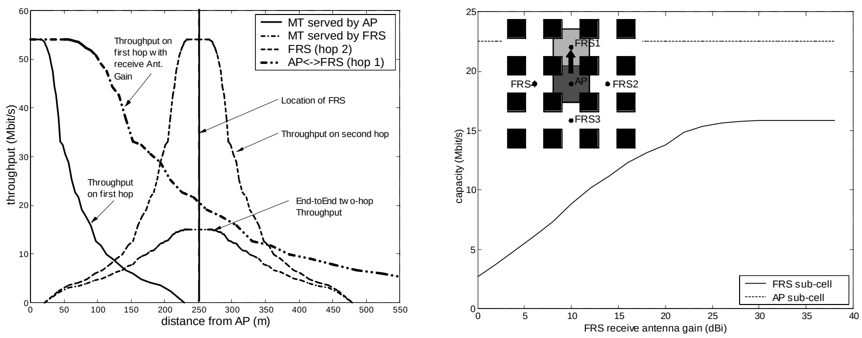


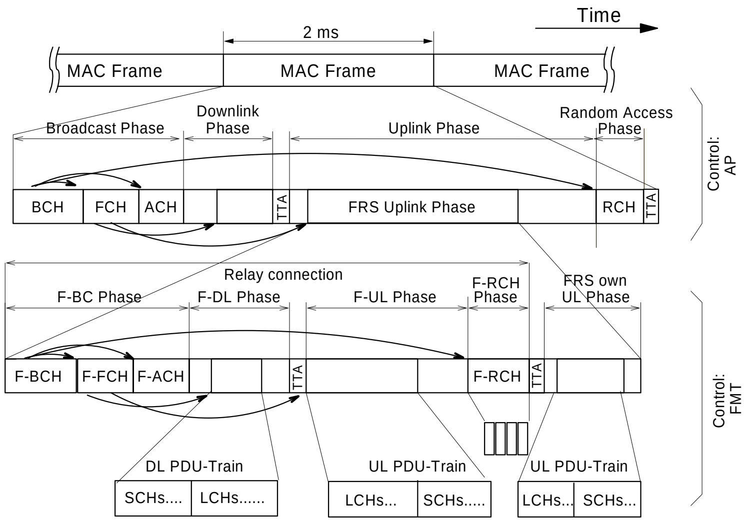
Figure 4-7: Left: Throughput for separate hops and end-to-end for MTs served by FRS (16dBi FRS receive antenna gain)
Right: Capacity of the AP in single-hop mode and capacity of a FRS

In this case, a FRS serves another FRS according to the needs besides serving the MTs roaming in its local environment. It is worth mentioning that we focus on relaying in layer 2 by means of what is called a bridge in Local Area Networks (LANs).

Realisation of MAC frame based Relaying - Example: HIPERLAN2

The HIPERLAN2 (H2) system is used here as an example to explain how MAC frame based protocols as 802.11e, 802.16a (HIPERMAN) and the recently adopted 802.15.3 can be applied to realise relaying in the time domain. All the MAC and PHY functions addressed here are existent in all these wireless standards and no changes of the existent specifications are needed for relaying. However, either the Logical Link Control (LLC) or MAC layer now needs a store-and-forward function like that known from a bridge to connect LANs to each other. In the description of a H2 relay we also use the term Forwarding when referring to Relaying. H2 specifies a periodic MAC frame structure, Figure 4-8. In the Forwarding Mode (FM) both signaling and user data are being forwarded by the FRS. An FRS operating in FM appears like a directly served MT to the AP. **Therefore, this does not preclude the possibility of allowing any MT to act as relay to become a Mobile Relay Station (MRS).** MTs are referred to as Remote MTs ((R)MTs) if they are served by a FRS.

The capacity of the MAC frame (see Figure 4-8, upper part) is assigned dynamically in a two-stage process [68].



BCH Broadcast Channel	SCH Short Channel	F-DL Forward DL
FCH Frame Channel	LCH Long Channel	F-UL Forward UL
ACH Access Feedback Channel	TTA Transceiver Turn Around	F-ACH Forward ACH
RCH Random Access Channel	F-BCH Forwarder BCH	F-RCH Forward RCH

Figure 4-8: Standard-conformant enhancements of the H2 MAC frame

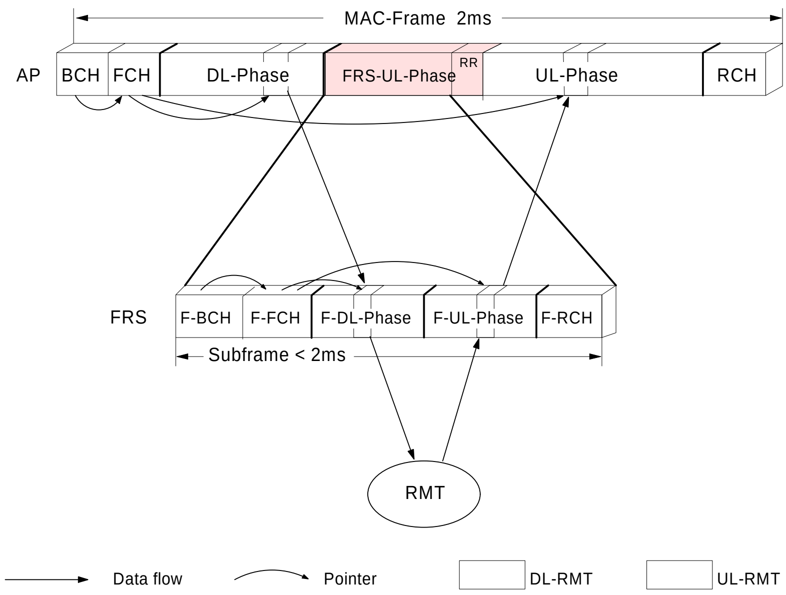


Figure 4-9: Data flow using a sub-frame in 2-Hop mode

The transmit capacity for terminals directly associated to the AP (FRSs and MTs) is allocated by the AP. An FRS appears to the AP like a MT but sets up a Sub Frame (SF) structure, which is embedded into the H2 MAC frame structure of the serving AP (refer to Figure 4-8, bottom).The SF structure has available only the capacity assigned by the AP to the FRS.

This capacity is dynamically allocated by the FRS to its RMTs according to the rules of the H2 MAC protocol. Using this scheme, the FRS needs one transceiver only. The SF is generated and controlled by the FRS (shown in Figure 4-9) and it is structured the same as the MAC frame used at the AP. It enables communication with legacy H2 terminals without any modifications. It implements the same physical channels as the standard H2 (F-BCH, F-FCH, F-ACH, F-DL, F-UL and F-RCH), which carry now the prefix "F-" to indicate that thy are set up by the FRS. A RMT may also set up a SF to recursively apply this relaying concept in order to cascade multiple relays.

Figure 4-8 shows the functions introduced to the H2 MAC frame to enable relaying in the time domain. The capacity assigned in the MAC frame to the FRS to be used there to establish a SF is placed in the UL frame part of the AP. When the FRS is transmitting downlink, the data is addressed properly to its RMT and the AP will discard this data accordingly. The same applies for data transmitted from the RMT to the FRS. The capacity to exchange the data between AP and FRS has to be reserved as usual in both UL and DL directions on request by the FRS [68]. A very similar operation is possible by using the Hybrid Coordinator Access in IEEE802.11e[69].

4.2.4.1.3 ARQ-Throughput 1-Hop vs. 2-Hop

The question arises under what circumstances relaying would be beneficial, i.e. when a 2-hop communication is preferential to one hop. Figure 4-10 shows analytical results [70] comparing the throughput achieved with 1-hop and 2-hop transmission for the two scenarios depicted in the upper right corner of the figure under Line of Sight (LOS) radio propagation.

It is assumed that the FRS is placed at half the distance between the AP and the (R)MT. It turns out that from a distance of 370 m onwards, the 2-hop communication delivers a somewhat higher throughput than 1-hop, as marked by the shaded area.

Relay based 2-hop communication provides another considerable benefit already mentioned in Section 0: it is able to eliminate the shadowing caused by buildings and other obstacles that obstruct the radio path from an AP. An example of this is given by the scenario in Figure 4-10 together with the throughput gain (shaded) resulting from relaying.

In this scenario, the AP and the (R)MT are shadowed from each other by two walls that form a rectangular corner, e.g. a street corner. The COST259 propagation model (see Section 0) was used and the walls were assumed to have an attenuation of 11,8 dB each. The shaded area highlights that the 2-hop communication gains over one hop, starting at a distance of 30 m only.

The two examples establish that relaying is of advantage for both, increasing the throughput close to the cell border of an AP (under LOS conditions) and for bringing radio coverage (and throughput) to otherwise shadowed areas.

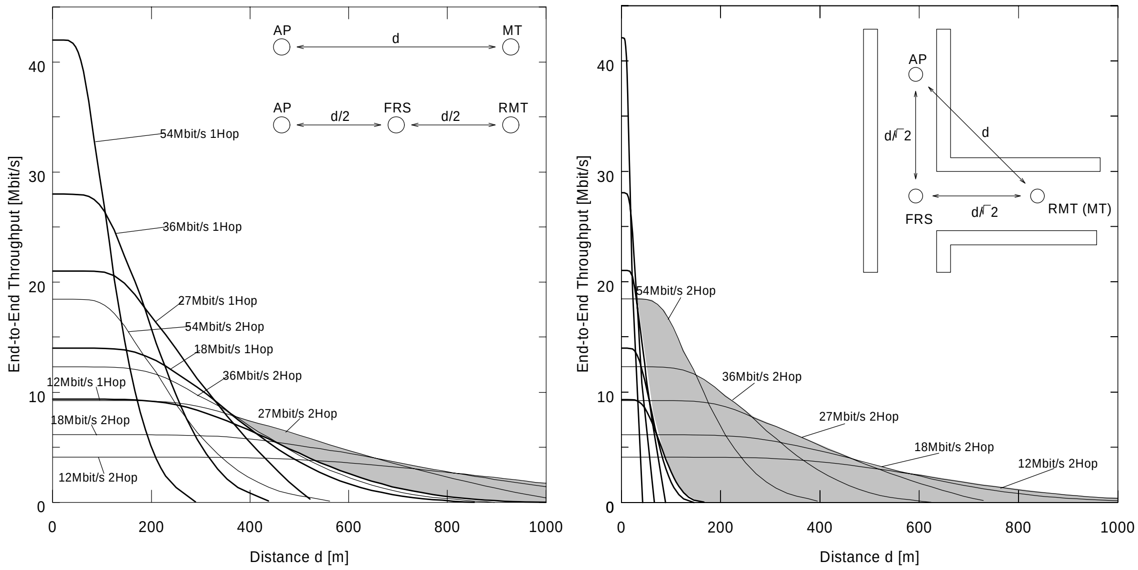


Figure 4-10: Comparison of the maximum achievable End-to-End Throughput over Distance for a 1- and 2-Hop Connection with ARQ

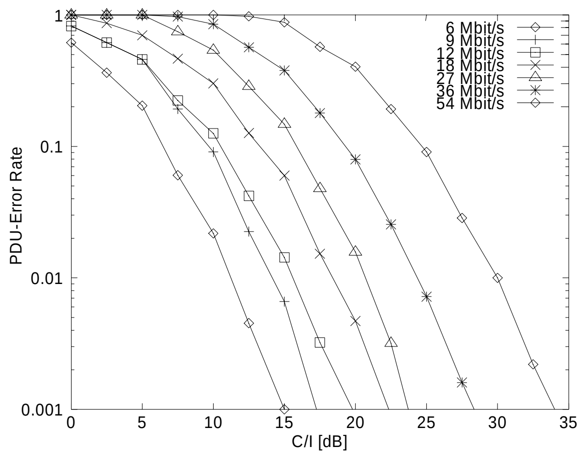


Figure 4-11: PDU-Error Probability for varying C/(I + N) and PHY-mode [22]

It has been explained by means of Figure 4-7 that relaying is consuming part of the capacity of an AP, since the relayed data has to go twice over the radio channel. It has been shown [71] that for relay based deployment concepts (like the one shown in Figure 4-13, left) MTs served at different relays that belong to the same AP can be served at the same time, whereby the capacity loss introduced by 2-hop communications can be compensated to a great extent.This capacity loss can even be turned into a substantial gain, if directive antennas are used at FRS as is shown in Section 4.2.4.1.6. Even if there is still a capacity loss resulting from a relay based system, this concept is able to trade the capacity available at an AP against range of radio coverage [72].

The trend towards increasing transmission rates resulting from further developed radio modems tends to provide an over capacity in the cell area served by an AP, especially in the first months/years after deploying a system. Relays substantially increase the size of the service area thereby increasing the probability that the capacity of an AP will be used effectively.

The next sections present a simulation-based performance evaluation of a relay-based system in a Manhattan-type environment.

4.2.4.1.4 ARQ throughput under delay constraints

The question arises under what circumstances relaying would be beneficial, i.e. when a 2-hop communication is preferential to one hop. Figure 4-12 shows analytical results comparing the throughput achieved with 1-hop and 2-hop transmission for the two scenarios depicted in the upper right corner of the figure under Line of Sight (LOS) radio propagation.

Again it is assumed that the FRS is placed at half the distance between the AP and the (R)MT. It turns out that from a distance of 370 m onwards, the 2-hop communication delivers a somewhat higher throughput than 1-hop, as marked by the shaded area.

For a maximum end-to-end delay of 10 ms the 2-hop connection (red curve) outperforms the 1-hop connection (blue curve) in an open space scenario (left part of Figure 4-12) for distances of more than 375 m. Again the relay based system shows its real strength in the shadowed scenario (right part of Figure 4-12), where the 1-hop case (blue curve) is outperformed by the 2-hop system (red curve) already after a few meters.

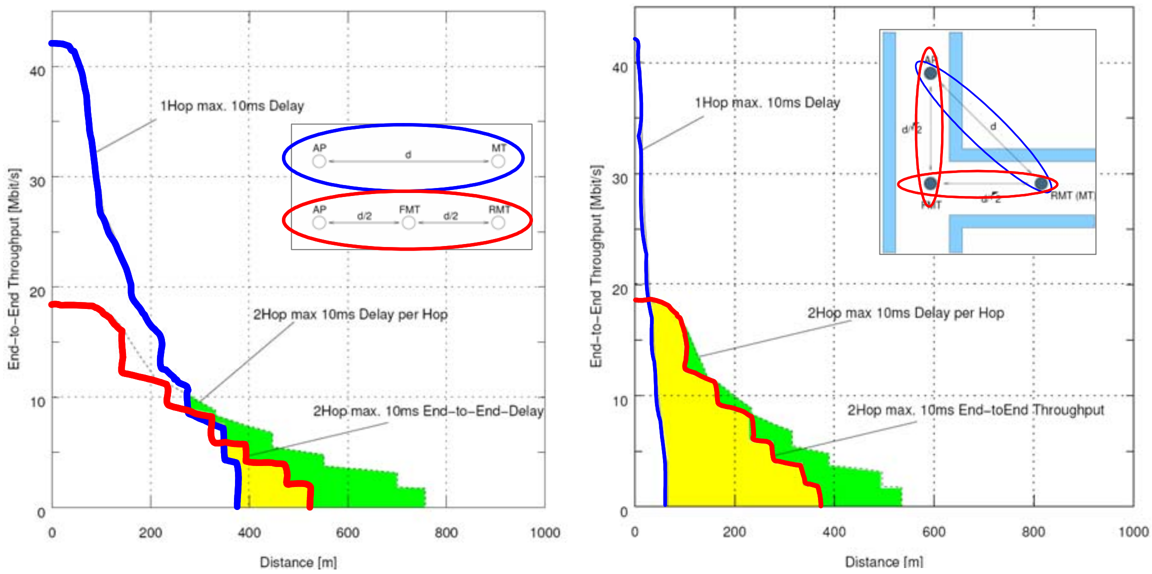


Figure 4-12: Comparison of maximum achievable end-to-end throughput for 1-hop and 2-hop system under delay constraints

4.2.4.1.5 Scenario and Simulation Environment

Scenario: Dense Urban Hot Area Coverage

The dense urban environment with a high degree of shadowing has been identified as a scenario especially suited for deploying a relay based wireless broadband network. The Manhattan grid scenario [74] has been taken for the following investigations, see Figure 4-13. The most important parameters of the scenario are the block size of 200 m and the street width of 30 m. The deployment scenario without relays (see Figure 4-14) is explained in detail in the benchmarking annex.

All three single-hop shown in Figure 4-14 require that a cellular coverage in the Manhattan scenario would have to rely on LOS, leading to a high number of APs. After having performed benchmarking simulations with a single-hop system in this scenario, we study the impact of covering the the same area with a system based on relaying. The basic building block, which consists of an AP and 4 FRSs is shown in Figure 4-13 (left). It has the potential to cover a much larger area than one Single-Hop AP. Figure 4-15 (top) shows the cellular deployment of these building blocks for various cluster sizes. Owing to the high attenuation caused by the buildings, only those co-channel interferers have to be taken into account that are marked in the figure in black and the reuse distance is indicated by the black arrows. For the cluster sizes N=2/3/4 we obtain reuse distances D=1380 m/2070 m/2760 m.

Scenario: Wide Area Coverage

The low coverage range that wireless broadband systems exhibit at high bitrates is shown in Figure 4-10 (left). In a conventional 1-hop hexagonal cellular approach, this leads to a large number of APs required for continuous coverage. It has already been suggested that the use of fixed relays can help to increase broadband radio coverage and thus reduce the number of APs needed. Figure 4-13 (right) shows the basic element (further referred to as "cell") used to achieve wide-area-coverage in a cellular approach. It consists of an AP and 3 surrounding FRSs which can be embedded into a hexagonal cell structure. We consider a coverage radius for a single AP or FRS of R=200 m. The result is that a relay based cell, which consists of 4 sub-cells has a radius of R=346 m. According to Figure 4-15 (bottom), different cluster sizes (N=3,7,12) can be realized just like in a traditional hexagonal cellular approach.

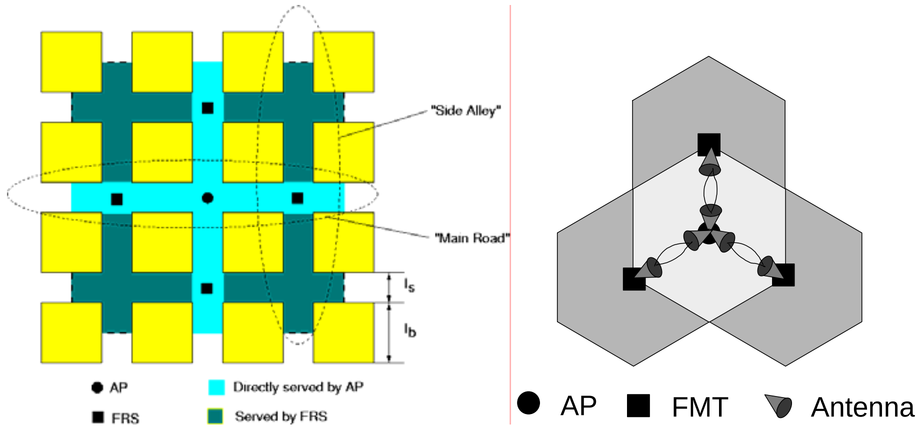
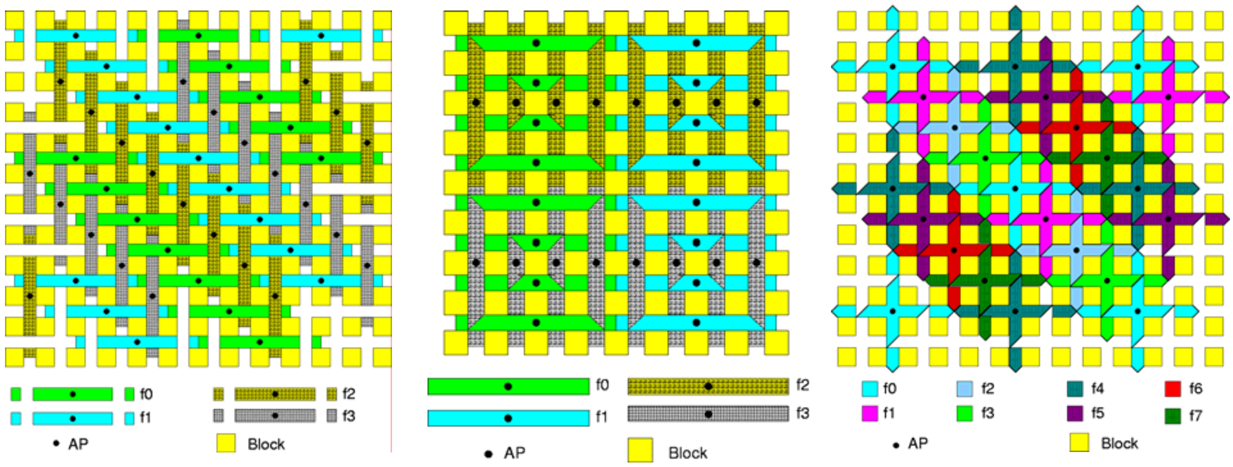


Figure 4-13: Left: Relay-based cell with four relays (below rooftop) in the Manhattan scenario Right: Relay-based cell with three relays (above rooftop) in a wide-area scenario

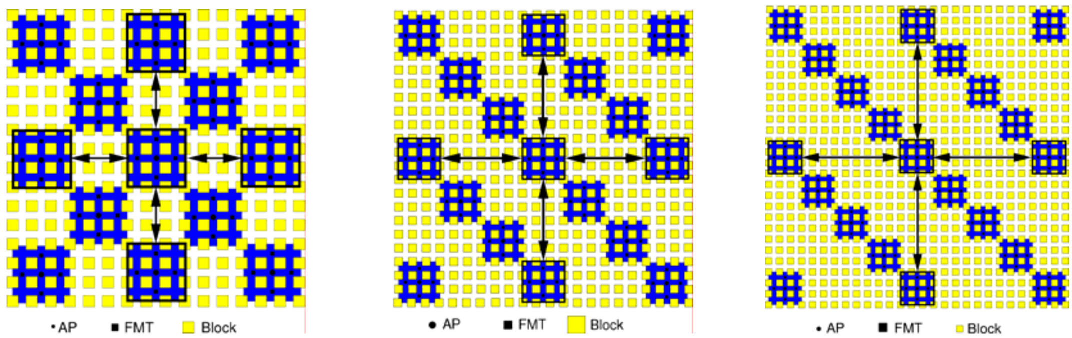
Air Interface

All of the MAC frame based air interfaces mentioned above will operate in the 5 GHz licence-exempt bands (300 MHz in the US, 550 MHz in Europe, 100 MHz in Japan). We assume for the following studies that the physical layer (PHY) uses an OFDM based transmission with 20 MHz carrier bandwidth subdivided into orthogonal sub-carriers. The modem is assumed conformant to the IEEE802.11a standard. As indicated in the introduction of this section, the 5 GHz frequency range is characterised by high attenuation and very low diffraction, leading to low radio range, which is one of the key problems addressed by the proposed relaying concept.



(a) UMTS30.03 deployment (b) horiz./vert. deployment (c) APs on intersections

Figure 4-14: Possible single-hop deployments for coverage of a Manhattan scenario



(a) Cluster Size N = 2 (b) Cluster Size N = 3 (c) Cluster Size N = 4

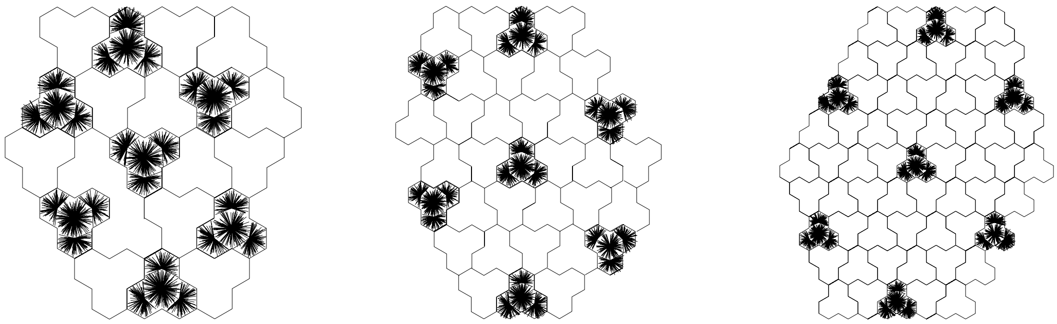


Figure 4-15: Top: City-wide coverage with relay-based cells for different clustersizes N=2,3,4
Bottom: Wide-area coverage with relay-based cells for different clustersizes N=3,7,12

Link-Level Performance: The same assumptions as in the H/2 benchmarking Annex have been made.

Propagation Models: Also the propagation conditions have been chosen similarly to those in the Benchmarking Annex, to ensure the comparability of results.

Other Parameters of the Simulation Model: To determine whether a MT should be served by the AP directly or via a FRS, the path loss between AP and MT is assessed. If it is higher than a certain threshold, the MT is associated to the closest available FRS ("closest" in terms of pathloss). The traffic load is assumed to be constant bitrate, which is a reasonable assumption when investigating the maximum achievable end-to-end throughput.

4.2.4.1.6 Simulation Results

This section presents the performance evaluation results obtained by stochastic-event driven simulation. Results for the Downlink (DL) direction are presented here only, since the main effects that can be observed are quite similar in Uplink (UL) and DL directions, a result which is partly due to the Time Division Duplex air interface studied. We compare these results against single-hop benchmarking results that can be found in the appropriate annex.

Simulation Results with Fixed Relay Stations: Manhattan Scenario

Simulations with fixed relays as introduced in Section 0 have been performed for the cluster sizes N=2/3/4, cf. Figure 4-14. Figure 4-16 (left) shows two sets of curves in one graph:

The C/(I+N) versus the distance of a MT from the AP (marked with 1. hop) and the C/(I+N) encountered by MTs being served by a FRS (marked with 2. hop). The FRS is located at a distance of 230 m from the AP on the "Main Road" (cf. the pictogram in the figure and Figure 4-13). This explains the peak of the C/(I+N) curve visible at that distance. Each set of curves has the cluster size N as a parameter. As expected, the curves with N=2 show the lowest C/(I+N) values. Figure 4-16 (right) shows the C/(I+N) situation in the "Side Alley" of the relay based cell. Like on the first hop, the situation for the MTs is almost similar to that of the MTs served directly by the AP in the single hop case, with the difference that the next LOS co-channel interferer is more than 780 m away, leading to lower interference and thus to a C/(I+N) which is approx. 4 dB higher than in the single hop case.

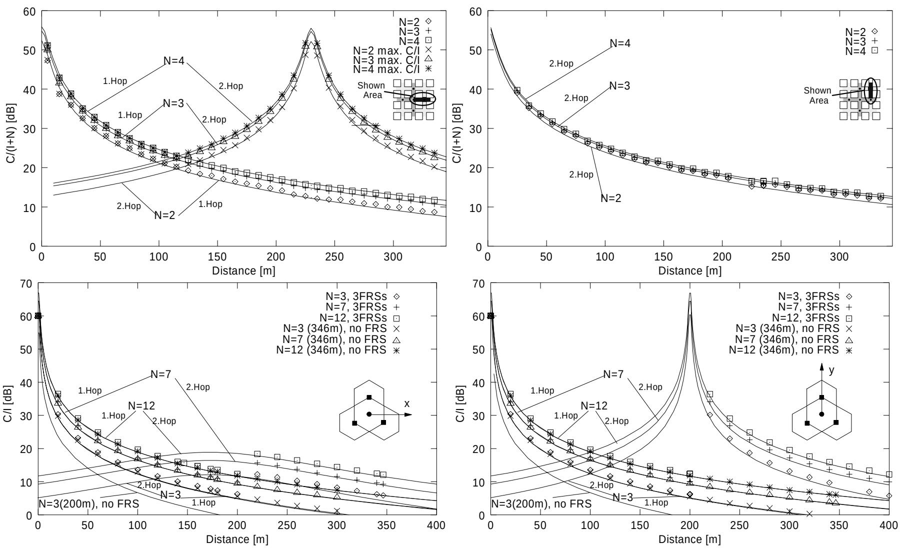


Figure 4-16: Top: DL C/(I+N) vs. Distance from (R)MT to AP respectively FRS for varying cluster sizes (2, 3, 4) using relays (Left: “Main Road”, Right: “Side Alley”, Lines: analysis, Markers: simulation)
Bottom: DL C/(I+N) vs. Distance from (R)MT to AP respectively FRS for varying cluster sizes (3,7,12) using relays (Left: “x-axis”, Right: “y-axis”, Lines: analysis, Markers: simulation)

A maximum TP of approx. 4-5,5 Mbit/s (depending on N) can be made available even at the cell border of the second hop, in an area which has no direct coverage of the first hop at all and which would require an additional AP in a single hop scenario (see Figure 4-14, right). Figure 4-16 (upper left and right) shows the resulting 2-hop TP for MTs on the "Main Road" and the "Side Alley" respectively with omni directional antennas used at AP, FRS and MTs. Obviously, the TP on both the first and the second hop depends on the cluster size N.

The relatively flat slope of the curves for the second hop indicates that the TP is upper-bounded by the capacity available at the FRS from the AP. More capacity can be provided when using gain antennas at FRSs and omni antennas at AP and MT, with the FRS serving its MTs with an omni antenna. The improvement in TP for the outer range of the relay based cell with an 11,8 dB gain at the FRS can be seen when comparing the left and right hand graphs in Figure 4-17. As predicted in Figure 4-7 (right), the resulting higher TP on the first hop allows a FRS to have much more capacity available in its service area.

At a gain of 11,8 dB, which is an intermediate value according to Figure 4-7 (right), an increase in max. TP of up to 80% (from 8 Mbit/s to 14 Mbit/s) can be observed on the second hop, both on the "Main Road" and in the "Side Alley".

Simulation Results with Fixed Relay Stations: Wide-Area Scenario

Simulations with fixed relays have also been performed in a wide-area above-rooftop deployment for the cluster sizes N=3/7/12, cf. Figure 4-15 (bottom). Figure 4-16 (bottom left and right) shows the C/(I+N) over distance of the MT from the AP respectively the FRS. The FRS is located at a distance of 200m from the AP along the y-axis (see pictogram). This explains the characteristic peak of the curves denoted "2. Hop". It is further visible in both sub-figures that the impact of the cluster-size on the expected C/(I+N) values is considerable. For reference, the figures also show the C/(I+N) curve for the N=3 and R=200 m one-hop scenario. It shows that the relay deployment helps to considerably improve the C/(I+N) values. The left-hand side of Figure 4-18 shows the maximum achievable Downlink End-to-End throughput versus the distance (in x- and y-direction) of a MT from the AP (marked with 1. Hop) and the throughput encountered by MTs being served by a FRS (marked with 2. Hop).

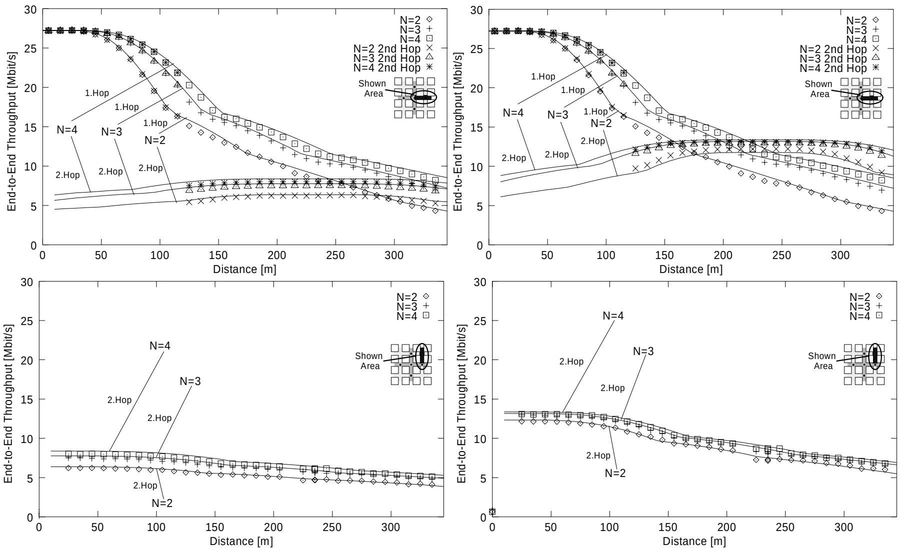


Figure 4-17: DL End-to-End-Throughput vs. Distance from (R)MT to AP respectively FRS for varying cluster sizes (2, 3, 4) using relays (left: using omni antennas only, right: with 11,8 dB receive antenna gain at the FRS (Lines: analysis, Markers: simulation)

The FRS are located at a distance of 200 m from the AP, e.g. in the y-direction (shown in the pictogram). This explains the maximum of the throughput curve for the second hop visible at that distance. Each set of curves has the cluster size *N* as a parameter. As expected, the curves with N=3 show the lowest throughput values, owing to the highest encountered interference. The right-hand side of Figure 4-18 shows the maximum achievable Downlink End-to-End throughput when an antenna gain of 11,8 dB is assumed between AP and FRS. Again, the upper figure represents the situation along the x-axis, while the lower figure refers to the y-axis of the relay based cell (also refer to the small pictograms included).

Like on the first hop, the situation for the MTs is almost similar to that of the MTs served directly by the AP in the single hop case (included for reference with a cell size of R=346 m).Depending on the cluster-size, the maximum End-to-End throughput along the y-axis improves for ranges greater than 220 m (N=3), 280 m (N=7) and 320 m (N=12) when relay stations are used instead of a single hop deployment. Along the x-axis improvements can be observed for N=3 and N=7 (ranges > 250 m and 325 m). In Section 4.2.4.1.3 we additionally show the result for the case where no co-channel interferers are present. In that case, improvements of the maximum throughput can be observed for distances greater than 370 m.

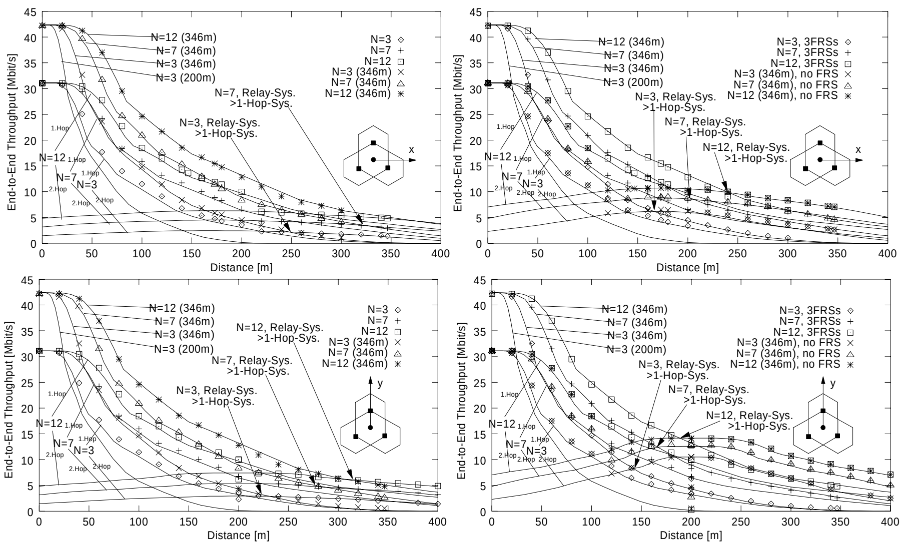


Figure 4-18: Maximum DL End-to-End-throughput vs. Distance from (R)MT to AP respectively FRS for varying cluster sizes (3, 7, 12) and sub-cell radii (200 m, 346 m) using relays (left: using omni antennas only, right: with 11,8 dB antenna gain between AP and FRS (Top: "x-axis", Bottom: "y-axis", Lines: analysis, Markers: simulation)

If an additional antenna gain is assumed between AP and FRS, the advantages of the FRS concept can already be observed at about 140 m (N=3), 170 m (N=7) and 190 m (N=12) along the y-axis, while - along the x-axis - the throughput of the two hop system outperforms the one-hop system starting at 170 m (N=3), 200 m (N=7) and 240 m (N=12). In general, a considerable improvement compared to the deployment without gain antennas can be observed. In addition, a more homogeneous distribution of the maximum achievable throughput can be noticed, which is especially beneficial in areas close to the cell border. The tighter the frequency reuse, the smaller becomes the minimal range where the use of FRSs is beneficial. Also, the number of necessary frequency channels is reduced with lower cluster sizes. This allows to use more frequency channels per cell and thus to increase an operators network capacity. When using FRSs, even in a cluster with N=3 the cell border can be served at sufficient quality due to the range extension. The gain obtained from the relaying scheme justifies transmitting the information twice.

The results given above are for the comparison of one- and two-hop cells with the same cell area (equal AP density). If an N=3-cluster with 200 m-cells is compared with a N=3 relay cell with 200 m sub-cells (equal site density), the advantages of the relay-based concept already become visible at distances > 30 m from the AP.

Table 4-3: Average Cell Capacity and spectral efficiency for a Cell with 10 MTs and Exhaustive Round Robin (ERR) Scheduling, comparing three Manhattan Single-Hop deployments (see benchmarking annex) with Multi-Hop deployment (with and without Receive Antenna Gain at the FRSs)

Scenario	Used # of Freq.	Cell Size [m2] / 103	Cell Capacity [Mbit/s]	Spect. Efficiency [bit·s-1·Hz-1·m-2]
1-Hop (UMTS 30.03)	4	25,8	21,04	10,19
1-Hop (horiz./vert. depl.)	4	25,8	20,01	9,69
1-Hop (APs on cross.)	8	53,4	20,24	2,37
2-Hop N=2	2	116,0	7,26	1,56
2-Hop N=3	3	116,0	9,03	1,30
2-Hop N=4	4	116,0	9,80	1,06
2-Hop N=2, +11,8 dB	2	116,0	10,72	2,31
2-Hop N=3, +11,8 dB	3	116,0	12,7	1,82
2-Hop N=4, +11,8 dB	4	116,0	13,34	1,44

4.2.4.1.7 System Capacity and Spectral Efficiency

In addition to the End-to-End throughput studied in the previous sections, the system capacity, i.e. the aggregate traffic that can be carried in a well-defined service area and a certain amount of used spectrum is an important measure to assess a system’s performance. To optimise a system, it is very important to have a clearly defined optimisation goal. The relay concept presented aims at providing a cost-efficient broadband coverage that can rapidly be deployed in a relatively large area. Table 4-3 shows the average End-to-End cell throughput for the different 1- and 2-hop deployments in the Manhattan scenario. The table also shows that the coverage area of one AP for the one-hop scenarios is relatively small, indicating that a large number of costly backbone connections is needed to cover the whole service area. From the small cell size and the high cell throughput results a relatively high area spectral efficiency. But a minimum of 4 carrier frequencies is needed in that case to provide continuous coverage. The AP deployment from Figure 4-14 (left) shows only small advantages over the horizontal/vertical placement Figure 4-14, middle). The placement on street crossings (Figure 4-14, right) has the advantage that a larger area is covered per AP, reducing the number of needed backbone connections by a factor of 2. At the same time, a minimum of 8 carrier frequencies is needed to enable continuous coverage. This and the larger cell size lead to a substantial reduction in spectral efficiency, while the average cell throughput changes only slightly.

Table 4-4: Average Cell Capacity and spectral efficiency for a Cell with 10 MTs and Exhaustive Round Robin (ERR) Scheduling, comparing the wide-area cellular Single-Hop deployment (see benchmarking annex) with the Multi-Hop deployment (with and without Antenna Gain between AP and FRSs)

Scenario	Used # of Freq.	Cell Size [m^2] / 10^3	Cell Capacity [Mbit/s]	Spect. Efficiency [bit·s^{-1}·Hz^{-1}·m^{-2}]

Scenario	Used # of Freq.	Cell Size [m^2] / 10^3	Cell Capacity [Mbit/s]	Spect. Efficiency [$bit \cdot s^{-1} \cdot Hz^{-1} \cdot m^{-2}$]
Standard 200m	3	104	6,84	1,10
Standard 200m	7	104	12,2	0,84
Standard 200m	12	104	16,42	0,66
3FRS	3	311	4,21	0,23
3FRS	7	311	7,27	0,17
3FRS	12	311	9,46	0,13
Standard 346m	3	311	6,53	0,35
Standard 346m	7	311	11,42	0,26
Standard 346m	12	311	14,82	0,20
3FRS +11,8dB	3	311	7,44	0,40
3FRS +11,8dB	7	311	11,14	0,26
3FRS +11,8dB	12	311	13,41	0,18

Another reduction of the number of APs needed (to a total factor of 4) can be achieved by using FMTs as proposed in Figure 4-7. This leads to a very cost-efficient cellular coverage of the service area. The 2-hop transmission obviously reduces the cell capacity, an effect that can be reduced through the use of higher re-use distances.

A substantial increase in throughput and cell capacity is achieved through the use of directive receive antennas at the FMTs.When using 2 carrier frequencies, the relay concept with directive antennas achieves roughly the same area spectral efficiency (2.31 bit$\times s^{-1} \times Hz^{-1} \times m^{-2}$) as the 1-hop deployment with APs on street crossings (2.37 bit×s-1×Hz-1×m-2), with the advantage of a lower number of APs and carrier frequencies needed.

Table 4-4 shows the average End-to-End cell throughput for the different 1- and 2-hop deployments in the wide-area scenario. Again, from the small cell size and the high cell throughput results a relatively high area spectral efficiency in the case of the 200 m-cells. However, the interesting observation is that the relay-based system achieves the same area spectral efficiency as a one-hop system with the same overall cell size. At the same time, as we have seen in Figure 4-18, the coverage quality at the cell border is superior in the two-hop case. Under dense frequency re-use (N=3), the two-hop system even exhibits a [14]% higher spectral efficiency (compare lines 7 and 10 of Table 4-4).

4.2.4.1.8 MAC-frame based approach: Conclusions

Modern wireless broadband air interfaces are based on MAC frames, the only exemptions being IEEE802.11a/b/g but 802.11e uses a MAC frame, too. MAC framed air interfaces have been established to be useful for relaying in the time domain by just using the functions available from the existing standards. Deployment concepts using fixed relay stations have been shown to be of high benefit to substantially reduce the effort of interfacing APs to the fixed network (owing to a substantial reduction of APs needed). Relays have been proven to substantially extend the radio coverage of an AP, especially in highly obstructed service areas. Gain antennas at FRSs have been established to substantially contribute to increase the throughput at cell areas far away from an AP.

4.2.4.1.9 Further enhancement for the WINNER relaying concept in the urban hot spot/area scenario

From the first MAC-frame based approach we can derive some pros and cons that can be used together with the introduced features as introduced in 4.2.4.1.1 to further develop the relaying concept. TBD

The simple MAC frame based relaying approach as shown in the sections before it was already demonstrated that the relay based deployment will have some advantages especially in urban scenario with many obstacles to overcome. The relay concept can be further enhanced by exploiting the shadowing effect by means of spatial reuse as multi-hop cells, e.g. in the proposed Manhattan scenario will be fully decoupled, which means they do not interfere each other. In Figure 4-19 a schematic MAC based is shown where always two FRS are able to serve there UTs at the same time [78].

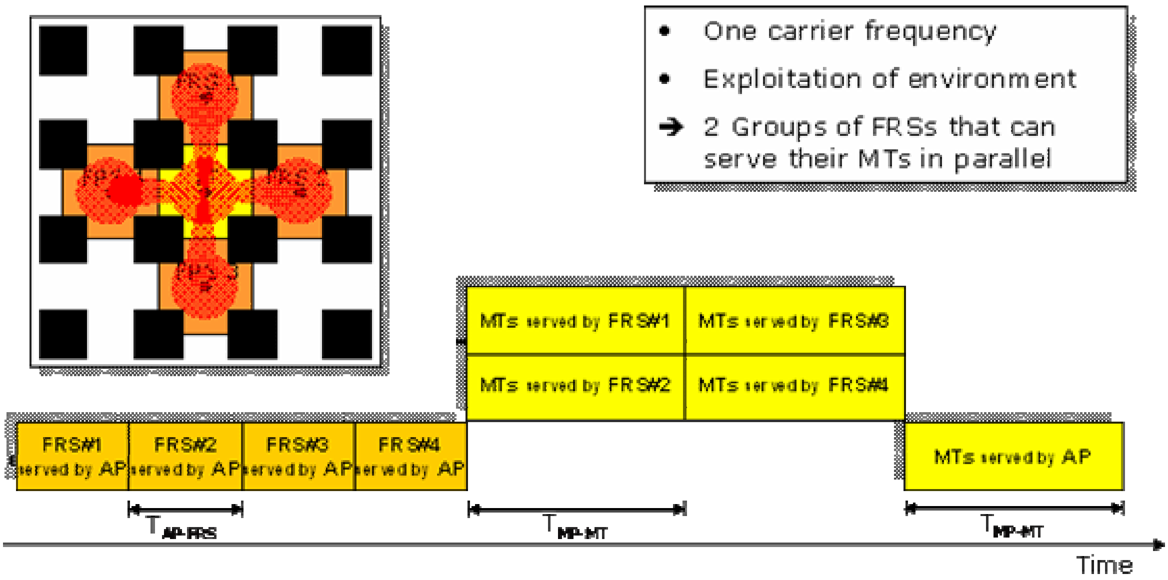


Figure 4-19: Relaying with spatial reuse

Also for further studies is the use channel oriented relays, which have advantages also for mesh networks.

4.2.4.2 Network deployment concept based on fixed relays and a Hiperlan/2 like centralised MAC

4.2.4.2.1 Introduction

With the ever increasing trend to higher bandwidths and the constrains that arise from the available spectrum it is likely that the cell sizes will be small compared to traditional cellular systems operating in the 1-2 GHz Bands. Therefore the number of cells will be large to provide a coverage solution for an area similar to regular cellular coverage. Since the fixed network access is currently seen as the main cost driver for the deployment, multi hop can be regarded as a possibility to counteract this problem in the novel WINNER network concept.

In this section a system concept with relaying capabilities is proposed which will be subsequently extended in the future. This approach will be investigated in terms of throughput and delay in an urban scenario with special focus on the comparison of both cases.

4.2.4.2.2 System Concept

The system concept described in the following sections is used in conjunction with the physical nodes which have been described in Section 2.9 where a detailed description of the node functionality can be found.

Frame Format:

In this proposal a centralized MAC architecture is used that is known as the "TDMA clustering" concept. One node per cluster, called Central Controller (CC), controls the collision free access to the medium of all the nodes within the cluster based on a Resource Request/Resource Grant method similar to 802.16a and Hiperlan. Furthermore it periodically broadcasts system information on its BCH channel.

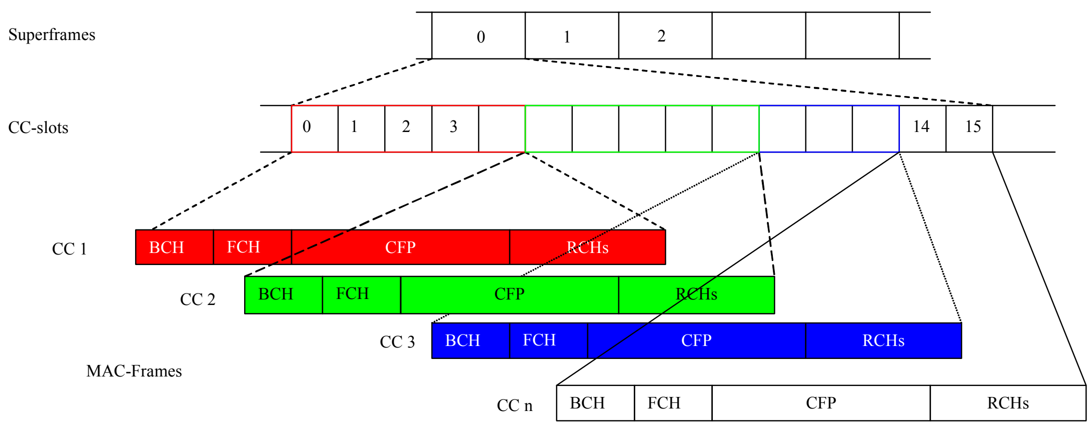


Figure 4-20 Medium sharing among neighbouring clusters

To identify TDMA slots a "SuperFrame" is defined which consists of SuFsize time slots (slot numbers 0 to SuFsize-1). These slots build the units of a resource pool that can be allocated to different CCs.

Unlike to a conventional TDMA scheme like in GSM, these slots do **not** itself build the eventual resource units for specific transmissions. They are just the basis to define the admissible position of a MAC frame structure that belongs to a certain CC. The configuration pattern that maps the slots to CCs repeats every SuperFrame. The re-configuration of the mapping pattern is performed on a long term time basis. I.e. in-between two re-configurations the same slot numbers are assigned to the same CCs.

The MAC Frame consists of 4 sub-frames:

The broadcast control channel (BCH) is transmitted in the first sub-frame It contains general announcements and some status bits announcing the appearance of more detailed broadcast information in the contention free period (CFP).

The FCH carries the information about the structure of the ongoing frame (Resource Grants), containing the exact position of all following emissions, their usage and content type.

The contention free period (CFP) sub-frame can be further divided into a two phases. The first phase contains user specific control information and user data which is transmitted from the CC to another CC/TE. It corresponds to the downlink in the single hop case. Additionally, this phase may contain further broadcast information which does not fit in the fixed BCH field. The second phase of the CFP is used to transmit control and user data from the TEs or CCs to the CC. The TEs have to request bandwidth for one of the following frames in order to get resources granted by the CC. This phase corresponds to the uplink in the single hop case.

The Random access channel (RCH) sub-frame is used by TEs and CCs to which no capacity has been allocated. It is used for the transmission of control information. Non-associated TEs get in first contact with an CC via the RCH and TEs performing handover to have their connections switched over to a new CC using the RCH.

Connectivity between CCs (Relaying support):

Relaying in this concept is based on Mac frame by frame relaying (Figure 4-21). The CCs in one hop distance communicate directly with each other in a way that a CC behaves like a normal node in the neighbouring cluster.

Data that needs to be send to another CC is transmitted in the downlink resource of the sending CC to avoid unwanted interference. Each CC listens to the control channels of neighbouring CCs and will therefore notice that data that needs to be received. The data is send in the DL partition of the CPF by the transmitting node at a position that has been pointed to in the FCH. The receiving node knows the position from evaluating the control channel information and is therefore able to receive the data accordingly.

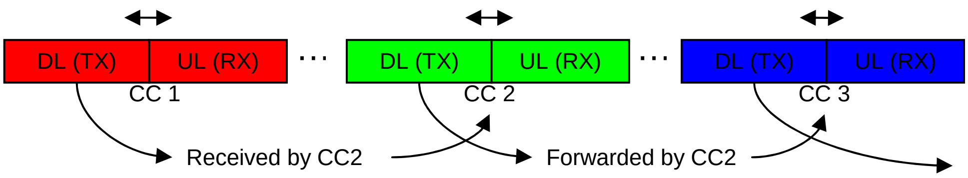


Figure 4-21 Forwarding Concept

In this concept the next hop is determined by a routing entity in the Layer 3. This routing functionality could be greatly simplified compared to Manet Routing algorithms because of the static nature of the system concept. All data designated towards the AP is send to a default route, known as the next hop in the AP direction. Data designated to the MT will be forwarded from the AP and intermediate FRRN based on dynamic routing table information. This information is gathered by location update messages that are transmitted by the mobile node each time a handover to another CC takes place. To detect broken links e.g. in bad transmission conditions a periodic timer and timeout handling is used.

An alternative method in this concept involves the use of transit nodes i.e. nodes that belongs to more than one cluster. Transit nodes relay the data received in one cluster and forward this data into other clusters. They do not have any scheduling and routing abilities and these tasks are performed in the CCs which control the resources that the transit nodes belong to.

4.2.4.2.3 Simulation Environment

Scenario:

The scenario used in the simulations is based on the Manhattan structure introduced by UMTS 30.03 [79]. In order to cope with the computational complexity of the investigations only a small part of the grid has been used for the initial simulations. The grid and the basic parameters of it are given below:

Street Width	30m
Block size	200x200m
Area	4x4 Blocks

Table 4-5 Basic parameters of the simulation scenario

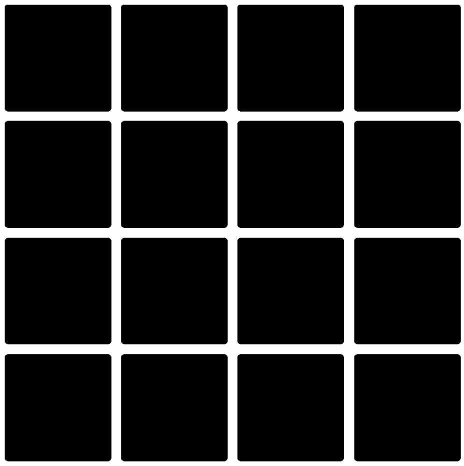


Figure 4-22 Simulation grid

Deployment:

Initial results for 2 possible multi hop scenarios were presented in [80]. It was shown that multi hop offers the maximum benefits if LOS between the source and the destination of the transmission is archived by the forwarding nodes and shadowing conditions are avoided. All access points and forwarding nodes are placed in the street canyons in the multi hop deployment case according to this result. The assumption is that the expected deployment costs of nodes placed above roof top level are considerably higher than nodes placed below rooftop level. In the single hop case the positions of the access points are left constant while the forwarding nodes are removed.

A central access point has been placed in the central street crossing of the simulation scenario (Node number 49). Eight Access Points have been placed around the central AP to simulate two interfering rings. Figure 4-23 shows the positions of all fixed network elements as an example. The respective resource allocated to these elements is indicated with different colours.

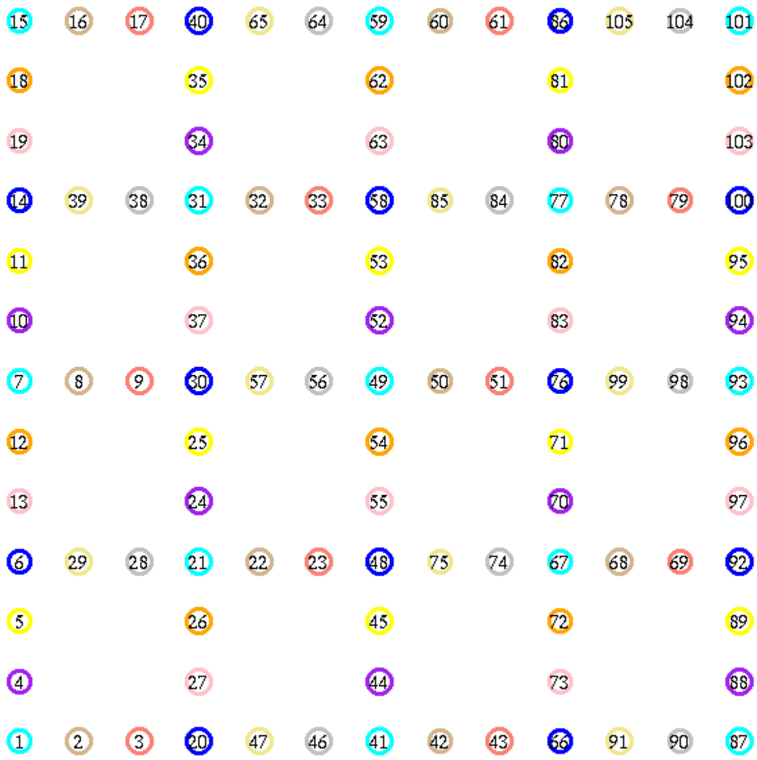


Figure 4-23 Fixed network element positions and resource allocation

In order to calculate the distance and reuse factor of the fixed elements a C/I calculation has been performed for one interfering cell. The result of this calculation is given in Figure 4-24 where R and D refer to the Cell Radius and distance of the co-channel interferer respectively. Distances and reuse was chosen in a way that the CIR at the receiver for fixed NW element communication if sufficient to use physical layer mode 6 and 7.

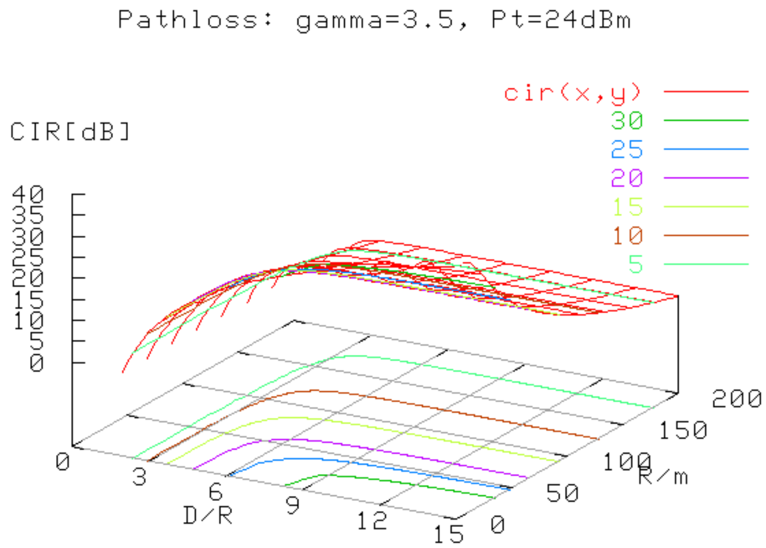


Figure 4-24 CIR Calculation

4.2.4.2.4 Simulation Results

The simulations for this concept are in progress and will be included in D3.2. It is planed investigate the concept and compare it to a traditional single hop deployment in terms of throughput and capacity. A simple investigation to determine the performance bounds based on simple calculation is currently performed and will be presented as well.

4.2.4.3 Network deployment concept based on 802.16a

4.2.4.3.1 Introduction to IEEE 802.16a

Broadband wireless access technologies have gained a lot of attention during the last years. At the same time multi-hop communication is expected to become a major technology in systems beyond 3G. The IEEE standard 802.16a provides either high speed wireless access as well as multi-hop capabilities through the two air-interfaces that are specified in the IEEE standard 802.16a: Point-to-Multi-Point (PMP) and Mesh mode (see Figure 4-25 and Figure 4-26).

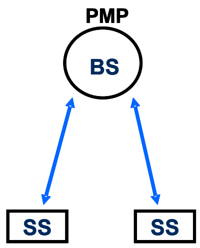


Figure 4-25. Traffic flows for PMP mode

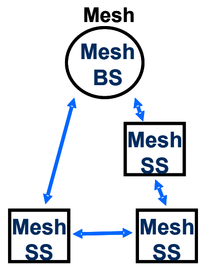


Figure 4-26. Traffic flows for Mesh mode

In PMP mode connections are established among the Base Station (BS) and Subscriber Stations (SSs), hence, data transmissions between two SSs are routed through the BS. Within the Mesh mode, traffic can occur directly among SSs since the protocol permits to set up data connections among neighbours, and supports multi-hop communication. The BS in the Mesh mode is termed Mesh BS. The Mesh BS is the entity that interfaces the wireless network to the backhaul links. It acts like a BS in PMP mode except that not all of the SSs have to be directly connected to the Mesh BS. The MAC protocol for the Mesh mode has been designed to support both centralized (Mesh CS) and distributed (Mesh DS) scheduling, or a combination of both [82], [92].

While the PMP supports both duplex schemes, Time Division Duplex (TDD) and Frequency Division Duplex (FDD), the Mesh mode only supports TDD because of its multi-hop support, i.e., communication between arbitrary SSs.

In the following the description if provided only for the TDD mode of operation for both, PMP and Mesh mode. The FDD mode of operation has no major impact on the frame structure of the PMP mode, except that the transceiver turn-around times (transition gaps) are not needed.

The IEEE Standard 802.16a [82], [92], approved in January 2003, complements the physical (PHY) layer specification for frequencies between 10 and 66 GHz published in the IEEE Standard 802.16 [91][92] with advanced PHY layer design in the frequency band 2 - 11 GHz to cope with non-line-of-sight (NLOS) conditions and multi-paths propagation effects. Different to the single-carrier (SC) version above 11 GHz, it is proposed OFDM for the lower frequency range.

In addition, respective changes to the MAC protocols have been specified for the new PHY specification and the Mesh mode has been added. All PHY and MAC versions of IEEE 802.16 are described in the latest version of the standard [82]. Originally, IEEE 802.16 has been specified as Wireless Metropolitan Area Networks (WMAN). However, in the latest specification there is also a further PHY layer specified, which is based on OFDMA. This version is known as IEEE 802.16e and should support mobile users.

The IEEE 802.16a air interface is described in more detail in the appendix. Motivated by the variety of modes and physical layers that are provided by the IEEE 802.16 standard, it is investigated in the following sections, which mode is best suited to fulfil the requirements of the WINNER air-interface, specifically to support multi-hop communication, and how the PMP mode has to be changed to support multi-hop connections.

4.2.4.3.2 Multi-hop Approach for PMP Mode

One basic feature of the 4G air-interface will be the multi-hop support. The PMP mode does not support multi-hop communication, yet. For that reason a modification of the standardized PMP mode is proposed, which is shown in figure below.

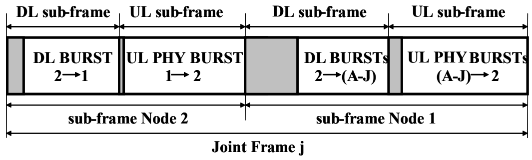


Figure 4-27. Sub-frame structure in PMP mode to support multi-hop communication

It is assumed that SSs are divided into different groups based on their position. Every group corresponds to an area in the network, an SS is elected as responsible of the data scheduling in the served area and is termed PMP Base Station (PMP BS). The PMP BS has direct radio contact to the SS in the respective area. This concept is similar to the Central Controller (CC) selection in HIPERLAN/2 [86]. It is further assumed that signalling among PMP BSs is possible due to direct connections among them, which are necessary to coordinate the scheduling for all links in the wireless network. The idea is to join n frames for n-1 connections between n PMP BSs, respectively hops, obtaining a unique Joint Frame, which is composed by n sub-frames, which closely corresponds to the concept described in [86].

The management mechanism for the Joint Frame is explained in detail for the example scenario depicted in Figure 4-28 considering n = 2 hops.

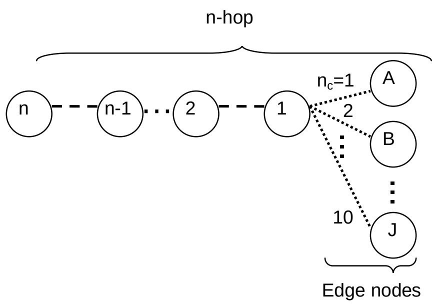


Figure 4-28: Simple tree network topology

In Figure 4-27, the Joint Frame is composed by two DL sub-frames respectively for the data connection between node 2 and node 1 (PMP BS 2 and 1, respectively) and data connections among node 1 and n_c edge nodes. Every MAC PDU transmitted on the link between node 2 and node 1 requires information on the final node that has to be addressed. We have introduced the Mesh sub-header in the MAC PDU, which contains the CID of the last hop towards the edge nodes (SSs). A short preamble is required for every uplink connection, hence a short preamble is used for every UL PHY BURST in the UL sub-frame for data transmission among edge nodes and node 1. On the contrary, the node 1 transmits only one UL PHY BURST, thus only one short preamble is introduced. The Mesh sub-header is introduced in every MAC PDU transmitted from node 1 to node 2 in order to address the source SS. Within the sub-frame handled by node 2 only one Bandwidth Request opportunity is used by node 1 for requiring resources from the node 2.

4.2.4.3.3 Analytical Performance Evaluation of PMP and Mesh Mode

In this section the theoretical achievable throughput for IEEE 802.16a will be derived. To become an impression about the performance of the MAC protocol the efficiency for different packet sizes will be defined and the maximum value is graphically displayed. Furthermore, the efficiency for the PMP mode, Mesh CS, and Mesh DS are compared.

The “maximum theoretical throughput per OFDM symbol” is defined as:

$$\Theta_{PHY/symbol} = \frac{N_{SD} \cdot R_C \cdot \log M}{T_{symbol}} \tag{1}$$

where N_{SD} is the number of data carriers per OFDM symbol. In the PHY layer specification N_{SD} equals 192. R_c is the coding rate and $\log M$ the number of bits per carrier. T_{symbol} is the duration of the OFDM symbol and depends on the chosen OFDM parameters.

The mathematical expression for the maximum "net throughput on MAC layer" is the sum of the payloads per frame duration:

$$\Theta_{Net/MAC} = \frac{\sum Payload}{T_{FRAME}} \tag{2}$$

The efficiency shall be defined as the ratio between the "maximum net throughput on MAC layer" and the "maximum theoretical throughput per OFDM symbol":

$$\eta = \frac{\Theta_{Net/MAC}}{\Theta_{PHY/symbol}} \tag{3}$$

The maximum net throughput on MAC layer depends on the payload length $L^k_{payload}$ of the MAC PDU and the PHY mode BpS_m (m=1, 2, 3, 4, 5, 6) according to **Table 7-1**, on p. 232.

$$\Theta^{k,m}_{Net/MAC} = \frac{N^{k,m}_{MACPDU} \cdot L^k_{Payload} \cdot 8}{T_{FRAME}} \tag{4}$$

The maximum efficiency for every PHY mode m is derived as:

$$\max[\eta]^m = \max_k \left\{ \frac{\Theta^{k,m}_{Net/MAC}}{\Theta_{PHY/symbol}} \right\}, \text{ with } k \in [13;2051\ byte] \tag{5}$$

The overhead introduced by the MAC protocol is strictly related to the number of MAC PDU $N^{k,m}_{MAC\ PDU}$ that fit in one frame. $N^{k,m}_{MAC\ PDU}$ depends on the set of parameters of the PMP and Mesh air-interface, which is derived in the annex **7.1.5**, respectively. The packing and fragmentation process has an impact on the protocol overhead. However, in the analytical evaluation it is assumed that these procedures are disabled.

Performance Comparison of Scheduling in Mesh Mode in Fixed Relay Scenario

In this section an analytical performance comparison between the centralized and distributed scheduling in Mesh mode (Mesh CS and Mesh DS) is carried out. As reference scenario a very simple *n*-hop topology is considered, see figure below:

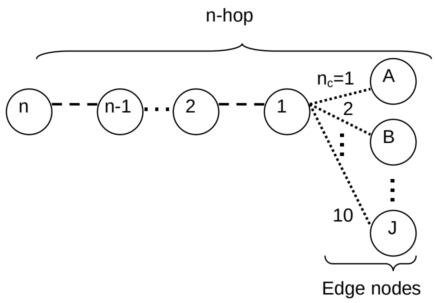


Figure 4-29: Simple tree network topology

The MAC frame is characterized by the OFDM parameters shown in Table 4-6. Within this topology one node directly serves the edge nodes via 1-hop. The node *n* acts just like a Mesh BS if the Mesh CS air-interface is considered. The edge SSs (A – J) do not announce MSH-CSCH or MSH-DSCH messages since it is assumed that assigned resources do not change.

Table 4-6: OFDM parameters

N_{FFT}	256
N_{SD}	192
BW	20 MHz
Frame duration	10 ms
K	¼
Q	7/6
T_{symbol}	13.7 µs
N_{symbol}	729

The efficiency on MAC layer is investigated for comparison purpose. In according to the overhead estimation proposed in the previous sections **7.1.5** and **7.1.5.2** the parameters fixed in Table 4-7 and Table 4-8 affect respectively the performance of the Mesh CS and the Mesh DS air interface.

Table 4-7: Mesh CS parameters

n_c	Variable
N_{CH}	1
N_{NODE}	n_c+n

Table 4-8: Mesh DS parameters

n_c	Variable
N_{SCHED}	0
$N_{REQUEST}$	0
$N_{AVAILABILITY}$	0
N_{GRANT}	n_c+(n-1)

The MAC protocol introduces low overhead as long as the edge nodes do not announce the MSH-DSCH or MSH-CSCH messages. Results in following sections show an upper bound since the Network Control sub-frame for SSs that want to gain initial access is neglected.

Furthermore, it is not necessary to transmit the MSH-CSCF message in every frame, see also the respective formula (66), on p. 237. Two versions have been considered and compared:

The MSH-CSCF message is transmitted in each frame, which we refer to as Mesh CS.

The MSH-CSCF message is transmitted only in the first frame: the overhead due to the MSH-CSCF message is not considered, and we refer to this as Mesh CS (without MSH-CSCF), i.e., $L_{MSH\text{-}CSCF} = 0$.

A solution in between both extreme scheduling mechanisms for MSH-CSCF messages might be applied in a real system. However, the analytical results for these extreme cases provide the upper bound of the performance of the Mesh CS air-interface in a real system.

P*erformance for 1-hop connections*

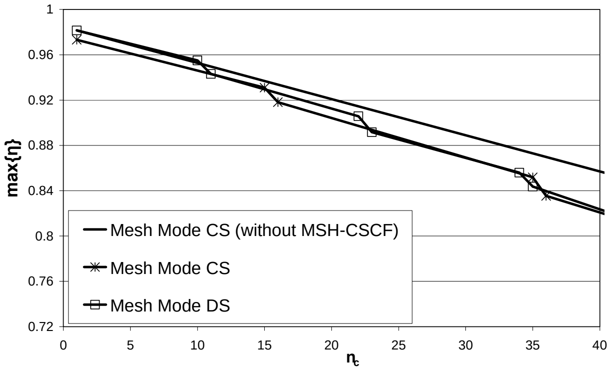


Figure 4-30: QPSK ½, 1-hop connections: comparison between Mesh CS and Mesh DS for different number of connections (nc)

In Figure 4-30 the maximum efficiency as function of the number of active connections is presented for 1-hop connections, i.e. n = 1, assuming modulation QPSK and coding rate equal to 1/2. The behaviour for different combination of channel coding and modulation is similar to that shown in Figure 4-30. Performances of Mesh CS (without MSH-CSCF) are obtained without considering the overhead introduced by the MSH-CSCF message. We can observe that the Mesh DS slightly outperforms Mesh CS for active connections below $n_c = 20$.

Plots show a non-linear behaviour due to the Control sub-frame. For instance, considering the Mesh CS and the step between 10 and 11 active connections, if the number of connections increases the MSH-CSCH message fulfills more than one transmission opportunity in the Control sub-frame. The Mesh BS reacts increasing the MSH_CTRL_LEN value and consequently the Control sub-frame duration. The Control sub-frame affects the overhead introduced by the protocol. Analogous considerations can be done for Mesh DS.

Performance for 2-hop connections

In Figure 4-31 it is shown the maximum efficiency for 2-hop connections assuming modulation QPSK and coding rate equal to 1/2. The behaviour for different combinations of channel coding and modulation is similar to that shown in Figure 4-31. We can observe that if the number of connections increases the maximum efficiency of Mesh CS air-interface decreases slower than the maximum efficiency of Mesh DS air-interface. The crossing point between curves is in correspondence to number of active connections below 10, the same behaviour is not noticeable in Figure 4-31.

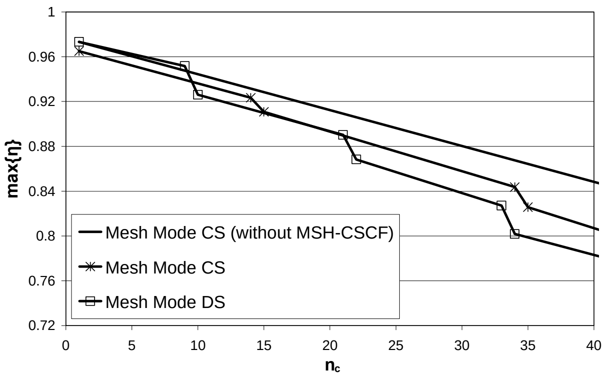


Figure 4-31: QPSK ½, 2-hop connections: comparison between Mesh CS and Mesh DS for different number of connections.

If the number of hops increases the Mesh CS outperforms the Mesh DS. The centralized scheduling mode combines the low overhead required by a centralized scheduling message with the Mesh connectivity, which provides efficiency (or equivalently low overhead) in a multi-hop scenario.

4.2.4.3.4 Spectrally Efficient Deployment Concept based on PMP and Mesh Mode

It is commonly understood that OFDM as transmission scheme is one important building block of systems beyond 3G. OFDM-based systems can be further enhanced with respect to flexibility and granularity by means of extending the orthogonal multiplexing by a multiple access component, resulting in OFDMA (Orthogonal Frequency Division Multiple Access). OFDMA allows to exploiting the frequency selective nature of the radio channel for multiple users, too. Another important feature of future systems is multi-hop to increase the coverage with high data rates and, at the same time, with low infrastructure costs. Hence, it can be expected that both components, single-hop (SH) and multi-hop (MH), will be part of B3G air-interface. A further trend indicates a higher demand with respect to the bandwidth to support the very high data rates. Target bandwidths of 100 MHz are envisaged in WINNER. This leads most probably to one available frequency band only per service provider, and hence, a cluster size of one. Moreover, because of scarce radio resources and already heavily exploited frequency bands it can be expected that there are only unpaired frequency bands with that huge bandwidth available.

Moreover, Frequency Division Duplex (FDD) does not efficiently support multi-hop communication if a continuous full-duplex mode is used. Incorporating a TDMA-component in the FDD system to support multi-hop destroys many of the advantages of FDD. Consequently, the most promising duplex scheme for the envisaged system is TDD (Time Division Duplex) not only because of the reasons mentioned before but also based on the advantages of reciprocity of the channel that can be exploited for intelligent and smart antennas and MIMO (Multiple Input Multiple Output).

MH communication can be realized in time, frequency or a combination of these (see [87]). In case of an extension for HIPERLAN/2 it is proposed an approach which incorporates a frame inside a frame for scheduling along a MH path to the final destination [88]. This approach uses the whole bandwidth for all traffic between the MT resp. subscriber station (SS), independent whether the users are reachable via one hop (SH) or via several hops (MH) with respect to an Access Point (AP) / base station (BS) or relay station (RS). This is one traditional approach in MH communication, and the resulting spectrum usage is exemplary shown in Figure 4-32.

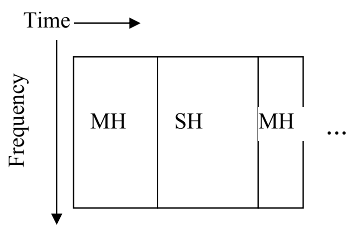


Figure 4-32. Example for multi-hop (MH) and single-hop (SH) communication sharing one frequency band

However, if the MH and SH is realized in the same frequency band, like in the proposals for HIPERLAN/2, these approaches result in large overhead, a strong dependency between the MH and SH resources, and in bad scalability, i.e. the number of supported hops is limited to a small number only, as this has been shown in [88].

In IEEE 802.16a separation of the MH network and the last hop (SH) is foreseen. This results in a hybrid approach. For the multi-hop connection a different frequency band and different technology is used than for the last hop. E.g., for the MH communication the Mesh-mode of Fixed Broadband Wireless Access (FBWA) IEEE 802.16a is proposed, as described beforehand in this document. For the last hop IEEE 802.11 or ETSI-BRAN/HIPERLAN/2 is envisaged. An example for this kind of allocation is shown in Figure 4-33.

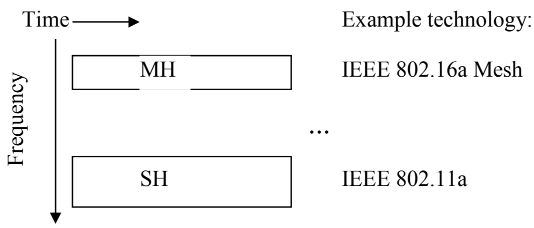


Figure 4-33. Example for multi-hop (MH) and single-hop (SH) communication in separated spectrum

That means, up to the next to the last hop IEEE 802.16a Mesh mode is utilized, whereas on the last hop a WLAN technology is deployed. The AP/BS or relay, serves as control station to share the medium between the stations on the last hop. Moreover, the operation of the MH transfer by means of IEEE 802.16a and the communication on the last hop via, e.g., IEEE802.11, IEEE802.16a PMP mode, or ETSI-HIPERLAN/2 is realized in different frequency bands as shown in Figure 4-9. Hence, the distributed wireless MH communication and the centralized SH communication is separated in the spectrum, requiring two separated frequency bands to realize this kind of deployment concept. However, the logical separation of MH and SH operation takes into account the differences in requirements and characteristics of the different modes, as explained in the next section.

Differences in Multi-hop and Single-hop Communication

The multi-hop traffic is realized between one access point (AP), respectively base station (BS), which is connected via wire to the fixed backbone network, and fixed relay stations, so called Fixed Multi-hop Nodes (FMHNs). The last-hop traffic takes place between the AP/BS or FMHN and mobile nodes (MNs), respectively subscriber stations (SSs). A typical deployment concept for the multi-hop air-interface is depicted in Figure 4-34.

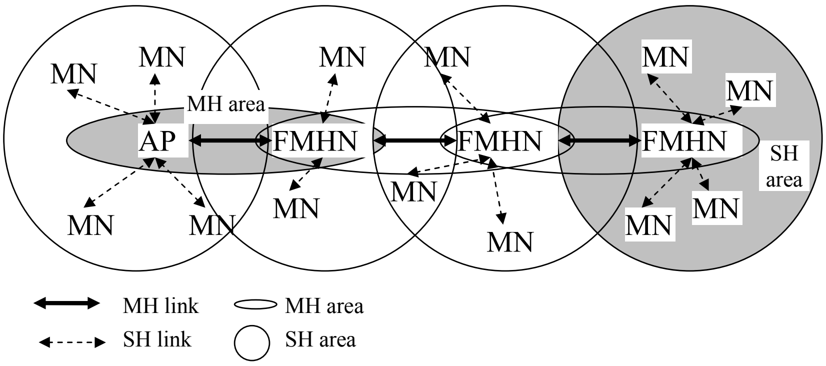


Figure 4-34. Deployment concept for a multi-hop topology

MH and SH communication is separated in different links between different devices. MH links exist between APs and FMHNs. SH links are established between APs respectively FMHNs and the SSs. Because different physical devices are part of the communication relationships the MH and SH links are separated in space. The MH area in Figure 4-10 comprises the AP and FMHNs that realize the MH connections as a kind of wireless high-speed backbone. SH areas exist around each AP and FMHN, respectively, to serve the MN on the last hop. There are quite different requirements for MH and SH communication. To get a clear insight into the differences in the following Table 4-9 the typical characteristic of MH and SH communication is listed. It is clearly visible that there exist basic differences, which most probably leads to different protocol design and different physical layers. However, it can be expected (that also becomes evident in the standardization efforts during the last years) that OFDMA will be the common denominator.

Table 4-9. Differences between multi-hop and single-hop communication

Characteristic	Multi-hop	Single-Hop
General		
UL / DL characteristic	Symmetric access	Asymmetric access
Traffic dynamic	Slowly varying	Very dynamic
Traffic characteristic	Multiplexed traffic streams	Depending on applications
Data rate	Very high	Low to high
Mobility	Static (no mobility)	Low/no to high velocities
Physical Layer		
Propagation conditions	LOS	LOS/NLOS
Delay Spread	Small (<0.5us)	Medium/large (1-5us)
Antennas	Directional	AP/FMHN: >4, MN: 1-4

Data Link Control		
Control of medium access	Distributed or centralized	Centralized
Power saving	Low relevance (power supply most prob. available)	Important (battery limited)
Scheduling	Distributed or centralized	Centralized
Resource Allocation	Slow to medium variations	Fast and dynamic allocation
Switching	Circuit-switched /packet-switched	Packet-switched

The access in UL and DL on a MH link is symmetric, i.e. there is no basic difference in the allocation of resources in the one or other direction. Even more, the DL and UL cannot really be distinguished, except that the traffic in the DL is directed to the MNs and in the UL to the AP, resp. Internet. To avoid misunderstandings it is worth to mention that there is a difference in access and traffic a/symmetry and that a symmetric access does not imply that the traffic cannot be asymmetric in a MH link. Different to that in the SH area there is one AP/FMHN and several receivers in the DL, while on the UL there are many transmitters and one receiver only. This should be taken into account, e.g., in the design of the physical layer but also on the data link control layer.

Another aspect is the traffic characteristic on the different links. A flexible and fast reaction on resource demands (on-demand resource allocation) on the last hop for many users (many SSs) with respective high protocol overhead should be supported in the SH area. Different to that, a slower resource allocation for multi-hop connection with a minimum overhead is envisaged for the MH area. This approach results in balancing the tradeoff between protocol overhead and fast resource allocation between a high number of connections in the SH area and the low overhead and small number of relays (FMHNs) in the MH area. Another aspect that justifies the lower dynamic on the MH links can be explained as follows: The traffic of the high number of MNs, which have to be served very flexible and fast, will be multiplexed at the FMHNs, resp. AP. The traffic multiplexing results in a multiplexing gain, and hence, in a better utilization of the available bandwidth, and at the same time reduces the dynamic in traffic demands. The latter effect results in a lower effective bit rate.

Closely related to the multiplexing is the required data rate on the SH and MH links. Since all traffic of the SH area has to be transferred to/from the AP/Internet, except for peer-to-peer traffic, the aggregated data rate of the SH link has to be managed by the MH link, resulting in much higher data rate demands on the MH link. But independent from this demand, the mobility characteristic supports a much higher data rate on the static MH link, which fits quite well to the required data rate.

Moreover, the MH link is most probably line-of-sight (LOS) with a very small Doppler spread and Delay spread opposed to the SH link with many times NLoS and a high Doppler and Delay spread. The difference in the physical channel characteristic on the SH and MH link calls for a different physical layer design. This design criteria becomes even more obvious from the antenna design. On the MH link directional antennas can be easily introduced, whereas on the SH link smart-antennas needs to be deployed that supports either MIMO, diversity techniques, or spatial-multiplexing / beam-forming.

A major difference between MH and SH communication can be expected in the design of the data link layer (DLC). One reason for this is the medium access control, which is most probably distributed on a MH link since there exist no central instance that can directly control all individual links on a multi-hop connection, and consequently, the access on the medium has to be managed between the instances within direct radio distance in a kind of self-organized and distributed manner. Different to that a centralized access scheme is the best on the SH link to manage the dynamic behaviour of several links at the same time within a limited range, which is denoted by the SH area. As most probably most of the traffic is exchanged between the AP/FMHN and the numerous MNs (to/from AP/FMHN from/to MNs), it is the logical consequence that the AP/FMHN takes over the central control.

Closely related to the medium access control is the scheduling, and it can be expected that distributed scheduling might be the choice for the MH links and centralized scheduling for the SH links. However, as long as the scheduling information can be disseminated fast enough to follow the changing traffic demands, a centralized scheduling approach might be appropriate even for MH connections.

The variability in the traffic demands has also an impact on the selected type of switching. If the same traffic demand exists between, e.g., two FMHNs, it might be advantageous to establish a circuit-switched connection for quite a long time. This allows other FMHNs in the neighbourhood to predict the interference and optimally exploit the resources based on the optimal selected modulation and coding scheme. However, this becomes less efficient in the SH link, since the data requirements of individual MNs changes much faster. Packet switching on a frame-by-frame basis should be the choice to take into account the changing traffic demands and, which is even more important, the changing channel conditions. They most probably change much quicker on a SH link than on a MH link, which also gives good reasons to select a different switching type for SH and MH communication.

Summarizing the characteristics of SH and MH links it becomes clear that these quite different connection types should be individually treated. In the following section a proposal is introduced that takes into account these differences in the design of the deployment concept.

OFDMA-based Multi-hop and Single-Hop Approach

In this section the individual realization of SH communication and MH communication over fixed relay stations (FMHNs) will be explained. OFDMA, as most promising access scheme, allows to separating a large frequency band in smaller fractions to serve different stations. As logical consequence from the differences of SH and MH connections, OFDMA is used to separate/split the available frequency band into two basic parts: one part is used for SH communication and the other is used for MH communication. It is foreseen that adjacent sub-carriers are assigned to the MH and SH traffic, resulting in two adjacent sub-bands, one for MH, the other for SH. That means, the air-interface of the target system uses one whole frequency band, e.g. of 100 MHz, and separates this whole frequency band in two parts made up of disjoint sub-carriers. An example for the allocation of Nc sub-carriers of an OFDMA-based air-interface is depicted in Figure 4-35.

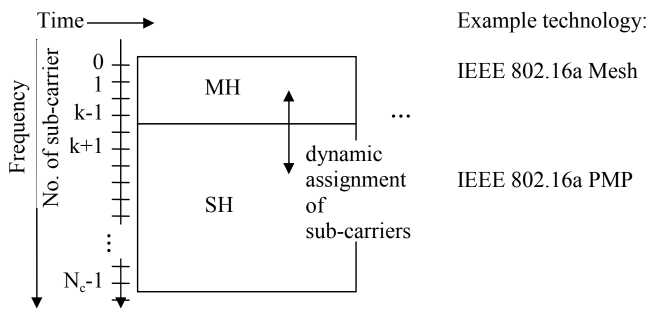


Figure 4-35. Dynamic assignment of sub-carriers to multi-hop (MH) and single-hop (SH) communication in an OFDMA-based multi-hop system (MS-OFDMA)

The separation in two sub-bands is dynamically performed in a flexible way by utilizing the features of OFDMA. OFDMA allows to assign different sub-carriers to different users, respectively links. It is proposed to perform a block-allocation of sub-carriers to the two sub-bands on demand, i.e., the sub-carriers in the upper frequency band are allocated to the MH sub-band, whereas the remaining sub-carriers are allocated for the lower sub-band that is used for the last-hop traffic (SH). The number of sub-carriers assigned to the MH and SH traffic can be changed dynamically on demand. The resulting approach that combines MH and SH in one coherent frequency band is called in the following MS-OFDMA (Multi-hop Single-Hop OFDMA).

Depending on the required resources for MH and SH (last-hop) traffic, the sub-band separation will be changed. E.g., if there is heavy local traffic between stations that are served by one FMHN, there is only a low demand of bandwidth to transport data between the Internet, resp. AP, and MN via FMHN. Hence, the MH sub-band is reduced to a very small number of sub-carriers. However, if more bandwidth is needed for MH traffic some carriers used so far for SH traffic will be allocated to the MH band in the future. For instance, if every MN has a connection with the Internet, the MH sub-band should be increased to support the huge traffic demands that are relayed via the FMHNs.

The separation of DL / UL between the AP/FMHN and MNs and the AP and FMHNs is realized by means of TDD. However, FDD operation for the SH links is supported by this concept, too, and a hybrid FDD approach can be used for the MH connections. Nevertheless, because of complexity reasons in the hardware is it not recommended here.

In the next Figure 4-36 the dynamic assignment of the sub-carriers to the MH and SH traffic is exemplary shown for a deployment concept with two FMHNs.

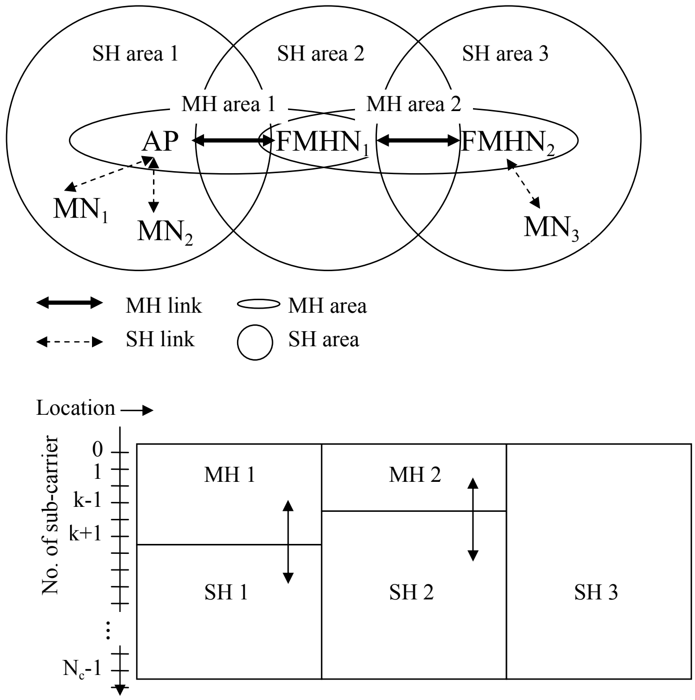


Figure 4-36. Deployment Concept for a multi-hop topology for the OFDMA-based multi-hop scheme MS-OFDMA

In the scenario connections with up to three hops are supported. The first two hops are realized in the MH sub-band between the AP and the two FMHNs, $FMHN_1$ and $FMHN_2$, and the third hop is realized in the SH band no. 3 (SH 3) between the second $FMHN_2$ and MN_3. In Figure 4-36 under the resulting topology the respective bandwidth allocation is depicted, too. The MH traffic of the MH area 1 (MH1), which corresponds to the bidirectional traffic between the AP and FMHN1, is sharing the sub-carriers with the SH traffic in SH area 1 (SH1), which is served by the AP. The MH and SH traffic of the MH area 2 (MH2) and SH area 2 (SH2), resp., is also sharing the sub-carriers in a dynamic way. The traffic within the SH area 3 can be exclusively utilizing all sub-carriers (SH3), since no further MH connection between FMHNs exists.

Bandwidth Dimensioning in MS-OFDMA

Since the traffic on the downlink (DL) from the AP to the MNs is distributed to the SH areas and on the uplink (UL) from the MNs to the AP it is aggregated, the bandwidth requirements for the MH links is increasing towards the AP, respectively to the Internet. This is taken into account by means of a higher number of sub-carriers allocated to the MH links closer to the AP, as shown in the example in Figure 4-36. However, a different allocation of sub-carriers is possible, e.g. less sub-carriers for the sub-band MH 1 than MH 2 in case of heavy local traffic or traffic between SH areas that are connected over FMHNs without crossing the AP. Since it is further expected that the traffic on the MH link is realized by line-of-sight (LOS) conditions with high antenna gains, the data rate is much higher than on the last hop between AP/FMHN and MN. Thus, the number of carriers allocated to the MH links can be lower than the number needed for SH links, if we assume that all traffic goes to/ comes from the AP, respectively Internet. Furthermore, the MNs served over the last relay, i.e. FMHN 2, will see the highest number of hops towards / from the Internet. The largest allocated bandwidth for these MNs, i.e. SH area 3, partly compensates this disadvantage, and allows to reducing the delay experienced on the respective links.

Alternatively, the size of the cells can be appropriately chosen. The last cell (MH area 3) will become a larger cell (covers more MNs) than the previous cells served by FMHN 1. This kind of cell dimensioning results in a constant data rate per user all over the whole deployment scenario, which is one envisaged target of the deployment concept of WINNER.

With the novel scheme a flexible allocation of resources on an end-to-end connection becomes possible. The OFDMA-based MH concept introduces a logical separation of MH traffic and SH traffic, which can be served via different protocols above a common physical layer and a shared common frequency band. Since the MH communication puts different demands on the protocol design compared to the SH communication, most efficient protocols can be developed and deployed almost independently for the respective problem field. At the same time there is no need for separate spectrum, like in conventional solutions. Only one frequency band is required. The proposed approach supports the exchange of bandwidth between MH and SH traffic with the help of OFDMA in a flexible manner. Furthermore, the separation between MH and SH bands do not need large guard bands like in conventional Frequency Division Multiple Access (FDMA) approaches, since OFDMA allows closer separation due to its orthogonal sub-carriers in the frequency domain.

Summarizing the advantages of MS-OFDMA, the differences as listed in Table 4-9 above will be taken into consideration, though both types of communication share the same bandwidth in a dynamic way on a demand-basis. MS-OFDMA combines the gains of a separation of MH and SH traffic due to their different characteristics with all benefits of OFDMA in a very efficient way.

Organizing UL and DL Transmission in MS-OFDMA

So far the distinguishing factor for the approach was MH and SH. The direction of traffic flow, i.e., downlink (DL) or uplink (UL), has not been considered so far. In this section the direction of traffic flow will be introduced and solutions will be presented for the MS-OFDMA approach. Specifically, the special interference constellations due to variable switching-points that results from a TDD approach in combination with an OFDMA approach (the logical bandwidth split in MH- and SH-sub-band) have to be taken into account. Another important aspect for the selection of a suitable solution are the applied technologies to combat interference, e.g., the performance of the filters to suppress adjacent channel interference, the synchronism of the sub-bands to avoid inter-carrier interference (ICI), and the used antennas. Especially, the selected antenna technologies allow different degrees of resource exploitation. In a first step it is assumed that only omnidirectional antennas are used. In this case, the simultaneous transmission in the SH- and MH-sub-band might result in strong sub-band interference. In the next step introducing directional antennas for the MH transmission, this kind of adjacent channel interference will be considerably reduced. In the extreme case of beam-forming, links can be spatially separated and the same frequency resource can be used at the same time on several links. Depending on the hardware effort to increase the spectral efficiency as explained before, the different solutions are presented in the following sub-sections.

MS-OFDMA with Omni-directional Antennas

In the following figures the following notation is used. The sub-band used for MH communication between AP and FMHN, respectively between the FMHNs, is denoted by “MH x”. The number “x” denotes the link corresponding to the topology in Figure 4-36. The lower sub-band is used for the SH transmission, i.e., to/from a MN, and is denoted by “SH x” with respect to the SH area. The station that is active occupies the respective sub-band and at the same time defines the direction of the transmission, i.e., DL or UL. E.g., if the AP is active on the SH sub-band, it transmits to the MNs in the DL, whereas the transmission in the same band by a MN results in UL transmission. Different MNs in a specific SH area “x” are denoted by “MN area x”. The time is quantized in frames of a fixed length. However, the individual length within a frame between UL and DL, i.e. the switching point, as well as the split between SH and MH band can be changed. If this can be done independent of the other transmissions this will be indicated in the graphics with an arrow. If the boundary should be fixed no arrow is introduced.

In the following Figure 4-37 one resource allocation example for the basic MS-OFDMA scheme for the topology depicted in Figure 4-36 is shown. The AP starts in this example with a DL transmission to the MNs in the SH-area 1 (SH 1), and simultaneously with a MH transmission to the FMHN 1 in the upper sub-band (MH 1). Since the AP as transmitting station can individually change the allocation of sub-carriers to the upper and lower sub-band, a vertical arrow is introduced to indicate this flexibility. The switching point between DL and UL in SH area 1 can be arbitrarily assigned, too, and a horizontal arrow indicates this flexibility. The following UL transmissions of MNs of the SH area 1 is accompanied by the MH transmission in the UL from FMHN 1 to the AP (MH 1). An exchange of sub-carriers between the MH and SH transmission has to be coordinated between the MNs and the FMHN, i.e., stations in two different areas. Hence, a vertical arrow is missing in this case. However, the switching point between the UL and DL transmission of the FMHN 1 can be individually changed. Thus, a horizontal arrow is introduced on top of the MH sub-band. The next transmission cycle takes part between the second SH area 2 and the second MH link in a similar way. During the last frame the UL of the second MH link and the third SH area has to be served. Since there is no further MH link the MNs can use the upper frequency band for SH communication, too. In total the exchange with all stations requires a time equals three frames. The challenge in this approach is the simultaneous reception of two different links, e.g., the AP has to receive on the UL the MNs in SH 1 and on the UL on the first MH link MH 1. However, this corresponds to the basic challenge of OFDMA in the UL and, hence, can be realized without any problems (if OFDMA in the UL is supported).

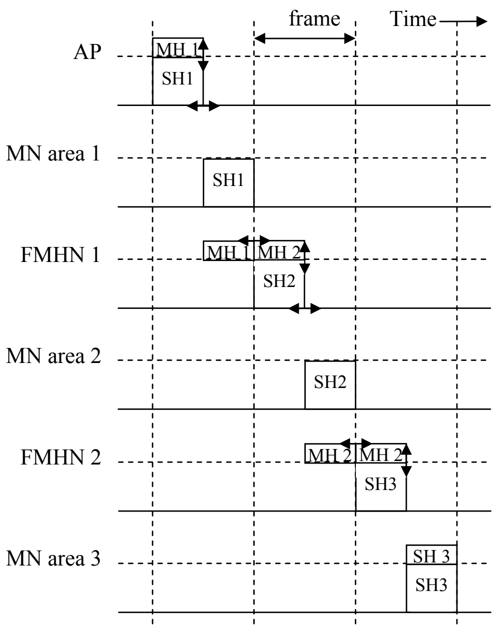


Figure 4-37. Basic MS-OFDMA with omni-directional antennas

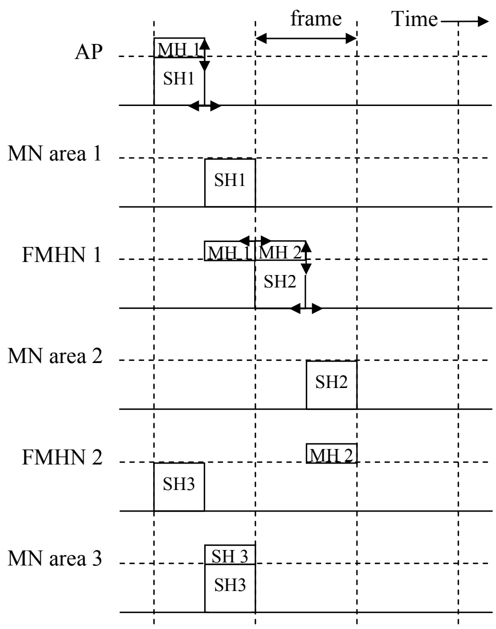


Figure 4-38. Frequency re-use in the basic MS-OFDMA with omni-directional antennas

In the next Figure 4-38 it is assumed that simultaneous transmissions in SH area 1 and SH area 3 can be realized due to frequency re-use. The whole transmission time is reduced by one frame to two frames. The advantages with respect to flexibility and challenges are the same as for the basic MS-OFDMA approach.

In the next Figure 4-39 a typical cellular-like deployment approach is used, where the DL and UL transmission of each SH area is simultaneously exploited. However, for the MH communication at least another frame is needed for exchange of data on the MH connections. What becomes also obvious is the waste of capacity due to the restriction in the second frame of the MH communication on the MH sub-band only. This kind of restriction is not recommended for this deployment concept. Instead, the communication over the MH links should utilize the whole bandwidth. In turn the transmission time on each link can be reduced, resulting in exactly two frames for this kind of solutions. The increase in bandwidth and reduction in transmission time is marked with the arrows in Figure 4-39. Since the DL and UL in neighbouring SH areas is exploited the typical challenge of variable switching points of TDD exists. However, different solutions are known to mitigate the critical interference constellations. A more detailed discussion about variable switching points in TDD and respective solutions can be found in [Ref: Del. WP2 “Duplexing Schemes”]. Moreover, this approach allows a long and predictable inactive time of at least one frame for the MNs for power saving and scanning of other frequencies for, e.g., interference measurements or handover preparation.

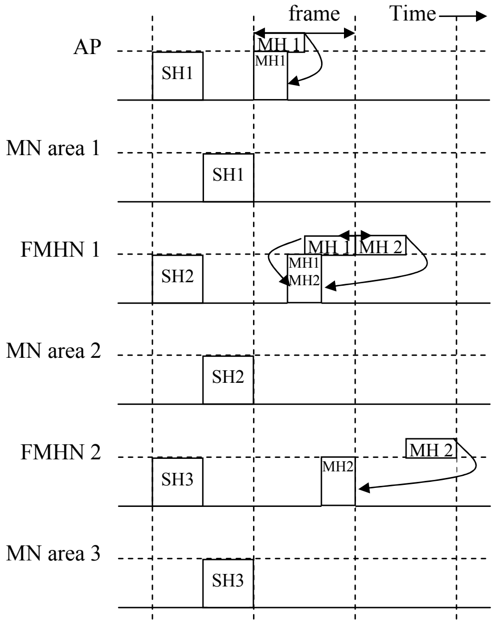


Figure 4-39. Cellular-like MS-OFDMA with omni-directional antennas

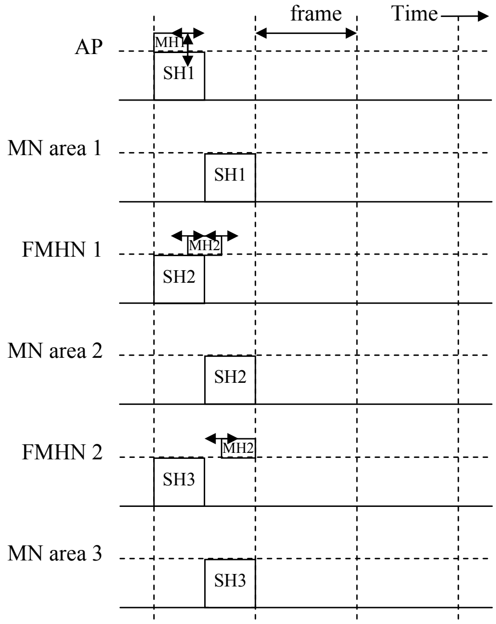


Figure 4-40. Cellular-like MS-OFDMA with omni-directional antennas and strong sub-band separation

In the next Figure 4-40 the restriction with respect to transmission and subsequent reception in adjacent sub-bands is relaxed. E.g., the FMHN 1 can transmit on the DL over SH 2 to the MNs and receive at the same time from the AP data over MH 1. This increases the spectrum utilization considerable and can reduce the whole exchange cycle to one frame only, if we slightly reduce the transmission time on the MH links. However, this kind of deployment concept requires strong separation of adjacent sub-bands during transmission and reception. The demand can be relaxed if some of the sub-carriers between the sub-bands are left free. Still, the hardware capabilities in the AP and FMHNs will be the restricting factor for this approach. On the other side this approach allows an independent concept deployment for MH and SH communication and supports the adoption of the most suitable solution for the respective communication type.

Here the typical tradeoff between spectral efficiency and flexibility on the one side and hardware complexity on the other becomes obvious. Nevertheless, the introduction of antennas might relax this tradeoff and definitely can reduce the requirements with respect to suppression of adjacent-channel interference.

Directional Antennas for Multi-hop Connection

In the following examples directional antennas are deployed for the MH connections. This does not change the assumptions and transmission order in the basic MS-OFDMA approach, but can reduce the transmission cycle when utilizing the capabilities of frequency re-use, as shown in Figure 4-41 compared to Figure 4-38 above. In addition, the DL and UL for the SH area 2 can be extended to the whole bandwidth, since no more MH communication is needed during that time in the second frame. The UL and DL transmission of FMHN 1 can be initiated in parallel (MH 1/2) due to directional antennas and a clear separation of the respective receivers, the AP and the FMHN 2.

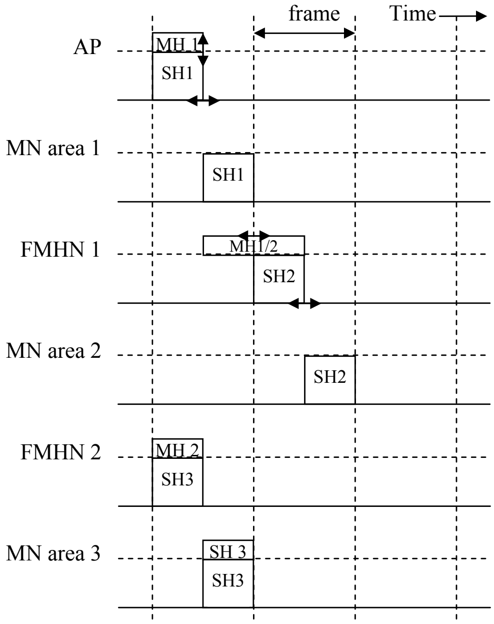


Figure 4-41. Basic MS-OFDMA with frequency re-use and directional antennas

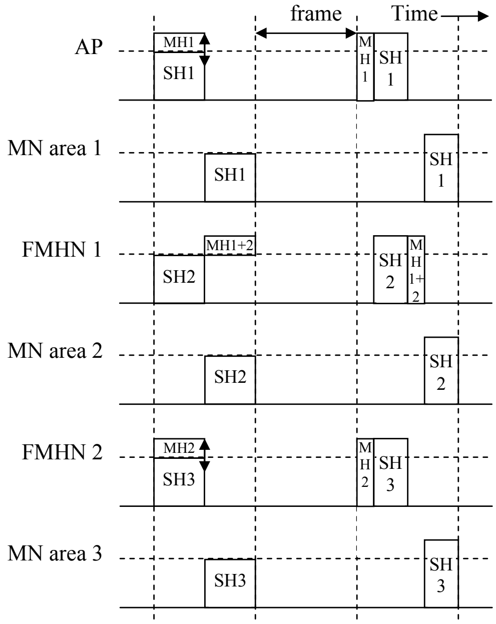


Figure 4-42. Cellular-like MS-OFDMA vs. TDD-OFDMA with directional antennas

Better separation between the different links can be reached in the cellular-like deployment, too see Figure 4-42. In contrast to Figure 4-40 the simultaneous transmission and reception in adjacent sub-bands is now only necessary for FMHN 2.

One alternative approach to MS-OFDMA is the separation of the MH and SH links in the time domain, which is here referred to TDD-OFDMA. Both variants are depicted in Figure 4-42. Using TDD-OFDMA no simultaneous transmission and reception becomes necessary at FMHN 2, reducing the hardware complexity. MS-OFDMA on the other side separates the SH and MH communication and, hence, decouples the resource allocation in the frequency domain. This allows to adopt a different frame structure to the MH and SH communication, which is not possible in case of separation in the time domain as in TDD-OFDMA, where the MH and SH parts make up and influence the structure of the frame.

Impact of Neighbour Deployments

So far the impact of other MS-OFDMA deployments has not been taken into account. This is important for smooth extension of the deployment area towards wide area coverage as evolutionary strategy. For the MH communication highly directional antennas will be used and it can be expected that the MH links are spatially well separated and do not interfere with each other. Nevertheless, the impact of directional interference will be considered if it comes to the topological design, i.e. the placements of relay stations, which will be discussed in a section. But for the SH communication the optimization is more complicated.

Based on the cellular-like MS-OFDMA concept typical interference situations of TDD systems will occur on the SH links. Solutions like MNs, which are closely positioned at the AP and have good link conditions, will be scheduled at times close to the switching point can be applied. Furthermore, different scrambling codes can be assigned to neighbouring cells that are served by an AP or an FMHN, etc.

One major topic that becomes more important for future systems is the usage of antenna techniques. Besides the commonly used sectorized antennas, beam-forming and MIMO techniques should be incorporated in an efficient way. Especially, for inter-cell interference suppression beam-forming can be advantageously be adopted to the DL transmission towards MNs. In the UL direction also beam-forming should be used again at the BS at the receiver, since it can be expected that the MNs are equipped only with less then or equal to four antennas, which makes a good beam-forming less efficient. Beam-forming enhances the overall system capacity due to simultaneous transmission to/from several MNs, but it does not considerably increase the data rate on an individual link like MIMO does. For MIMO multi-path diversity is needed for uncorrelated channels seen from the different antenna elements. This is the case, e.g., for NLOS conditions. At the same time it is known from investigation on MIMO that the C/I should be sufficiently high to achieve large capacity gains.

Topological Design Criteria for MS-OFDMA

The deployment concept of MS-OFDMA comprises also the topology. Specifically, the strategic placement of relay stations in the different scenarios will have a strong impact on the overall system performance with respect to capacity and costs.

Manhattan Scenario

For typical Manhattan like scenarios the following spectral efficient deployment concept is envisaged. For the spectrally efficient connection of relay stations (AP/FMHNs) a displaced positioning of relay stations is proposed. A typical topology along a street is shown in the following figure.

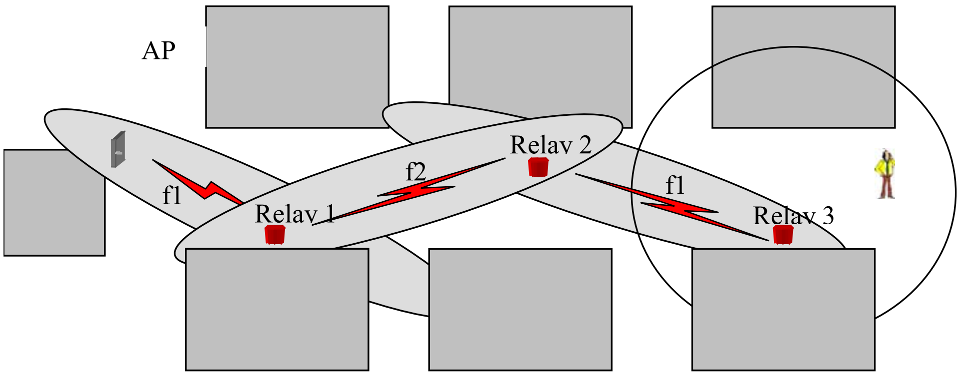


Figure 4-43. Spectrally efficient topology with relay stations along a street

In this topological structure the relays are positioned displaced to the logical line the relay stations build along the street. Subsequent relays are placed on the opposite side of the street. Coupled with the geographical placement of the relays is the deployment of highly directional antennas to connect the AP with the relays. Based on this topology and the directional antennas only two frequency bands (carrier frequencies f_1 and f_2) are needed to support a simultaneous transmission on different links within a frequency re-use distance of two hops. Thus, with only two frequency bands an arbitrary multi-hop connection with simultaneous operation of all individual links can be realized. One typical mode of operation in this topology might be the "pumped" forwarding as shown in the following Figure 4-44.

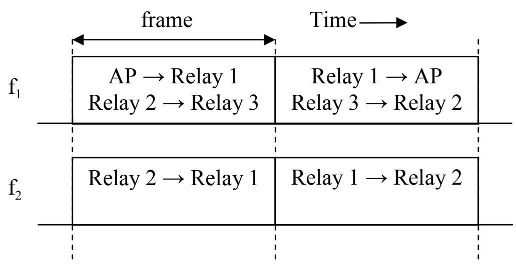


Figure 4-44. Schedule for two frequencies in a spectrally-efficient MH topology

The AP and next but one relay, relay 2, will transmit at the same time on the same frequency. The AP will operate on f_1 and relay 2 will simultaneously serve relay 1 on f_2 and relay 3 on f_1. In the next period the remaining relays, relay 1 and relay 3 will transmit concurrently. Relay 1 will serve the AP on f_1 and relay 2 on f_2, while relay 3 will transfer data to relay 2 on f_1.

In a further step the SH area around the relays can be served with directional antennas, too. This is depicted in the next Figure 4-45.

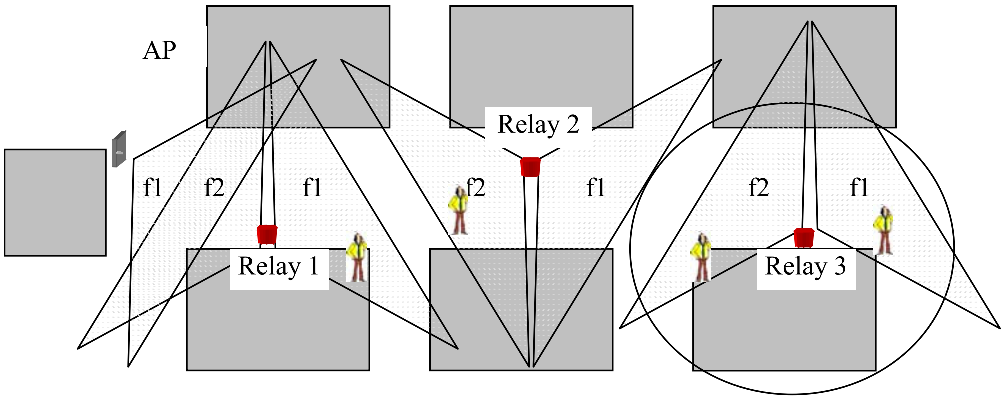


Figure 4-45. Deployment of directional antennas to serve the SH area around the relay stations

The main directions of these antennas are the same as the one of the directional antennas of the relays for the MH connections. And similar to the MH connection only two frequencies are foreseen for the simultaneous operation and wide-area coverage. Only with two frequencies all MNs can be served in the street canyon. In comparison with the previous Figure 4-43 the DL and UL for the SH communication will use the same frequency as the MH link between the AP/relays. This is assumed as possible solution because of the used directional antennas. However, if the co-channel interference becomes too large the MH and SH frequency bands can be separated, as explained and proposed for the MS-OFDMA approach, i.e., the MH connection uses the upper OFDMA sub-band, whereas the SH connections will use the lower sub-band.

The main advantage of this topology is the minimum number of frequency bands to serve a multi-hop connection of arbitrary length. This represents an extremely spectrally-efficient solution for multi-hop communication. The co-channel interference is minimized in spite of the low frequency re-use distance due to the intelligent placement of relays and the deployment of directional antennas. In addition, it is proposed a bidirectional operation for the radio connections. This allows to using TDD scheme for the optimized support of asymmetric and time-variable traffic flows. The service of MNs around the relays can be realized with the same two frequencies used for MH communication for a complete coverage along a street. Despite of the low number of frequencies, co-channel interference is not a problem because of the directional antennas for MH and SH communication. This kind of deployment concept for MH and SH links represents an efficient adoption of Space Division Multiple Access (SDMA) technique.

4.2.4.4 Mobile Relay-Based issues and deployment concepts

4.2.4.4.1 Introduction

Characteristics of future mobile/wireless radio networks will be quite different to those of current networks. Compared to a “in general” uniform spatial distribution of similar types of traffic (voice-data) in a GSM cell, a much more non-uniform traffic will be witnessed in future deployments in terms of spatial distribution within the cell of those needs - traffic requirements and even higher demands for a much broader bandwidth of bit rates stretching from low to high bit rates for higher mobility.

At the same time, the more we move to higher frequencies, the more the range of the cell shrinks to smaller values compared to e.g. GSM cells. This means that in order to cover the same geographical area, more BSs/APs need to be deployed, which leads inevitably to much higher deployment costs for operators.

Apart from those “new” needs, future network deployments will need to solve more adequately problems that already exist in current conventional network deployments and are related to the issue of coverage; either extend/stretch the coverage of cells further away or provide better coverage within the coverage area of a cell in terms of e.g. better QoS.

These requirements/constraints will need to de addressed either by new technologies e.g. new air interface with conventional wireless/cellular approaches and systems or with new deployment concepts which will introduce new network elements.

One simple and cheap solution has been already proposed within 3GPP for 3G networks. This is introducing a new network element called “repeater”. Repeaters are basically analogue, amplify-and-forward network elements which receive, amplify and transmit a signal without performing any baseband processing to the signal. [95][96][97] For 2G networks, in Europe they are used to provide better coverage within the cell (e.g. in black spots/shadowed areas), whereas in the USA they are used to extend the coverage of cell borders due to the large and usually scarcely populated areas that need to be covered. However, as we will see further, there are some implications associated with them, such as amplification of both the signal and the noise, mobile positioning etc.

An extension to the approach of repeaters is that of fixed relays. Fixed relays are network elements which can be as simple as repeaters, but also as complex as an AP. The end goal of fixed relays is basically the same as that of repeaters - better coverage – extension of coverage. However, they target more complicated and more demanding scenarios and as such, they tend to be more expensive with regards to cost compared to repeaters.

The main idea behind fixed relays is that having incorporated more functionalities compared to a repeater, they can, “in a way”, replace an AP. So, they can increase capacity/coverage more efficiently than a repeater, but with much less cost compared to an AP, which is actually their main advantage. Thus, as long as in future networks we need to provide more APs/BSs, then fixed relays (effectively a “cut-down” version of an AP) could be a cheap and economically viable solution to cover our needs adequately.

4.2.4.4.2 Mobile/Movable Relays

Mobile Relays

As stated above, new needs such as coverage/capacity will need to be addressed by new network elements, such as fixed relays. However, fixed relays will cover more deterministic/“expected”/“uniform” needs. For instance, a fixed relay might be placed in an area with poor reception, because a building that was newly built prevents proper reception.

However, there are cases that fixed relays might not be able to provide adequate solutions especially under the view of cost. Additionally, there might be cases that fixed relays might not be able at first place to solve some needs.

Examples of saturated capacity due to high concentration of terminals and calls, lead to a need to provide that additional capacity, but only when this is required.

Let us assume the case of an accident (1) or a terrorist act (2) or football stadiums (3). In those circumstances the network should cope with

- An exceptionally high volume of voice or data calls (people informing friends/relatives of the news or sending SMS/MMS/video clips)→(1)
- Need to guarantee resources (high bit rates) for emergency services e.g. fire brigade, ambulances, and police for reasons of e.g. telemedicine or cooperation of emergency services. →(2)
- High capacity but only for few days per week.→(3)

In these cases, fixed topologies might not be able to cope with increased needs and, thus, become saturated. On the other hand, the need for additional capacity/coverage could be handled by new network elements which could e.g. be switched on demand and also be very cheap to deploy, thus not cause a “financial” problem to operators. Those could be fitted to any “carriers” like cars (ambulances, police cars, buses etc). What is more, those elements would be in the “target” area only when we need them, because inevitably police cars or ambulances will be there in the case of an accident.

In general, events like the ones we described above, are mainly characterised by

- Unpredictability. Cannot predict when an accident will happen
- Non-uniformity. Football match occurs only 2-3 days per week. Capacity is not required in a stadium all the time.

Of course new schemes are proposed to cope with some of these types of events. Proposals include reservation of resources for emergency services, similar, for instance to the MESA project that ETSI is involved. [98] However, these schemes follow the approach of restricting normal users from using the network (e.g. voice calls) and not focusing on trying to accommodate them.

Additionally, there might also be cases that purely for reasons related to cost, it would be more beneficial to deploy a number of mobile relays on buses to provide coverage on streets, rather than deploying fixed relays.

From the above it is obvious that in cases like these, fixed relays either could not cope with some events or would not be economically justifiable to be deployed e.g. football grounds. Thus, in those cases it might be beneficial to deploy mobile relays as a complementary option for the network.

Figure 4-46 shows a generic representation of a mobile relay scenario. It depicts an AP, a user terminal (T – cylinder) which is moving and a mobile relay (R – rectangular) which is used to provide coverage either to that UT or to a specific area (bold shaded area). What is also depicted is the direct and the relayed paths.

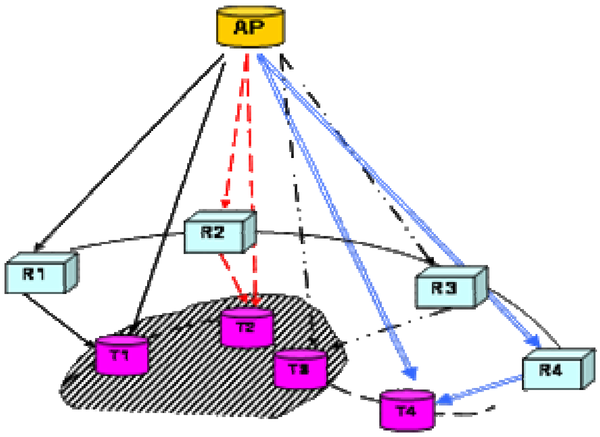


Figure 4-46: Generic mobility scenario for mobile relays

Note: T→Terminal, R→Relay, Ri, Ti → Positions in time i=1,2,3,4.

Thus, the main needs that mobile relays could solve (not only on the technical level) could be the following:

- To provide a cheaper solution compared either to conventional topologies or fixed relay-based topologies. This solution will be cheaper not only in terms of deployment but also in terms of maintenance. A typical example is that of MRs being fitted on buses passing by a park. Thus, with a proper scheduling we can provide better coverage of the park without the need for deploying fixed relays or additional APs.
- To provide coverage in unexpected events where high concentration of UTs occurs e.g. road accident or terrorist act. Those relays could be fitted on police cars, ambulances etc and be used “on the spot”.
- To provide coverage for events which occur infrequently but can be “known” in advance e.g. football match. It might not be economically wise to deploy an AP next to a football stadium. In these cases “on demand” MRs could be used.
- To provide coverage where fixed relays “cannot go”. Imagine a scenario where we need to provide coverage in terminals which are on top of a riverboat or in a train. Thus, we rely on fixed relays which, on such a carrier, become inevitably mobile relays, instead of relying to, probably, poor reception from fixed relays or APs close to the banks of the river.
- To reduce levels of power of transmission. By deploying mobile relays (in the same way as in fixed relays) we can afford to reduce the total transmitted BS/AP power. The direct gain could possibly be increased capacity in a CDMA-based interference-limited system.
- To reduce the impact in the environment/humans. A plethora of GSM BSs and UMTS Node Bs has triggered a lot of controversy and complaints from communities, NGOs, environmental organisations etc. Thus, through mobile relays we
    - make deployment
        - easier and ease any controversy.
        - cheaper. Less cost is required to incorporate them on “carriers” rather than deploy them on top of building (more expensive solution related to renting those areas)
    - provide less risk for humans in terms of transmissions
    - provide better aesthetical “intervention” and less impact to environment.

The general idea for mobile relays is not to totally replace fixed relays. The goals for both types of relays are the same. However, mobile relays can be seen as a complementary technology, used to provide incremental gain in cases where fixed relays cannot provide full/adequate solutions. An initial "idea" would be to "deploy" a large number of relative cheap mobile relays, so the cost remains low and use them just on top of conventional and fixed relay-based topologies.

In general, mobile relays should be seen under the view that the target is not to cover a specific area, but rather the "ad-hoc" needs which arise in an area, like e.g. the high concentration of UTs.

A first idea of how mobile relays could be introduced came under the technology of ODMA (Opportunity Driven Multiple Access). ODMA was introduced in 3GPP in order to provide coverage outside the coverage area of the BS. [99] The main idea was that a user terminal (in the edge of the cell) would be used and act as a relay to provide a link with a terminal that is out of the coverage area of the BS.

However, we must bear in mind that nothing is done for free. The trade off for using mobile relays is the relatively complicated processes that mobile relays require due to some inherent problems like not guaranteed minimum services (emergency services), not guaranteed area of coverage etc all of those related to mobility and presence, as we will see in the next sections.

Movable relays

Mobile relays are assumed to be constantly moving. However, we have to consider two other cases for the sake of completion of this study.

- Scenario 1. Assume a mobile relay that is fitted on a bus and for a relatively long period of time is stationary next to a football stadium e.g. for 2-3 hours. In this case, and for all this period of time, this relay can be assumed to be fixed. Thus, as long as in any way this relay will incorporate mobility functions, care should be taken so that during this period of time the mobility functions are either disabled or running with e.g. very low updates, so that when the full mobility commences again, normal mobility function are up and running again.
- Scenario 2. The other scenario would be of a type of relay that
    - Is brought in area
    - Is switched on
    - It provides relaying functionalities
    - Is switched off
  - This type of relay is effectively a fixed relay and as such it should not be required to incorporate any mobility functions. Thus, the only requirement would be that it is switched on, it configures itself to provide relaying functions. This type of relays can be characterised as "movable" relay.

As we see, movable relays are similar to fixed relays and in that sense do not posse so much difficulty in terms of the actual relaying functions being affected by mobility. Following the above points, we believe that at this point the mobility issues should investigated. As such, investigation of mobile relays will cover Scenario 1.

4.2.4.4.3 Types of mobile relays

Classification based on the relay mobility/ownership

An initial classification of mobile relays is the following:

- Mobile Relays type A (MRtA):→ User terminals (e.g. mobile phones) acting as relays (Typical owner the user)
- Mobile Relays type B (MRtB):→Relays being built only for relaying purposes (Typical owner the operator)

The reasons for doing so are the following

- MRtA will have built-in functionalities which can be used by the network for the relay functions. On the other hand, MRtB are relays that need to be designed from scratch and as such, it is up to us to define what functionalities we want to incorporate.
- Following this point, different processes should/could be taken by the network if it wants to use either MRtA or MRtB.

- Moreover, we should not only rely on user terminals used as mobile relays. Issues related to security or not granting access to relaying access, from the user's point of view, need to be considered and, thus, alternatives should be provided.
- Finally, different parameters need to be taken into account for simulation purposes depending on the type of relays. For instance, power availability is a problem for MRtA, whereas this is not the case for MRtB.

- Mobile Relays type A

Presence of user terminals in specific areas cannot be guaranteed. However, statistically we can expect that in a specific time of a day, in a specific area some terminals will be present. Of course, the larger the area or the deployed number of terminals is, the higher the probability is for terminals to be present in a specific time of day in that area. However, issues arise for the cases of presence during the night, which means that user terminals might not be able to provide basic/minimum emergency services during those times.

Additionally, we have to bear in mind that capabilities of user terminals either are limited or might not be applicable/suit us for relaying purposes, so it might be a frequent case that we cannot use them for complicated/demanding processes e.g. broadcast services. Typical example (and one of the main problems) would be the limited availability of power.

Furthermore, availability of user terminals might not "be there". A user might switch his phone off and thus in similar cases, we need to "find" another suitable one. This process might consume network resources.

Possibly of highest importance, is the issue of users not granting access to their mobile to be used as a relay for reasons of security, fast power drain etc. Thus, operators should at least provide some incentives to users if they want to use their terminal for relaying purposes.

In general, the main advantages of MRtA is that statistically we will always expect to have some terminals available to be used as relays, but the main disadvantages are related to the not-guaranteed availability which crystallises in the forms of "non-user permission" and limited functionalities, such as battery power.

- Mobile Relays type B

As with A-Type, MRtB presence cannot be guaranteed.

The number of deployed mobile relays type B would be far less than that of user terminals and thus, statistically, we have less probability of MRtB being in a specific area. However, a solution is to "schedule" mobile relays on certain "carriers", so that we increase the probability of their presence. By the term "carrier", we define all those mobile elements that mobile relays can be mounted on.

Let us, imagine a scenario of a park where there is no good coverage. Mobile relays could be fitted on buses, taxis or on other public transport carriers, that frequently pass by the park, so that statistically we expect at any time of day that at least a few (or just one) to be present. The more of them we deploy on e.g. buses, the higher the probability of presence is.

Other means of carriers are trains, tubes, ambulances, police cars, riverboats etc. The more scheduled/frequent a carrier is, the more probable it is for mobile relays to be present in the area of interest e.g. buses Vs tube.

Compared to MRtA there are some favourable advantages such as

- Potentially higher degree of incorporated functionalities, thus ability to cope with more complex and demanding needs
- Higher availability. "Always there-always on" approach. We do not depend on user interaction.

However, the main "bottle neck" is again the "not-guaranteed" presence, which is similar to that of MRtA, but apparently more severe.

- Discussion

What we have seen from the previous classification is that

- Presence cannot be guaranteed

- Depending on some approaches taken, statistically we can expect to have some "minimum" presence of a mobile relay
- We can increase the probability of presence by e.g. using frequent means of public transport, but this will still not solve the problem totally.

Thus, a number of issues which should be addressed are the following.

- **Presence**. How do we know that there are mobile relays present in the area of interest? Do we need to know? Do we transmit from the AP without knowing the presence of relays, thus do we use dedicated or broadcast RBs with those relays?
- **Location**. Do we need to "use" specific relays which are in a certain area, thus do we need to associate mobile relays with a specific location/area? Do we need to know their location and how often? Can current mobile location techniques provide adequate results? Are there any implications with current mobile location techniques?
- **Relay Capabilities**. Should the AP use all mobile relays that are within its vicinity? If not, based on what criteria will it make a selection of those to use and will it need to take into account the type of relays present and their capabilities? What will those capabilities be and how will (and with what implications) the AP be informed of those capabilities? What other information/measurement will be required?
- **Deployment scenario/Type of area**. If we assume we have different types of relays, with different functionalities, should we also take into account the deployment scenarios? For example, MRtB might be more suited to cover an area with e.g. broadcast services, whereas MRtA, might be more suitable to cover a much smaller area (for broadcast services) or just a small number of out-of-coverage terminals. Additionally, when we know in advance (from cell planning) that an area suffers from bad coverage then MRtB might be more applicable (due to some preplanning/dimensioning). On the other hand, MRtA might be applicable for cases of routing traffic (low bit rate preferably) for events of unexpected, but high concentration of UTs in the target area.

Classification based on their capabilities/functionalities

As stated above, the main target is to make mobile relays a cheap alternative, thus incorporate relatively limited functionalities. However, we expect that due to the diverse deployment concepts that mobile relays will need to cover, different types of relays will be built, from very simple to complex ones. There will be a trade off between flexibility and gain Vs cost, which needs to be addressed in the future.

Thus, another classification of mobile relays could be based on their capabilities, the functionalities incorporated, in the similar way to the fixed relays. For instance, based on what Layer 1, Layer 2 or Layer 3 functionalities we incorporate into them, we can classify them in the relevant categories. A "user-perceived" classification for MRtA based on their capabilities could be the following. High/Medium/Low-end MRtA equivalent to a mobile phone/PDA/laptop respectively. For a similar list of relays, please refer to [100]. Figure below shows a simple classification of relays depending on the AP/UT functionalities incorporated into them (effectively their complexity).

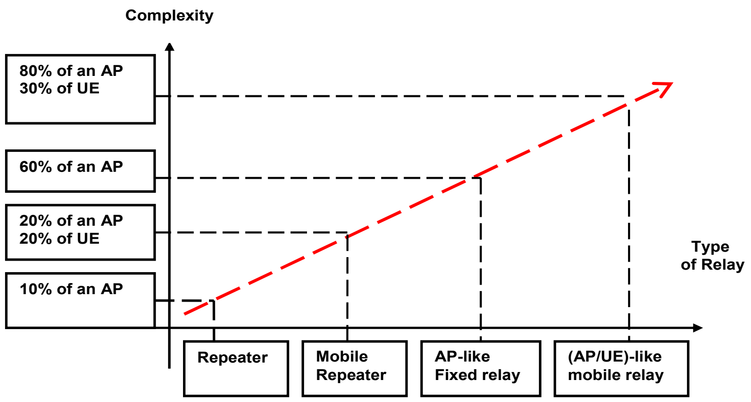


Figure 4-47: Types of relays with reference to the functionalities incorporated

4.2.4.4.4 Scenarios – Issues to take into account – Assumptions

Within WINNER WP3 a number of Scenarios have been identified, as depicted in figure below:

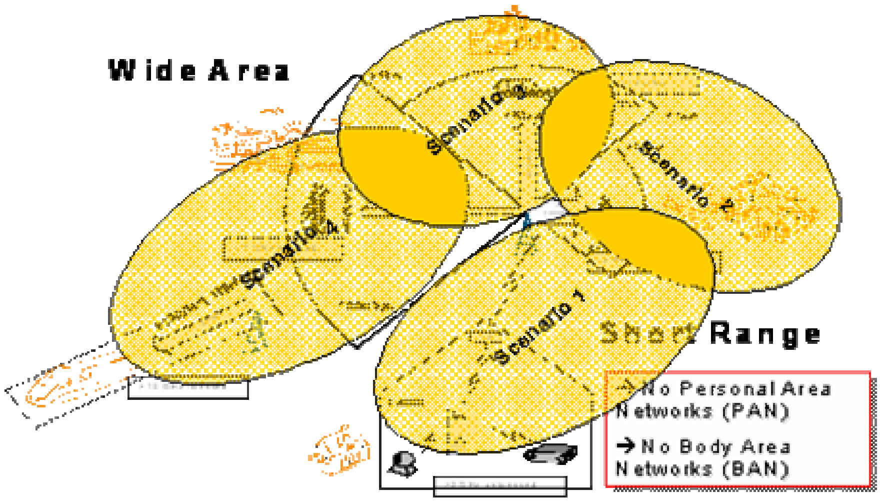


Figure 4-48: Scenarios as identified in Internal Report IR3.1

Out of those (4+1) scenarios (four depicted in Figure 4-48 and one which is the feeder scenario), two are more applicable for mobile relays.

- Scenario 3→ Urban / Suburban
- Scenario 4→Rural / High Speed

More information on those scenarios is included in [101].

Elaborating more on those scenarios, thus moving to a higher degree of detail in order to identify the users cases/deployment concepts we can list a number of issues with reference to mobile relays-related scenarios. These are some of the degrees of freedom/parameters we need to take into account and which will be used for evaluating certain deployment concepts and techniques.

These parameters need to be defined so that for each of the combinations of scenarios/deployment concepts/user cases we define *realistic* values that are required further for simulation purposes, so that we make sure that for each of the deployment concepts we have a business case. For instance, we could not expect from an MRtA to be used for providing coverage on a train due to e.g. many handovers or limited capabilities.

Thus, some of those issues are the following:

- Areas (Cities, Rural areas)
- Type of Land (Land, Sea etc)
- Presence (MRTA and/or MRtB and/or fixed relays)
- Number of APs, Relays, terminals, terminal “able” to act as relays
- Types of mobile relays and their relevant capabilities
- For mobile relays, issues related to inter-arrival time e.g. on buses
- Power of AP/Relays/UT and coverage
- Environments (e.g. urban, pedestrian)
- (AP-Relay), (Relay-UT), (AP-UT) channels e.g. propagation conditions.
- Mobility (Random→ Tourist in Trafalgar square, Predefined→buses/tube, Straight line or circle (Highway, train line)
- Mobility of UTs/Relays with reference to Relays/APs accordingly (e.g. UTs moving but always within the vicinity of the Relay)
- Velocity (Very low→Pedestrian/User terminal, Low→Ships/Riverboats, Medium→Cars, Tubes, High→ Trains

- Type of traffic/Bit rates (High, Medium, Low bit rates/ Voice-Data traffic) handled

Table 1 shows some values for some basic parameters. Although the values are not "exact", the intention is to show the diversification of those parameter for any of the scenarios or types of mobile relays. A similar, but much more detailed table, should be considered when building system level simulations.

Table 4-10: Initial assumptions on basic dimensioning parameters

	Scenario 3 – Urban/Suburban			Scenario 4 – Rural/High Speed		
	Power	**Coverage**	**Velocity**	**Power**	**Coverage**	**Velocity**
AP	5W	500m	NA	15W	2km	NA
MRtA	0,2/0.5W	20m	3km/h	NA	NA	NA
MRtB	2W	100m	10/30km/h	10W	1km	80km/h

[Note: The parameters in the table are merely indicatory.]

Similar and more detailed assumptions (for fixed relays though) can be found in the relevant sections of [102] and [103].

4.2.4.4.5 Deployment concepts

Introduction

Following the discussion in the previous sections we can identify some general deployment concepts for mobile relays with reference to some priorities we assume.

Deployment concepts with reference to the needs to cover

The first classification (two cases) is based on the initial needs for introducing Relays

- Case A (Figure 4.1/a) →Area of interest inside the AP coverage area (UT in the coverage of the AP)
- Case B (Figure 4.1/b) →Area of interest outside the coverage area of an AP (UTs outside the borders of the AP)

For case A the main idea is to provide better quality, higher capacity and data rates in an area that was supposed to have adequate coverage within the border of the AP's coverage area. Case B addresses the issue of providing "new" coverage to areas that previously there was no coverage at all. Thus, the main strategy for Case B is to reduce infrastructure cost, as long as in conventional approaches we might consider deploying additional BSs/APs.

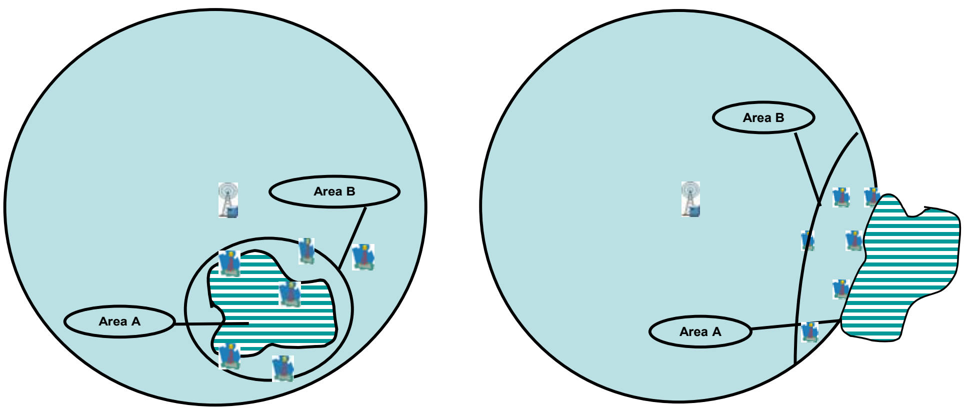


Figure 4-49a/b: Problematic area well in/outside the AP coverage area

Note: Area A → Problematic area, Area B → Area within which relays can be used to cover the problematic area

Deployment concepts with reference to the relay type

As stated previously, different strategies/parameters will be considered depending on the use of mobile relays type A or type B.

- Terminal acting as Relays (MRtA)

In this case we have no prior information about the location/capabilities of the potential relays. Thus, one possible approach would be for the network to "look" for terminals willing to act as relays in the vicinity of the area it wants to cover. This process would probably include paging and positioning of those relays. A more "sophisticated" approach would be to look for potential relays with certain capabilities.

A scenario that would fit this approach is a relative small area, with urban characteristics, with terminals of, on average, low velocity (or stationary), with relatively limited capabilities, of "in general" unknown (but possibly statistically predicted) mobility or availability.

- Relays built only for relaying (MRtB)

In the case of Type-B relays, the same general concepts as above apply. However, some differences exist. We "can know" or assume, based on some initial pre-planning, that in the area of interest mobile relays will be present (e.g. public transport). Another difference is that due to the type of "carriers", we expect to have fewer MRtB compared to MRtA, the latter expected to spend more time (due to their "on average" lower mobility) in the cell. However, MRtB can be used for free – no "selfish" UTs.

A scenario that would fit this approach would be a large area, with MRtB having higher velocity, high relaying functionalities/capabilities, high power availability, known or predictable mobility.

Deployment concepts with reference to the mobility of the UT population in relation toi the Mobile Relay

Another classification of concepts is with reference to the Relay-UT mobility

- UTs - Relay mobility uncorrelated: Conventional case of relays targeting to cover a specific geographical area.
- UTs - Relay mobility correlated: Stationary UTs with reference to the MRs. Relay is used to cover a moving (along with the Relay) population of UTs e.g. train, riverboat

The main difference between the two cases is that in the first one, the main goal is the *area to cover*, however, in the second it is the *population of UTs*. Basically, the second case is simpler as long as the UT population/area-to-cover is moving along with the Relay.

A simple example to show the difference in those approaches, is that while in the second case the transmitted relay power might be fixed (or at least with small variations), in the first case a different power control scheme must be employed. While the MR is moving away/close/away from the target area, the transmitted power should be high/low/high respectively, so that we ensure the "same" reception levels in the target area, but also minimum interference when the MR is not needed (away from the target area).

Conventional case – The goal is the "area"

The generic figure that can portray the movement of relays within an area is the following. In this case, the mobility of relays and UTs is uncorrelated.

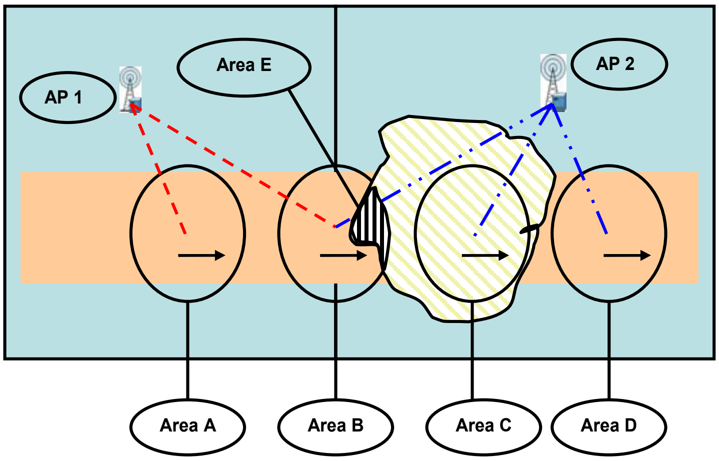


Figure 4-50: Coverage of problematic area with MR

The MR is moving along the horizontal axis (on a road) and at different times its coverage is defined by Area A/B/C/D. (We assume fixed transmitted power). During the movement, it maintains links with two APs (AP1 and AP2) (one at a time or simultaneously). The lightly shaded area is the target area to cover.

Thus, the following cases can be seen in the above Figure. (Areas← →Cases):

- A: MR can listen only to AP1. Mobile relay outside the "target" area
- B: MR can listen to both AP1 and AP2.
- E(sub-case of B): Partial coverage of the "target" area.
- C: MR can listen only to AP2. MR within the area of interest.
- D: MR can listen only to AP2. Mobile relay outside the "target" area.

Stationary UTs (with reference to the MR) – the goal is the "UT population"

Another approach is to use the MRs to cover areas/UT populations which are in moving vehicles e.g. trains, buses, riverboats etc.

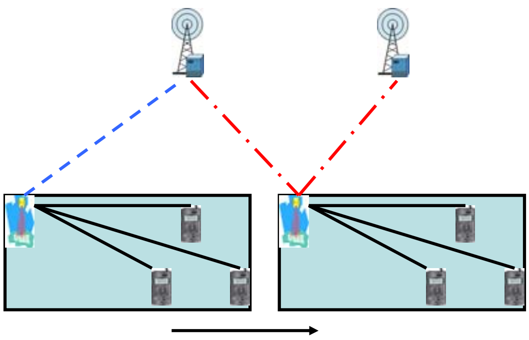


Figure 4-51: UTs stationary with reference to the Relays (Train / Ship scenario)

In these scenarios, the UT-Relay mobility is correlated (in this case moving with the same velocity). The main difference in this scenario is that the UT population can be assumed to be stationary compared to the MR, as Figure 6 portrays. However, we have to bear in mind that due to certain constraints we have to take into account that in some cases antennas fitted on top of trains could not provide coverage within the train, so some other solutions e.g. feeders might be required.

Conclusion on deployment concepts

The main intention in the section was to show that there are different parameters we need to take into account depending on the type of scenarios/deployment concept/user cases we follow.

4.2.4.4.6 Issues for consideration / Implications of Mobility

Introduction

A number of issues need to be resolved and are being addressed for fixed relays. [104] However, mobile relays take this a step further. The main issue with mobile relays is that of mobility, and all the implications/consequences that this brings, which effectively points to the main issue which is the duality of the mobile relay, acting like an AP for the UTs and at the same time as a UT for the AP.

This means that a lot of issues need to be considered in advance so that we have a more broad idea of the implications of mobile relays. Thus, we are able to evaluate certain techniques and approaches better, before deciding what types of mobile relays we need to consider for each of the scenarios/deployment concepts.

Thus, in this section we will address in short some of those issues/areas which we will take into account for the link/system level simulations.

List of items

Specifically we have the following items some of which we touched upon on previous sections.

- Multiple Access (MA) techniques

Each time within a cell we expect to have a number of mobile (N) and fixed (M) relays. This number will frequently be changing and will not be fixed. This means that resources need to be allocated to those relays based on e.g. the type of the Multiple Access (MA) scheme applied like CDMA/TDMA. For instance, if we assume that each relay is assigned a different scrambling code, then effectively we have a CDMA scheme. If we assume a TDMA scheme, then due to the scarcity of the time domain, we need to split the time frames into slots and allocate them to each relay. A trade off between e.g. duration of time slot Vs number of time slots needs to be thought of. The MA schemes will inevitably also affect the interference issues. [104][105]

For fixed relays what has been proposed and seems more applicable is a TDMA approach. [106]

Also within each MA scheme any other related techniques should be investigated like reservation (dynamic or not) schemes for resources when allocating e.g. time slots to the mobile relays.

- Duplex scheme (FDD/TDD)

Duplex scheme needs to be decided. For instance, if we use FDD, this means that a separate frequency band needs to be assigned for uplink and downlink. Advantages and disadvantages need to be considered.

- Registration/Paging/Relay discovery/Identification of relays

With “pure” mobile relays (MRtB) there is always the need to provide some coordination. This means that we need to be able to count their numbers, gather their capabilities, associate them with an area/location etc. In order to do that we should be able to identify each one of them separately (or in groups) and be able to address them separately (or in groups) which brings up the issue of identification and paging.

Thus, two approaches exist

- A “discovery mechanism” which will be a “paging-like” process (wake-up mechanism) through which the AP could collect information for the existence of mobile relays.
- A registration process which will occur with the AP/APC whenever a mobile relay (applicable for MRtB - “pure” relays) enters the coverage area of an AP/APC.

These processes are very similar to UT processes. However, we should investigate what functionalities we need to omit or add – compared to the typical UT functionalities.

Another issue is how we distinguish each relay from the other. Do we use temporary or permanent identities and at which level e.g. within the AP (cell level) or within an APC area (collection of cells)? For instance, if we assume handover processes/soft combining in handover, then we need a global identity to distinguish each MR not only within a cell but within an APC area.

- Handover

As it was obvious in the previous section, mobile relays need to perform some UT-like processes. Thus, each mobile relay will act as a “terminal” to its access point. This means that mobile relays may be needed to monitor a number of APs for handover processes, in the similar way to UTs. Of course, handover processes will increase the complexity level of a mobile relay, but in the end it might be beneficial to apply a scheme, probably, though, not as complicated as in the UT case.

So a number of issues need to be addressed

- Need for HandOver (HO)
- Types of handover (hard / soft)
- Number of APs a mobile relay needs to monitor. Similarly to UMTS will we have an active/monitoring set of APs?
- Measurement that need to be performed to support HO
- Functionalities and physical location of the related mechanisms
- Signalling overhead
- Triggers
- Channels that mobile relays need to monitor (CPICH-like channel?)

- Power control/Allocation schemes

In CDMA systems power needs to be controlled very accurately. Otherwise, interference rises especially due to the near-far effect. For conventional systems e.g. UMTS, different schemes are employed, like open-closed/fast-slow power control. For relay-based systems, power control seems even more crucial. In these systems we have three degrees of freedom to take into account, two of which are "mobile"; the mobile relays and the user terminals. This inevitably means that more measurements and signalling will be required and one parameter which will probably be affected will be how rapidly power can be adjusted to the new conditions.

With mobile relays, due to the mobility we have to decide what our priorities are. Let us take the examples of a deployment concept with reference to the UT-Relay mobility. For a train scenario (UT and Relay moving with the same velocity), the DL required power for the relays will be expected to be either fixed or with very small variations. However, in the case where the target is a problematic area (mobile relay mobility different from that of the UTs to accommodate), then we expect that as the mobile relay approaches/moves away from the target area it will increase/decrease it transmitted power, so that we provide "stable" receptions levels at the target area while not inducing interference in other areas.

Thus, we should investigate strategies for power control/allocation of mobile relays and what is the effect on interference (as a result of possibly degraded power control schemes).

Types of mobile relays – Classes/Capabilities

As stated above, we need to have simple mobile relays that cut down the cost of deployment. At the same time, however, we need to have relays that can support some minimum required functionalities e.g. mobility. Thus, building up on those simple functions, we might end up with relatively complicated mobile relays, whose cost might be even comparable to that of APs. Thus, we should examine the pros and cons of simple or more complicated mobile relays.

Additionally, due to the diverse needs we need to cover we expect to have plethora of types of mobile relays in relay-based deployments. The network, as we will see further, needs to take into account the capabilities of those relays. Thus, some issues to consider are

- What type of capabilities do we need to monitor? (static or dynamic[1] ?)
- What type of signalling will be required (DL/UL?)
- Can we reduce signalling requirement through other means? (define classes of relays instead of capabilities?)
- Which are the processes that capabilities will be taken into account? (RRM?)

[1]By the term dynamic we define capabilities which change in the course of time e.g. power availability, velocity of a relay etc.

(AP- Relay) – (Relay-UT) signalling issues (UL/DL)

Issues related to general UL/DL signalling between the AP and the relay need to be considered.

- What type of parameters will be required to be transmitted over the AP-Relay or Relay-UT physical layer link?
- What type of signalling? (Dedicated or broadcast or multicast radio bearers?)
- How often will this signalling be occurring (periodic or trigger based?)
- Can we deploy techniques to reduce the amount of signalling?
- What will be the impact to the PHY resources?

Cooperative Relaying

Cooperative relaying is another attractive technique which can be used to provide incremental gain in the quality of the reception signal at the UT. A separate section is dedicated to cooperative relaying especially in relation to mobile, so for more information please refer to that section.

Mobile Positioning

As we touched upon this issue, mobile positioning will be required for several reasons e.g. RRM algorithms. What is more, compared to positioning in 3G, more accurate, fast and reliable algorithms will be required. Thus, it is important to investigate

- Current and new positioning techniques
- Requirements e.g. accuracy
- Problems that might arise and possible solutions

Let us take the following example. When using timing techniques for positioning, the additional delay that is introduced through a relay on the path that will be used to measure the e.g. RTT (Round Trip Time) between the AP and the terminal, does not correspond to the distance between the terminal and the AP

Measured RTT: RTT'(AP-UT)=Time(AP←→Relay) + ***D(Relay)*** + Time(Relay←→UT)

Correct RTT: RTT(AP-UT)= Time(AP→UT) + Time(UT→AP)

AP-UT distance: D(AP-UT)=c*RTT

Note: D(Relay) is the total delay introduced by the relay processing. Time(x-y) is the RTT time it takes for a signal between those two elements

The larger the D(Relay) is, the more degraded the positioning becomes. This factor should effectively be zero for accurate positioning. This issue was originally identified in 3GPP [13] [14]. Thus, positioning is severely degraded and solutions should be proposed to cope with this problem. [15] However, more concrete solutions should be considered for more complex scenarios e.g. more than two hops.

Channel models / Propagation issues

For each of the relevant scenarios/deployment concepts, different propagation cases need to be taken into account e.g. park scenario/ train scenario.

Even "within" one deployment concept different propagation conditions need to be considered for the (AP-Relay) and (Relay-terminal) channels e.g. height of antenna, antenna gain, shadowing etc.

Resource coordination/ Management / Allocation

As stated previously, in future network deployments we might have a large number of mobile relays to be used, with different capabilities and available resources. Additionally, the needs we will be required to cover will be diverse and frequently changing.

Thus, it might be beneficial to use just part of those relays because not only we want to use those that are absolutely needed and not waste resources, but additionally, not cause other problems like interference. Thus, coordination of resources and scheduling of which relays to use (and how) should be in place.

Thus, we should deploy a selection process, which will take into account a number of information and decide on what the best/optimum relays are to use, while making sure no side effects are produced e.g. interference, or minimise those side effects.

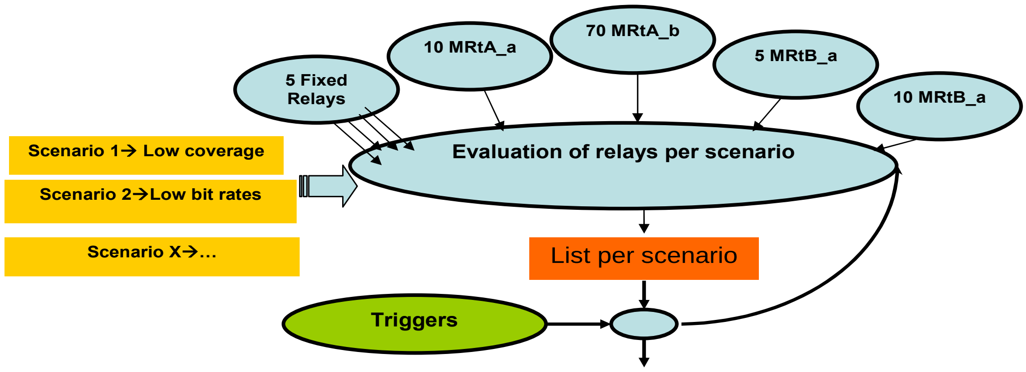


Figure 4-52: Flow chart of decision making process for optimum relay selection

Figure 4-52 portrays the selection process for deciding on the optimum relays to use for each of the identified scenarios within a cell. "*_a/b" defines a complex/simple relay:

Reconfigurability

As we saw in the previous section one way to fulfil our needs within a cell is to try to match our needs with the relay capabilities we have. Another approach could be that of reconfiguring those mobile relays, based on our needs. This might have gains in e.g. the number of mobile relays to deploy and the reduced need to provide mobile relays suited for each deployed scenario.[110]

Thus, we assume we have deployed a number of "flexible/generic" mobile relays which can be configured by the network, which is a more direct approach for re/configuring the network resources.

Some of the approaches could be the following

- The relays download the relevant "codes" and execute them. This approach requires heavy signalling, dedicated signalling, but not complex relays.
- The network signals the "type" of techniques e.g. convolutional code ½, but not the actual code. The mobile relay (having already either downloaded the relevant code or having it from its original configuration), just selects the appropriate one. This approach requires much signalling, dedicated signalling, more complex relays and is not as flexible as the previous approach.
- The network signals a pre-defined configuration. The relays uses a "look up" table (having downloaded it either offline through broadcast services or from its original configuration) and selects the relevant techniques defined by this configuration. For instance, the AP will signal a bit stream of 100 bits that will define the combinations of Layer 1/2 techniques to be used. This approach requires less signalling, more complex relays, but is not very flexible.

Other approaches could be proposed as combinations of the above or new ones e.g. association of needs with certain geographical areas.

This reconfigurability is applicable mostly to "pure" mobile relays (MRtB).

Heterogeneous Mobile Relaying

Heterogeneous fixed relaying is introduced and investigated within Task T3.3 of WINNER WP3. [111] Thus, the previous section of reconfigurability could be seen under the view of reconfiguring each mobile relay to be able to receive and transmit within a heterogeneous environment, without the need to deploy, *at first place*, heterogeneous mobile relays.

However, this might be a complex issue and should be for future studies.

Mobile relaying and MBMS

In the previous sections we have assumed that each mobile relay handles dedicated traffic which requires dedicated signalling and as a result certain MA technique might be more applicable e.g. TDMA scheme for fixed relays.[106].

One other approach where mobile relays could be used (and which is also applicable for fixed relays) is with broadcasts/multicast Radio Bearers (RBs). This will result in gains in signalling, resources (TDMA slot allocation), power saving etc.

Specifically,

- There might be no need for dedicated signalling, identification of mobile relays etc. The AP transmits the same information to the entire cell. Each relay reproduces the same info.
- If a TDMA scheme is applied (e.g. for fixed relays) time slots need to be allocated to each relay. With broadcast/multicast RBs this is not required. Figure 4-53.
- By having mobile relays reproduce the broadcast service to the areas close to the edge of the cell, the AP can effectively reduce its broadcast channel Tx power down to the level of just reaching the mobile relays. Thus, the edge of the cell could be "serviced" only by the mobile relays. As a result, we might have gains in interference, capacity etc. Figure 4-54

An example of multicast/broadcast modes is the UMTS MBMS RBs which are investigated within Release 6 of 3GPP. [241]

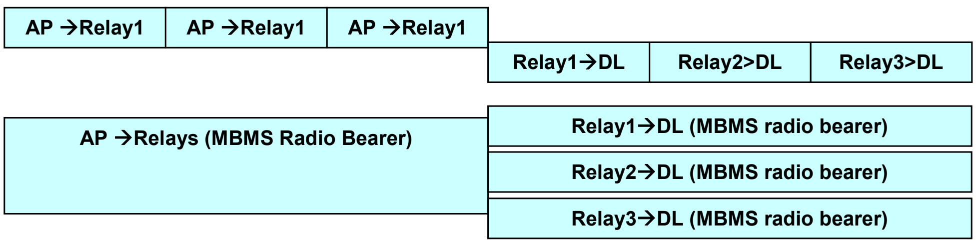


Figure 4-53: TDMA scheme with 3 relays and MBMS RABs

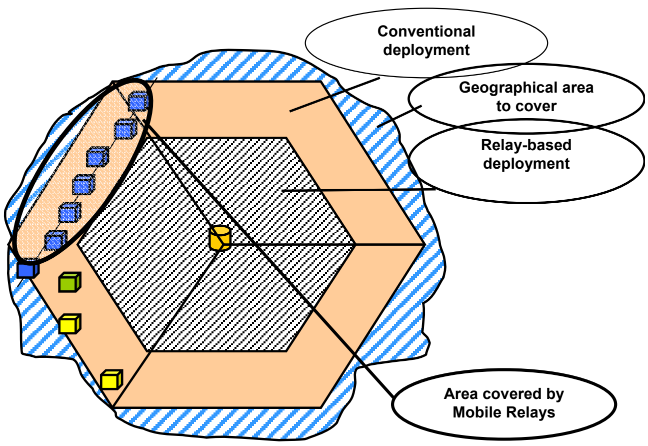


Figure 4-54:Extension of coverage through relays by deploying broadcast/multicast radio bearers

4.2.4.4.7 Conclusion

Mobile relays have been proposed as part of the relay-based topologies for future network deployments to provide better coverage in cellular/wireless environments.

In this section we presented a short overview on mobile relays. We presented the basic needs that mobile relays could cover, we investigated some typical deployment concepts and while we identified problems due to mobility, we presented some alternatives and ideas on how certain aspects related to implications of mobility could be resolved. The main intention, in general, was to have an introduction on the concept of mobile relays.

As we have seen, mobile relays pose a number of challenges mainly those related to the issue of mobility. At the same time, though, as we have analysed, mobile relays could potentially better cover certain needs compared to conventional or fixed relay-based topologies.

Following the above we believe that simple schemes incorporating mobile relays could be deployed to be used as “complimentary” solutions when required. The idea is not to totally replace fixed relays, but rather, in most of the cases, to be used along with them.

Initial simulations for evaluation purposes will include link budget / system level simulation e.g. SINR measurements.

4.2.5 Models and algorithms for routing/forwarding in multi-hop wireless networks

4.2.5.1 Introduction to Routing/Forwarding in Infrastructure-based Multi-hop Networks

Routing/forwarding algorithms for infrastructure-based multi-hop networks in the context of homogeneous relaying is the primary scope of this subtask. The main focus is the fixed relays deployed by the service provider; wireless terminal relays, which have inter-terminal communication (mesh) capability, are also considered.

The overall goal of this subtask is to facilitate two-hop or multi-hop communications through efficient and intelligent forwarding and routing algorithms, respectively, in relay-based cellular, WiFi, and WiMax networks. Pure ad hoc networks which do not rely on any infrastructure (Mobile Ad-hoc NETworks [MANETs]) are outside the scope of this subtask.

Since the basic routing/forwarding operation is relatively straight-forward in infrastructure-based networks in comparison to ad hoc networks, goals more advanced than the simple identification of routes between BSs/APs and wireless terminals will be sought such as QoS optimization (including throughput maximization, delay minimization, etc.) and traffic diversion through load-balancing.

What distinguishes the proposed work from the ad hoc routing literature is the following: the main goal in ad hoc routing is to establish and maintain connectivity; but this is a relatively easy task in infrastructure-based networks due to the presence of a common source or sink which may have considerable complexity and intelligence (BS or AP). Therefore, other more involved goals can be targeted in choosing routes in infrastructure-based networks, such QoS optimization (including throughput maximization, delay minimization, etc.) and traffic diversion due to load-balancing. It is also worth emphasizing that routing among fixed relay stations and routing among mobile/movable relay stations have quite different characteristics which should be taken into account in developing routing algorithms. However we could use the MANET protocols (e.g. DSDV, DSR) as comparison in order to show the expected improvement of the proposed solution.

The problem of routing in multi-hop wireless networks has been deeply investigated in the context of MANETs. These networks are characterized by a varying topology and require a completely distributed operation of routing protocols. Routing mechanisms for MANETs usually aim at minimizing the protocol overhead needed to manage frequent topology changes by means of on-demand route formation [112][113][114] or hierarchical schemes [115]. Therefore, route selection is based on inaccurate network state information, and the metrics adopted try to optimize the energy consumption rather then the use of radio resources.

In this hostile network environment the challenge of providing Quality of Service (QoS) guarantees to traffic flows has attracted attention by the research community [116][117] and several algorithms have been proposed [118]-[124]. QoS routing algorithms proposed so far for MANETs are tailored to specific MAC (Medium Access Control) layers able to provide information on resource availability and to control resources assigned to traffic flows. The more common approach is to consider a TDMA based ad-hoc network [123]. Each connection specifies its QoS requirement in terms of time slots needed on its route from a source to a destination. For each connection, the QoS routing protocol selects a feasible path based on the bandwidth availability of all links and then modifies the slot scheduling of all nodes on the paths. This task is not easy and even the calculation of the residual bandwidth along a path is NP-complete [123] since can be easily shown that the slot assignment problem is equivalent to the graph coloring problem [125].

Infrastructure-based wireless networks can be extended in their coverage by mobile nodes with relay-capabilities spontaneously acting as forwarder for other nodes outside the range of components belonging to the infrastructure, specifically APs. In particular, self-organizing networks are promising candidates to wirelessly extend the fixed network infrastructure. Though they cannot provide any guarantees with respect to quality-of-service and grade-of-service, they extend the coverage for best-effort services in a spontaneous, flexible, and cost-efficient way. In such networks nodes can communicate without the need of an infrastructure with each other by using intermediate nodes as relays, resulting in multi-hop communication. Thus, data from / to the Internet have to traverse several hops between the AP and those stations over a respective path. One basic challenge in such kind of networks is the routing. Many papers on routing in ad hoc networks have been published and good overviews are given in [126], [127], [128], [129]. In general, existing routing approaches can be classified in three categories: proactive, reactive and position-based algorithms. In proactive schemes, all nodes maintain routing information about the available paths in the whole network even if these paths are not currently used. Hence, proactive schemes do not scale well with network size, and frequent network changes will result in high traffic load caused by signaling messages used for route maintenance, making this approach less suitable for very flexible and dynamic network topologies.

Reactive routing schemes, also called on-demand routing, establish and maintain only paths that are currently in use, thereby reducing the signaling traffic. Nevertheless, they have two inherent limitations when the topology changes frequently: first, even though less routes have to be maintained than in proactive approaches, route maintenance may still generate a significant amount of signaling traffic, and second, packets on their way to the destination are likely to be lost if the route to the destination breaks.

The last category, position based routing algorithms, eliminate some of the mentioned deficiencies of proactive and reactive algorithms. They only require information on the physical position of the participating nodes. A comprehensive overview and investigation is given in [128]. Since it can be expected that in the near future all devices can be located respectively their position is known, e.g., for the purpose of location-based services, this requirement is fulfilled in the WINNER system. The forwarding decision in position-based routing is based on the destination's position contained in the packet and the position of the forwarding node's neighbours. Position-based routing, thus, does not require the establishment or maintenance of routes. The nodes neither have to store routing tables nor do they need to transmit messages to keep routing tables up-to date.

4.2.5.2 On the Efficiency of Using Multiple Hops in Relay Based Networks

It is necessary to establish strategies and methods for efficient deployment of fixed relay stations, in such a way the overall cost of the network is minimized. Efficient radio resource allocation to network elements is a critical part of the overall network cost optimization effort. The main goal of this section is to introduce a criterion to determine in which conditions a single-hop fixed radio link can be efficiently (from a spectral efficiency perspective) replaced with a chain of (multi-hop) links. Identification of such a criterion will be helpful in determining the advantages and disadvantages of multi-hop routing.

One of the open questions regarding the deployment of wireless networks using fixed relays is the optimal number of hops between the source and destination radio stations. It is important to be able to decide with reasonable accuracy in what conditions it is more advantageous to send a signal directly to destination (may be by increasing the allocated transmit power, bandwidth or time) or route the same signal over a number of relay stations, each using much less resources compared to the replaced link. Due to their ambitious requirements, the B3G systems will most likely use solutions with at least two hops (whenever necessary), one of the hops being between the fixed relay and the mobile terminal. The remaining radio link going back to the base station, referred to as the “feeder system” in [130], is comprised of a set of fixed relays interconnected through radio links, and can have one or more hops. The underlying scope of the following discussion is to determine the optimal number of hops of the feeder portion of the B3G wireless system.

In this work we take a novel approach by analyzing the aggregate spectral efficiency of the multi-hop communication system. The additional radio power introduced in the network by relays comes from the “wall plug” and should not be added to the radio resource costs. More significant is the effect of the relaying schemes over the aggregate end-to-end spectral efficiency. Higher overall spectral efficiency allows better use of the available spectrum license – the most expensive asset of the cellular operator.

4.2.5.2.1 System Description

We consider the multi-hop link S-R in the feeder part of a fixed relay network with an n-hop link, as shown in Figure 4-55, where S and R are the source and the recipient stations, respectively. The message can be either sent directly from S to R (single-hop operation), or the message can be sent via n-1 intermediate fixed relays over n hops.

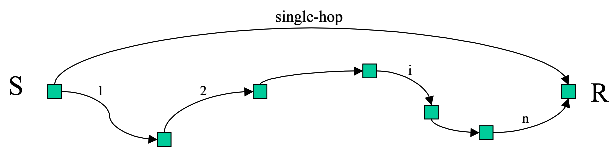


Figure 4-55: Multi-hop link topology.

The relays are of the digital regenerative type. Since the relays are fixed, the topology of the system is considered known and non-dynamic, so no routing algorithm is required. As the overhead messaging related to multi-hop functionality will not be significant, it was not considered as a factor in the discussion below. As well, the processing delays in the relays have been considered much smaller compared with the transmission time of relayed data packets.

The system uses a time-slotted resource allocation scheme, where each radio link is assigned a channel in the frequency – time domain. Each intermediate radio link adopts an appropriate modulation scheme resulting in the best possible spectral efficiency based on the given SNR conditions.

We adopt the following assumptions as stated in [131]:

- The amounts of data received and transmitted by each relay are equal.
- For simplicity we assume that all *n* hops of the multi-hop link, and as well the single-hop S-R are allocated the same amount of bandwidth *B*. The individual time required to pass a message over the hop *i* is

$$t_i = \frac{M}{Bx_i}, \tag{6}$$

  where *M* is the message size in bits and x_i is the spectral efficiency (in bits/sec/Hz) of the radio link over hop *i*.
- The timeslots t_i allocated to each hop in the multi-hop link are considered orthogonal to each other. Also for simplicity but without loss of generality, we assume that all links operate on the same carrier frequency. Although the orthogonality condition seems conservative, as it will be seen later, in most cases the number of hops envisioned for the multi-hop link is three or less; with such a low number of intermediate links and given the high SNR required by the desired data rates, it is unlikely that any frequency or time slot reuse would be possible, unless advanced processing techniques or antenna architectures are employed. In conclusion, the total time, *T*, required to pass a message of size M from S to R is the sum of all intermediate timeslots:

$$T = \sum_{i=1}^{n} t_i = \frac{M}{B}\sum_{i=1}^{n}\frac{1}{x_i} \tag{7}$$

- Since the channels used by each hop in the multi-hop link are orthogonal, no co-channel interference is present.
- All radio links in the system have similar radio propagation parameters.

4.2.5.2.2 Optimal Relay Locations

The analytical development for the optimal relay locations is presented in Appendix 7.2.1.1.

We consider the case of a *n*-hop link with "short" intermediate links, i.e., all intermediate hops lengths d_i respect the condition (82), and as a consequence all intermediate hop SNRs γ_i are as in (84). According with (87), such an *n*-hop link would have a better aggregate spectral efficiency (smaller message transfer time) compared to a *n*-hop link having the length of all intermediate hops the same and equal with the mean hop length.

Further, in the particular case when all relays are placed on the straight line SR, any configuration has a better aggregate spectral efficiency compared with the situation when the *n-1* relays are distributed along equal intervals on the line SR. In other words, if all relays are placed on the straight line SR, evenly spaced relays achieve the worst performance (lower bound) in terms of spectral efficiency.

Considering now the opposite case, of a *n*-hop link with "long" intermediate links, we assume that all hops lengths d_i respect the condition (81), the intermediate SNRs γ_i are as in (83), and $f(d_i)$ is strictly convex. In this case, (85) stands true.

Geometrically, the smallest value of the sum of all *n* individual hop lengths is reached when all relays are placed exactly on the straight line SR. In that case

$$\sum_{i=1}^{n} d_i = D. \tag{8}$$

As a result, the best possible geographical locations for the *n-1* intermediate fixed relays which would minimize the sum term in (79) and the total message transfer time *T* are along equal intervals on the straight line SR. In such a case,

$$d_i = \frac{D}{n} \tag{9}$$

for i = 1, 2, …, n.

We have reached the apparently peculiar conclusion that there is no unique optimal configuration for the locations of the *n-1* relays. For "long" intermediate hops, the configuration with evenly distributed relays is optimal, while for "short" intermediate hops, it is the worst. An example of a two-hop link is considered in Appendix 7.2.1.2 to graphically show this behaviour.

4.2.5.2.3 Multi-Hop Criterion

It is more advantageous to use the *n*-hop link if the end-to-end spectral efficiency is improved. With the message size *M* and bandwidth *B* being the same, a single-hop link should better be replaced by a multi-hop link if

$$\sum_{i=1}^{n} \frac{1}{x_i} < \frac{1}{a} \tag{10}$$

where *a* represents the spectral efficiency for the single-hop link S-R. The condition (10) simply states that in order for the multi-hop link to be more efficient, the time required to pass a message of a given size *M* from S to R over the single-hop link must be larger that the time required for the same operation over the multi-hop link.

Using (85), (86) and (8), we express the lower bound on the message transfer time using *n*-hops as

$$\begin{aligned} T &\geq \frac{M}{BK_1} \frac{n}{\log_{10}\left(K_2 \left[\frac{D}{nd_0} \right]^{-p} \right)} \\ &= \frac{M}{BK_1} \frac{n}{\log_{10}\left(n^p \gamma \right)}. \end{aligned} \tag{11}$$

The expression (11) shows the smallest possible message transfer time using *n*-hops, given that (83) is true. If the single-hop link can transfer the data in a shorter or equal time, then there would be no point in using relays. Using (11), the inequality (10) can be rewritten as

$$\frac{n}{\log_{10}(n^p \gamma)} < \frac{1}{\log_{10} \gamma} \tag{12}$$

which can be simplified to the expression

$$\gamma < n^{\frac{p}{n-1}} \tag{13}$$

If for a given single-hop link condition (13) is not met, a *n*-hop link replacement with a better overall spectral efficiency does not exist, no matter where the relays are located. The inequality (13) represents a quantitative criterion which can be used to decide in which situation a multi-hop link could be used.

The "break-even" SNR values in (13) are plotted in Figure 4-56 for various values of the path loss exponent *p*.

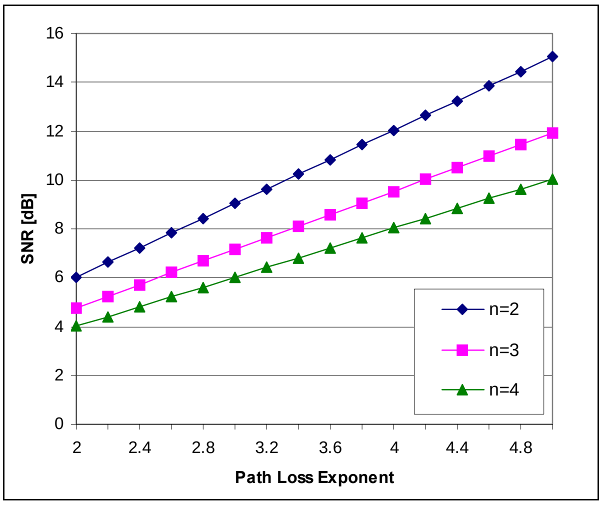


Figure 4-56: SNR values above which the single-hop has better spectral efficiency compared to any n-hop

The plots show the variation of the threshold (or “break-even”) SNR value at which a more efficient multi-hop alternative to the given single-hop link becomes feasible. For example, in Figure 4-56, if, for instance, the path loss exponent of all intermediate hops happens to be $p = 3.6$, then, according to (13), there exists a possible 3-hop configuration (the 2 relays evenly distributed along the line SR) with a better aggregate spectral efficiency, as long as the single-hop link SNR is less than 8.6 dB.

4.2.5.2.4 Remarks

The criterion developed above uses as a performance metric the bandwidth (frequency·time) needed for transferring messages from source to destination. The energy required for transfer is not considered as a metric; that is, the additional power inserted in the system by intermediate relays is considered “free”.

As the discussion above is based on mean SNR values not including shadowing, these results are only applicable for a statistical average over a set of multi-hop links; the results above are not binding on a given particular realization of a multi-hop link.

In the case of $n=2$ (one fixed relay between S and R), the inequality (13) becomes $\gamma < 2^p$, in concordance with the result in Appendix 7. For example, if the propagation exponent has a value of 3, one relay placed in the ideal location (right in the middle of the link S-R) would be efficient only if the SNR of the link S-R is less than 8 (9 dB).

The multi-hop criterion is in general applicable to multi-hop links with all individual hops having similar radio propagation characteristics, i.e. equal path loss exponents. In the special case when the mobile access link (the last hop of a B3G cellular link using fixed relays) has the same path loss exponent as the feeder system, the multi-hop criterion can be extended to cover the entire link between the base station and the mobile terminal.

4.2.5.2.5 Spectral Efficiency based on Shannon Analysis

To come to a more general view of the potential gains of multi-hop (MH) communication the Shannon capacity can be used as upper bound for the achievable throughput on a single link for a given signal-to-noise ratio (SNR) and for a given bandwidth, W [141]. This analysis is presented in Appendix 7.2.1.3.

Shannon Capacity with Frequency Re-use

In the following Figure 4-57 the resulting transmit-power-to-noise ratio (TNR) for the MH connection with and without frequency reuse over the respective TNR for a one-hop connection is shown. The capacity breakeven is indicated by the line through the origin with gradient one. At the respective points both approaches, multi-hop and one-hop, require the same transmit power. Curves above this line indicate that one-hop connections do need less power than MH connections, whereas curves below indicate scenarios in which MH connections should be favored.

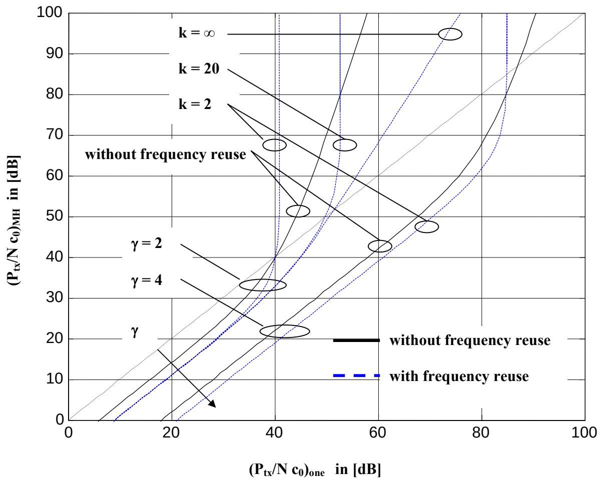


Figure 4-57: Required transmit-power-to-noise ratio for four-hop connections as function of the TNR for an one-hop connection with and without frequency reuse at 5 GHz for d = 100, N = -80 dBm

It becomes evident that the gain for a MH connection strongly depends on the transmit power, P_{tx}, and the attenuation exponent, γ. The smaller the transmit power, and the stronger the attenuation, the more attractive becomes the introduction of intermediate hops. It can further be seen that with increasing interference reduction, i.e. with increasing value *k*, the breakeven point is shifted to larger TNRs when introducing frequency reuse. Comparing the gain with and without frequency reuse, there is a constant offset of about 3 dB as long as the TNR is low enough. Beyond some points the required transmit power rapidly increases and there cannot be achieved any gains in increasing the transmit power even more because of self-interference. Nevertheless, high gains can be obtained under high attenuation, e.g. approx. 10 dB (20 dB) for γ = 2 (γ = 4) and n = 4 hops.

4.2.5.2.6 Shannon Capacity comprising Overhead

For realistic and fair capacity estimation the protocol, respectively physical overhead is taken into account in the next step [142]. The latter impact can be assessed by the pessimistic assumption that each transmission requires some fixed overhead, *ovh*, and *n* transmissions need *n*-times that overhead. In this case the exponent *n* in equation (95) becomes $n \cdot (n \cdot ovh + 1)/(ovh + 1)$. The relation of the TNR for the MH connection to the TNR for the one-hop connection as function of the introduced number of hops without considering frequency reuse is shown in Figure 4-58.

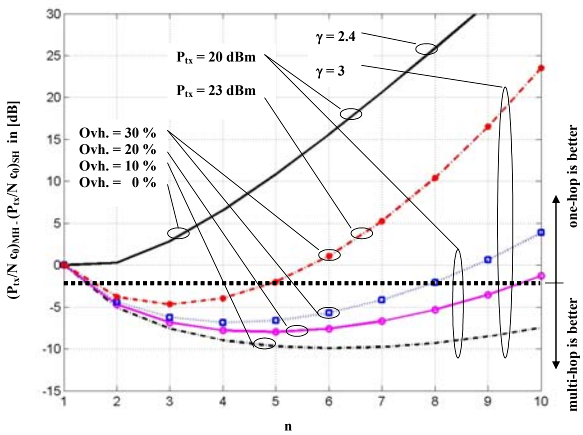


Figure 4-58. Relation of transmit-power-to-noise ratio for MH and one-hop connections as function of the number of hops, d = 100 m

For values below 0 dB the MH connection needs less power than the one-hop connection, whereas values above indicate scenarios in which a one-hop connection should be favored.

It becomes evident that the gain for a MH connection strongly depends on the transmit power, P_{tx}, the distance between source and destination, d, the attenuation exponent, γ, and the number of hops, denoted by n. The smaller the transmit power, the larger the distance, and the stronger the attenuation, the more attractive becomes the introduction of intermediate hops. Typical transmit powers of P_{tx} = 20 dBm (23 dBm), a distance of 100 m, attenuation exponents of γ = 2.4 (3), and different values for the overhead between 0% and 30% have been chosen. It can be seen that with increasing number of hops the gain for relaying is increasing up to a point where the overhead is dominating. E.g., for P_{tx} = 20 dBm and γ = 3 the highest gains can be obtained for a number of 5 and 4 hops for overhead values of 20 % and 30 %, respectively. For better link conditions, respectively higher transmit power (P_{tx} = 23 dBm) the highest gain can be achieved for 3 hops only taking into account an overhead of 30%. However, the most dominant factor is the attenuation exponent, which is expected to decrease to the free-space value with decreasing distance on the one-hop connection. In this case even for no overhead no benefits are expected from relaying for small transmit power of P_{tx} = 20 dBm and γ = 2.4.

All the aforementioned facts and results lead us to the conclusion that a number of 4 hops, which is called oligohop, should not be exceeded to benefit from relaying.

4.2.5.2.7 Conclusions and Future Work

Using Jensen's inequality, we have shown that for high signal to noise values, a single hop has batter spectral efficiency compared with a n-hop replacement; for small SNR values, a more efficient n-hop link is possible, with the optimal locations of multi-hop regenerative relays being at equal intervals along the straight line between the source and destination. A novel quantitative criterion is developed, which by the inequality given in (13), offers threshold mean SNR values below which a n-hop replacement should be considered over a single-hop link. Additional research into the statistical distribution of spectral efficiencies for multi-hop links may bring further clarifications on the properties of cellular systems using fixed relays for multi-hop communications.

4.2.5.3 Flexible Extension of the Infrastructure-based Network via Mobile Relays

A brief introduction to the different categories of routing schemes as well as commonly known examples for each category are presented in the Appendix 7.2.2. Taking into account the advantages and deficiencies of the aforementioned routing protocols motivated for the development of a hybrid routing scheme that combines the characteristics of on-demand and position-based routing algorithms to come to a more flexible and efficient scheme avoiding the drawbacks of both at the same time. This new location-aware on-demand routing scheme is described in Section 4.2.5.3.1. A comparison of the routing overhead for the on-demand routing scheme and the new hybrid scheme with the help of simplified analytical formulas is presented in Section 0.

4.2.5.3.1 Location-Aided AODV

From the previous description of AODV it becomes obvious that route breaks can have sever impacts on the system performance, especially in dynamic network topology where route breaks happen frequently. Furthermore, the local route-repair in AODV is not optimal, since it introduces delays and overhead owing to the repair process.

Motivated by the good performance of geographical routing schemes in highly mobile environments, it is proposed to combine both routing schemes, i.e., on-demand topology-based routing with location-based stateless routing. Location-based routing, also referred to as geographical routing, does not require route maintenance, since it determines the next hop towards the destination on-the-fly based on the location of the destination and the neighbours. If a station discovers a neighbour closer to the final destination it forwards the packet to it, leaving the further delivery up to that station. To select the next neighbour towards the destination, different methods can be used, e.g., greedy forwarding, restricted directional flooding, or hierarchical schemes. We have chosen greedy forwarding since this is a simple and efficient method with knowing only the location of the neighbours and the final destination. The combination of AODV and location-based forwarding for efficient route maintenance is called Location-aware AODV (L-AODV) in the sequel.

Different to LAR [139], which is used for the route discovery process, we concentrate on the route maintenance. In addition we use greedy-forwarding instead of directional flooding. In [140] it is also proposed to enhance AODV with geographical information. But the purpose is to realize geo-casting and not to improve route maintenance.

Detailed Description of L-AODV

The basic operation of L-AODV is described in the following Figure 4-59.

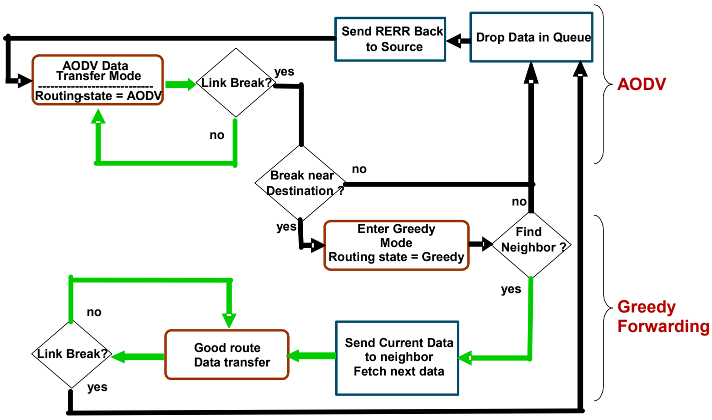


Figure 4-59 Flow-chart of Location-aware AODV

The main steps of L-AODV are the same as in conventional AODV, i.e. after route discovery and setup the packets are transferred to the destination. But when a link breaks close to the destination, L-AODV automatically switches to greedy-forwarding mode. In this mode it first checks whether there is a neighbour with a location closer to the destination. If this is true it sends the packets in the future to that neighbour and updates its routing table respectively. If the selected neighbour has no route to the destination in its cache, it also uses greedy forwarding mode. However, if the link breaks in greedy-forwarding mode, i.e. the next-hop neighbour previously selected via greedy-forwarding is not in communication range, the station sends an RERR packet back to the source and drops all packets for the respective destination in its queue. The source starts then again in AODV mode with a route discovery.

Changes in AODV to Use L-AODV

From the previous description of L-AODV two main questions arise: First, how does a station get to know the location of the destination, and second, how does it get to know the location of the neighbours?

The location of the destination is simply added in the RREP packet that is send back to the source and all intermediate stations cache this information. The absolute location is encoded in latitude and longitude, which are each 4 byte in length. The location of a station is derived, e.g., from the Global Position System (GPS).

The next challenge is to derive the locations of the neighbours to determine the closest neighbour to the destination, i.e., the station to forward the packet to in case of a link break in AODV mode. Two approaches are possible. In the first approach the location of a station is included in the HELLO packets that are used to check the availability of the next-hop neighbour in AODV maintenance. The second approach is based on reactive beacon packets that are sent out by the station that has recognized a route break. As response to this beacon each neighbour reports its current location back to that station by means of so called position-beacons (pos-beacon). It has been found out via simulations that the second approach based on on-demand beacon packets is more efficient than adding the 8 byte location information in every HELLO packet. Hence, only the on-demand beacon approach is used to gather the location information of the neighbours and to determine the candidate station for greedy-forwarding.

Performance Estimation of L-AODV

Based on the L-AODV approach with beacons in this section the analytical performance is roughly estimated to become a better insight into the pros and cons of the new approach. For an arbitrarily selected scenario with five stations the required routing overhead is shown in Figure 4-60.

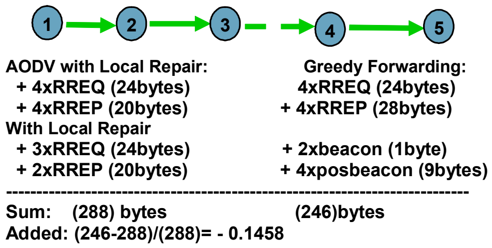


Figure 4-60 Routing overhead for AODV and L-AODV

The overhead needed to discovery the route is almost identical for AODV and L-AODV, except the RREP packets are 8 byte longer in L-AODV because of the additional location information.

In case of a route break between station 3 and 4 AODV needs three RREQ packets, one by station 3 and a retransmission by station 2. If we assume that another station does not know an alternative route to station 5, say station 7 not shown in Fig.3TODORef, this station also sends out a RREQ packet. Then station 5 and station 7 send back an RREP to station 3. In case of L-AODV station 3 sends one beacon packet, which is responded by all neighbours. The new station 7, which is assumed to be closer to the destination 5, responds with a pos-beacon. When station 7 receives the next data packet and it has not an active route to the destination, it also sends out a beacon, which is responded with two pos-beacon from station 3 and station 5, ending up in a total of two beacons and four pos-beacons for this example. Comparing the routing overhead for this example L-AODV needs approximately 14% less than AODV.

Moreover, in combination with location-aided routing (LAR) [139] to improve the route discovery, L-AODV is expected to be a promising candidate for further investigations. In addition, the integration of infrastructure components with fixed positions will further improve the performance of the proposed hybrid routing scheme.

However, for a realistic and fair performance comparison event-driven simulations will be performed in the future.

4.2.5.4 Wireless Fixed Relays routing (WiFR)

In this section we present an algorithm suitable for Fixed Relay Networks (FRNs) (see 4.2.2.2). In such network support of multi-hop is mandatory and efficient routing algorithms are required to exploit at best network resources and to efficiently manage internal connections and load from and to the core network resources. Even if FRNs are multi-hop wireless networks like MANETs, their characteristics make the routing problem quite different [143]. FRNs have a defined topology and topology changes are mainly due to node failures, so can be considered not frequent. Therefore, the distribution of network state information is not much costlier than in wired networks, and even a centralized control of route selection can be adopted [144]. Finally, energy consumption is not a problem for network nodes.

The new QoS routing algorithm proposed is named "Wireless Fixed Relay Routing" (WiFR) and it presents a solution (based on an ad-hoc heuristic) for a new model for the QoS routing problem in multi-hop wireless networks explained in section 7.2.3 of Annes II. The model is an extension of the well known multi-commodity flow problems [145] where link capacity constraints are replaced with new ones that takes into account interference constraints among different radio links. It guarantees that the rates of routed flows are compatible with radio channel capacity, but does not require to explicitly solve the scheduling problem. Hence the new approach proposed results in the complete separation between construction of route that satisfy the required QoS and the definition of proper scheduling at MAC level. This means it could be even applied to MAC that does not support strict resource control, like e.g. IEEE 802.11, taking a proper margin on the radio link capacity in order to take into accounts all overheads due to the protocol and the contention on the channel. In this case if strict control on the flows entering the network is enforced, the new approach assures that no persistent congestion occur in the network.

In addition we show how it is possible to derive an admissible scheduling (TDMA based) of packet transmissions on the paths provided by the routing algorithm. Finally, in the last part of this section, results on the performance of the proposed approach is presented.

4.2.5.4.1 WiFR: a new QoS routing algorithm for FRN

The new Wireless Fixed Relay routing (WiFR) algorithm proposed to solve the problem presented in section 7.2.3 of Annes II has the following main features:

- centralized algorithm, route computation performed in a central entity
- route choice based on precise knowledge of global network state and accurate global information
- flow-by-flow bases; i.e. packet belonging to different flows are treated separately even if the flows have the same source and/or the same destination
- end-to-end bases; i.e. our algorithms works considering source-destination parameter of all flows in the network and not only taking greedy decision hop-by-hop
- automatic load balancing through the network
- fully functionalities with any lower layer technologies, even if we propose a on purpose TDMA mac layer, optimized for FRN that shows to raise network performance

Suppose now to be in a certain network state characterized by a given number of routed flows, even zero; central entity maintains precise information about the global network state using the following data structures that are updated each time a new connection is admitted in the FRN or when an existing one is rerouted or stops to flow.

i) Topology matrix $\underline{\underline{T}}$ is a $N \times N$ matrix with $N = |V|$, i.e. number of relays constituting the FRN, where the generic element $t_{j,k}$ of matrix $\underline{\underline{T}}$ is set as follow:

$$t_{j,k} = \begin{cases} 1 \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E \end{cases} \tag{14}$$

Network is represented here as a mono directional graph where for each relay outgoing links are distinguished from incoming ones separating in this way transmission capacity from reception capacity and allowing a better use of network resources.

ii) The set of relays neighbours of given relay j, i.e. that can be directly reached by j 's transmission and the set of relays that are two hops away from given relay j, i.e. the set of relays that have a common neighbour with j. Respectively:

$$H_1(j) = \{k \in V | (j,k) \in E\} \tag{15}$$

$$H_2(j) = \{l \in V | (k,l) \in E \wedge k \in H_1(j)\} \tag{16}$$

iii) The residual capacity matrix $\underline{\underline{RC}}$ that is a $N \times N$ matrix where $N = |V|$ is the number of transceiver in the network. The generic element $rc_{j,k}$ of the matrix represents the residual capacity on radio link j to k; with term residual capacity is meant how fraction of bandwidth B, provided by lower layers could be transmitted from relay j to relay k without having overload of some relay and avoiding collision with other signals. Thus formalizing the generic element of the residual capacity matrix could assume the following value:

$$rc_{j,k} = \begin{cases} r \subset [0,1] \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E \end{cases} \tag{17}$$

if $rc_{j,k} = r$ it means that $r \cdot B$ bit/sec can be transmitted from j to k. It is assumed that the bandwidth provided by lower layers is a known parameter at routing level and therefore represents an input of our algorithm.

iv) Weight matrix $\underline{\underline{W}}$ is a $N \times N$ matrix with $N = |V|$, i.e. number of relays constituting the FRN, which is used to store the weight associated to each existing radio link in the network. This weights actually are the metric used in route computation. The generic element $w_{j,k}$ of the matrix can assume the following value:

$$w_{j,k} = \begin{cases} w \subset [1,\infty) \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E \end{cases} \tag{18}$$

To define numerically this new metric we have introduced firstly a new matrix $\underline{\underline{S}}$ whose generic element $s_{j,k}$ is a gauge of saturation level of radio link from j to k. The generic elements of this new matrix is set as follow:

$$s_{j,k} = \begin{cases} 1 - r_{j,k} \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E \end{cases} \tag{19}$$

Being $rc_{j,k} \in [0,1]$ also the generic element of matrix $\underline{\underline{S}}$ will be in $[0,1]$ with the following meaning:

$s_{j,k} = 0 \rightarrow (j,k) \in E \quad is \quad unloaded$

$s_{j,k} = 1 \rightarrow (j,k) \in E \quad is \quad fully \quad loaded$

We want our algorithm to avoid saturation of one or more links hence to traduce mathematically this constraint we have introduced a penalty function $PF(\underline{\underline{S}})$ of the network. A function, to be defined as a penalty function, must have the property of diverging to infinity as soon as even only one constraint it embeds is violated. Penalty function is defined as:

$$PF(\underline{\underline{S}}) = \sum_{(j,k)\in E} \left(\frac{1}{1 - s_{j,k}} \right) \tag{20}$$

Since $s_{j,k} \in [0,1]$, even a single link fully saturated causes the penalty function to diverge; besides penalty function is also a gauge of overall network condition, in fact the higher its value, the heavier network is loaded. Our algorithm set up routes trying to keep as low as possible the penalty function, i.e. trying to distribute as uniform as possible traffic in the network avoiding congestion. Weights are defined as follow:

$$w_{j,k} = \begin{cases} \dfrac{\partial PS(\underline{\underline{S}})}{\partial s_{j,k}} = \left(\dfrac{1}{1 - s_{j,k}} \right)^2 \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E \end{cases} \tag{21}$$

It can be notice that in the beginning when no connections flow in the network, is $w_{j,k} = 1 \forall (j,k) \in E$. Metric introduced allow to assign deeply differentiated weights to near-fully loaded links and to low loaded links; further, link's weight tends to increase more quickly as link approaches saturation and so the higher is the weight the nearer we are to saturation of a network portion and the faster we are moving towards it. Using of such metric allow WiFR to automatically balance load in the FRN and to keep end-to-end delay as low as possible.

v) Status tab $\underline{st}$ is a list of record that has $N = |V|$ entries, one for each relay in the network used to stored information on relay state. For a given relay j the parameters stored in its record are:

$$usedband(j) = ub \subset [0,1]$$
$$freerx(j) = fr \subset [0,1]$$
$$freetx(j) = ft \subset [0,1]$$

These parameters represent respectively the fraction of provided bandwidth B, normalized to 1, that relay j "sees" as yet consumed either for its own transmissions or for receptions of other signals both addressed to it and both not addressed to it; the fraction of provided bandwidth B, normalized to 1, that relay j has still free to receive without having collision with its own transmissions or with other received signals, this parameter depends only on what happens in relay j itself and in the set $H_1(j)$; last parameter is the fraction of provided bandwidth B, normalized to 1, that relay j has still free to transmit without causing collision with other relays transmissions and respecting all the constraints introduced in the mathematical model, this parameter depends not only on what happens in relay j itself and in the set $H_1(j)$, but also on what happens in the set $H_2(j)$.

Suppose now that a route request from relay s to relay t with a request bandwidth $\tilde{b}$ (normalized to the provided bandwidth B) is delivered to the central entity. The central entity applies the route searching routine that selects, just among the feasible paths between source node and destination one, the path that has the minor impact on network global saturation level according to the metric introduced previously. We use as basis for our route searching routine the mechanism of Dijkstra 's algorithm, where the metric used is the one described previously, but we have modified it in order to search the route just among the feasible paths and in a greedy manner, so that the optimal route is selected avoiding the exhaustive exploration of all existing feasible paths. This can be done thanks to the following exploring routine that we have developed to determinate if a potential next hop relay, say k, should be explored by Dijkstra 's algorithm or not. Following are fall-through controls executed by routine as we have implemented it; j is supposed to be the relay actually selected by the routine:

1. (Preliminary controls, only if j is the source). If j is the source node then it is immediately discarded and connection refused if it has not a $freetx(j) < \tilde{b}$.
    a. Otherwise, if j is the source node and k is not destination node, this means that node k will transmit either to destination or to a next hop toward destination. So it must be controlled that j can receive the signal retransmitted by node k without violating any constraints. This means that j should not have $usedband(j) + 2\tilde{b} > 1$; this latter condition derives from necessity that j should receive the signal retransmitted by node k, that consumes a fraction $\tilde{b}$ of its still not in use capacity, after that j itself has transmitted the signal thus having yet consumed a fraction $\tilde{b}$. If the last equation is true connection is refused.

2. If connection is not refused preliminarily, then the routine controls if node k is the destination node, if so node k could be selected since node j can surely transmit directly to destination having passed all controls before being selected.

3. Routine arrives at this step if and only if j is not the source node and k is not the destination node. If j is not the source node, then it can surely receive signal from its previous hop, transmit to its next hop and receive possible next hop 's retransmission, all this since node j has been yet selected by Dijkstra 's algorithm and hence has yet fallen through previous and following controls. Besides if routine as arrived till this control node k is different from destination, otherwise it will be selected in control 2).
    b. In this case a preliminary condition that must stand not to discard immediately node k as potential next Dijkstra 's selection, is that node k can forward packets of the new connection. Therefore if node k has $freetx(k) < \tilde{b}$ then it is immediately discard since it won't useless to make all other controls. If node k is not discarded here it means that potentially it could forward packets of the new connection but other controls must be passed before selecting it.
    c. Then routine controls if one or more common neighbour exists for node j and node k, such common neighbour will receive transmission both of j and of k (that, remember, is not the destination node). If such common neighbours cannot receive both transmissions without violating constraints, then node k is discarded.
    d. If node k has not been discarded before, then routine controls if destination node t is a neighbour of k or not. If t is a neighbour then it is controlled that k could receive signal from its previous hop j and that can retransmit to destination. So if

$usedband(k)+2\tilde{b} \leq 1$ then node k is selected, otherwise discarded. If destination node t is not a neighbour of node k, this means that node k must be able to receive from its previous hop j, transmit to a next hop different from destination and finally receive retransmission of its next hop. So if $usedband(k)+3\tilde{b} \leq 1$ then node k is selected, otherwise discarded. To notice that these conditions are more restrictive than control 3a).

If the routine states that node k can support the new flow then it is explored by Dijkstra 's algorithm as in the usual version, i.e. relay k is labeled with a temporary mark that indicates overall distance from the source node calculated using the metric we have introduced and the id of node from which it has been reached. If the routine instead states that relay k is not able to support the new flow no further actions are taken on relay k. In both cases, after having executed proper actions basing on routine 's response, a new neighbour of the actual selected node j is chosen, if exist, and routine is called again. Once that all neighbour of j has been controlled by exploring routine and, if necessary, explored by Dijkstra 's algorithm, then the algorithm selected the new node to be marked as definitive exactly as in the usual version of Dijkstra 's algorithm and on this node the algorithm reiterates itself. All these actions are repeated until destination is reached or until no node can be marked as definitive and so Dijkstra 's algorithm cannot iterate and stops. If a feasible path is found, central entity reads the sequence of relays forming the path $\underline{p}$ and performs the following operations in order to update the information about the global network state:

$$usedband(j)=usedband(j)+\tilde{b} \leftrightarrow j \in \underline{p} \wedge j \neq t \tag{22}$$

$$usedband(k)=usedband(k)+\tilde{b} \leftrightarrow k \in H_1(j) \big| j \in \underline{p} \wedge j \neq t \tag{23}$$

$$freerx(j)=1-usedband(j) \forall j \in V \tag{24}$$

$$freetx(j)=1-\max\left\{\max_{k \in H_1(j)}\{usedband(k)\};usedband(j)\right\} \tag{25}$$

Equation (22) represent resources consumption of relay forming the new path for transmitting the new flow. Equation (23) represent resource consumption due to the physical broadcast: all neighbours of transmitting relay receive the signal. Equation (24) updates the residual capacity for receptions of each realy, this parameter depend just on one hop cluster of the relay. Equation (25) updates the residual capacity for transmission of each relay, this parameter consider both one hop cluster and two hop cluster of the relay.

Then central entity sets value of other data structures so that they reflect the actual network condition and are ready to be used in next call to route computation:

$$rc_{j,k}=\begin{cases}\min\{freetx(j);freerx(k)\} \leftrightarrow (j,k) \in E \\ 0 \leftrightarrow (j,k) \notin E\end{cases} \tag{26}$$

At this point all elements are here to update the weight matrix $\underline{\underline{W}}$ following equation (19), (20) and (21).

Having proposed an heuristic QoS routing algorithm that, for its inner nature, finds a sub-optimal solution, implies that when no feasible path is found for a new flow, it is no necessary true that network capacity is insufficient to support the new flow, but it could exist a different sub-optimal solution that leaves room also for the new flow. Starting from this assumption, we have developed the following optimization routine. Central entity selects, according to specified criterion one of the yet routed flow and tries to reroute it along a different feasible path using the same route searching routine described above and obviously always keeping respecting all algorithm 's constraints and maintaining thus the QoS required by all flows in the FRN. If it isn't possible to reroute the selected flow optimization routine marks the selected flow as yet optimized and selects a new flow on which reiterate itself. If rerouting is possible then optimization routine tries to admit the new flow, if this is possible then rerouting is effectively performed and new flow admitted, otherwise optimization routine reiterates itself. To notice that no rerouting is effectively executed till the new flow may be admitted in the FRN.

4.2.5.4.2 Optimized Scheduling

We have presented a new mathematical model (see 7.2.3 in Annes II) in which the new constraints embedded in it allow our QoS routing algorithm to be completely independent from the particular mac layer effectively used in the network and not to need any information about scheduling if it is done. This is a great advantage also for a possible commercial use of our algorithm since no technology constraints must be considered, but it could be used upon any kind of mac and physical layer without any adjustment or modification. Further, even if no information about scheduling are needed, our algorithm allow existence of a proper scheduling that support the routing scheme for relay transmissions at mac layer, if this latter uses some form of slot as time slot, channel assignment or code assignment.

By the way, it has seemed to us that offering Quality of Service at routing level and then coupling this with a non QoS mac layer as one from 802.11 family is a very sub-optimal choice since many advantages conquered through routing algorithm are lost because not supported by mac layer reducing in this way the overall network performances. Following this consideration we have decided to develop a TDMA (Time Division Multiple Access) mac layer optimized for using in FRN with WiFR as routing algorithm.

The problem of slot assignment has been modelized as a variation of the well known Minimum Order Frequency Assignment Problem (MO-FAP) where a maximum given number N_S of slot may form the mac frame and each relay requires a given number of time slot according to the capacity it needs for transmission. The same slot is assigned to all relays whose transmission are not conflicting. The assignment is resolved through a heuristic algorithm of graph multi-coloring that keeps minimal the overall number of slot used so to keep as short as possible the mac frame and the end to end delays, too.

Since to each relay is always assigned a integer number of slot, bandwidth is assigned to relay in multiples of the minimum bandwidth quantum corresponding to bandwidth associated to one single slot. To have routing algorithm work properly guaranteeing QoS to all flows admitted in the FRN and in the meanwhile allowing the existence of a proper scheduling the TDMA mac we propose, it is necessary to transpose the fraction of provided bandwidth required by flow f from $\widetilde{b}_f$ to

$$b_f^* = N_{a,f} \cdot \frac{1}{N_S} \tag{27}$$

with

$$N_{a,f} = \left\lceil \frac{\widetilde{b}_f}{1/N_s} \right\rceil = \left\lceil \widetilde{b}_f \cdot N_S \right\rceil \tag{28}$$

The value $N_{a,f}$ is exactly the number of slot assigned to the relay for the transmission of flow f. So the overall number of time slot assigned to the given relay transmitting N_f flows can be expressed as:

$$N_a = \sum_{f=1}^{N_f} N_{a,f} = \sum_{f=1}^{N_f} \left\lceil \widetilde{b}_f \cdot N_S \right\rceil \tag{29}$$

Hence, the overall fraction of provided bandwidth and the overall bandwidth reserved for this relay can be expressed as follow:

$$B_a^* = N_a \cdot \frac{1}{N_S} \tag{30}$$

$$B_a = N_a \cdot \frac{1}{N_S} \cdot B \ \left[bit/\mathrm{sec} \right] \tag{31}$$

It may seems too conservative and incorrect to transpose the single flow 's requested bandwidth to a multiple of the bandwidth quantum, instead of transpose the aggregate relay needed bandwidth for transmission of all the N_f flow. In reality, if the latter possibility is followed, is no more possible to guarantee that for each solution to routing problem adopted by WiFR exists also a proper scheduling at mac layer. This is to due to the different bandwidth consumption that would be seen at routing level and at mac level for the same connection.

4.2.5.4.3 WiFR flow chart

In this paragraph is reported the flow chart illustrating how the different routines interact one with the others and how the routing algorithm accesses to external input it needs and to information stored in its internal structures. To notice that blocks and information regarding mac layer are effectively performed only either when our optimized TDMA mac layer or another slotted mac layer is used in the FRN.

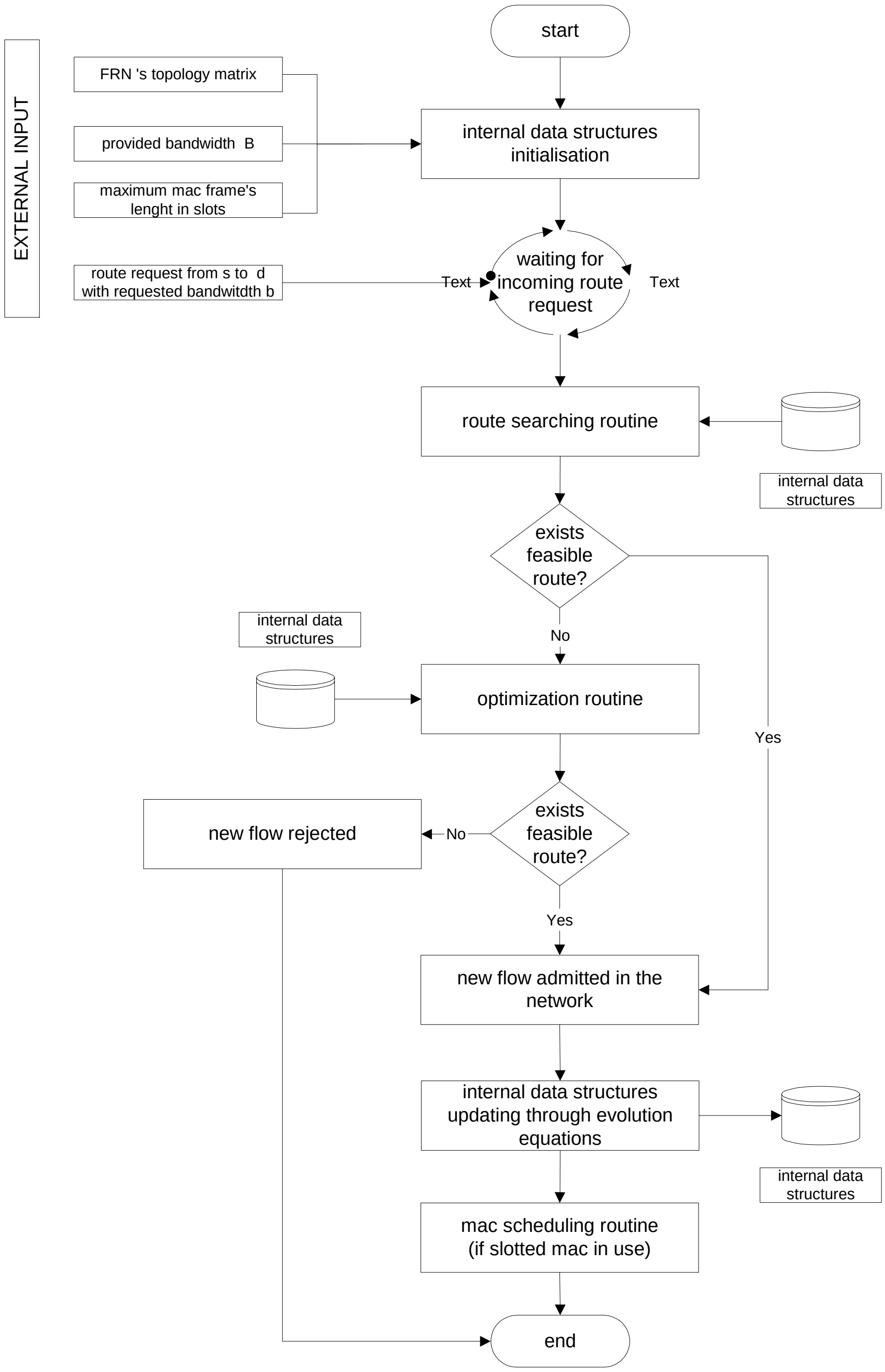


Figure 4-61 WiFR flow chart

4.2.5.4.4 Simulation results

To evaluate the new model and WiFR algorithm presented above, a benchmark QoS routing algorithm based on pure hop-count metric named WSPF (Wireless Shortest Path First) and the optimized TDMA mac layer as been added into the event-driven network simulator Ns2 [118]. Simulations presented here has been conducted using two ray (ground reflection) channel model, a provided bandwidth of 2 Mbit/sec, packets of 1 kbyte, exponential On/Off traffic sources with differentiated bandwidth request resembling voice traffic.

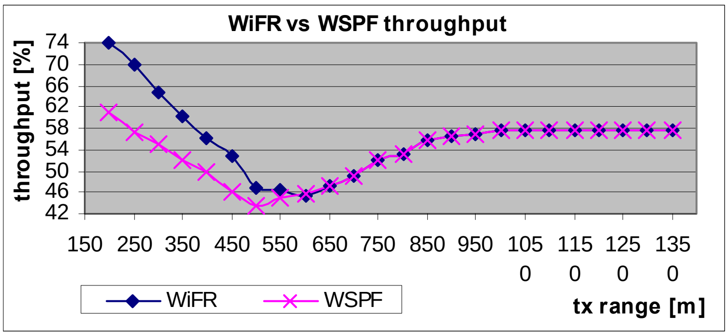


Figure 4-62 WiFR vs WSPF, random topologies, 60 relays, high traffic

Figure 4-62 shows throughput obtained in a FRN of 60 relays random deployed following a uniform distribution over a 1000m X 1000m area for various values of relays radio range. For highest values of radio range the network become fully connected (each source reaches directly its destination) and both algorithm gives the same throughput; instead it is clear that for a network with low/medium connectivity degree, our routing algorithm not only outperform the WSPF of about 13%, but also rises network capacity of about 19% with respect to fully connected situation. As the radio range increases, the throughout firstly is reduced till a minimum because every single transmission impact on a set G_i greater and greater rising resources consumption; but meanwhile, increasing radio range means to diminish the route length and hence the re-transmissions, for this reason after the minimum the throughput rises till settling to the final plateau. Results proof that the metric introduced is really efficient and able to distribute load in the FRN and that mesh topology may increase overall network performances if resources are efficiently managed. WiFR performance has been deeply investigated also in Manhattan topologies, see Figure 4-63, starting from a basic grid of 4x4 relay used as sources and/or destinations of the connections and adding a number n, from n=0 (basic grid) to n=5, of additional relays that has only the task of forwarding packets.

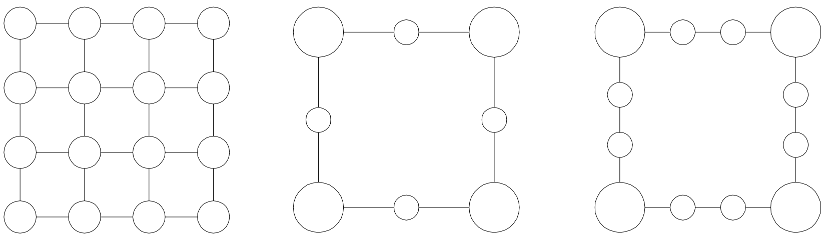


Figure 4-63 Manhattan basic grid (n=0) and particular for n=2 and n=5 topologies

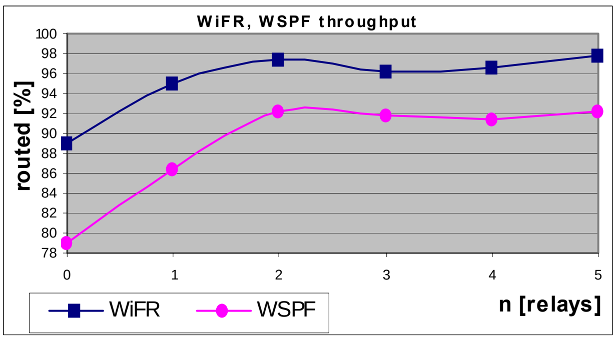


Figure 4-64 WiFR vs WSPF, Manhattan topologies, low traffic

Figure 4-64 shows that WiFR outperforms WSPF in all Manhattan topologies from 6% to 10%; further, impact of additional allows to increase throughput thanks to possibility of increase spatial reuse of shared radio resource, with only one additional relay capacity gain is about 16%, with two additional relays gain is about 20%. It can be noticed that with three and four additional relays throughput decreases since route becomes longer and re-transmissions negative effect are not completely balanced by higher spatial reuse.

It seems hence that one or two additional relays is the optimal compromise between performance and costs. We have also compared the performance achievable using different mac layer, in particular Figure 4-65 shows the relative throughput, i.e. the percentage of admitted load that is correctly delivered, obtained using our apposite TDMA mac layer and the IEEE 802.11 mac layer with RTS/CTS/Data/ACK mechanism.

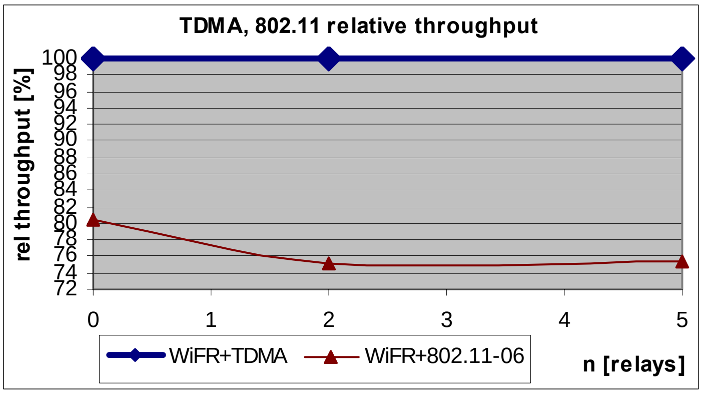


Figure 4-65 TDMA vs 802.11 relative throughputs, Manhattan topologies

To notice that in simulations with IEEE 802.11, the overall load admitted in the FRN has been kept lower of 20%, so it works in better conditions that TDMA mac layer. It can be seen as a centralized access mechanism as our TDMA mac layer gives no loss whereas the packet loss using the 802.11 mac layer, due only to the distributed way of accessing, are from 20% to 25%. It is clear that non-slotted mac layer as 802.11 are deeply inadequate to support the QoS guaranteed by WiFR, since many advantages gained by routing algorithm get lost during channel accessing.

WiFR has shown to be able to kept very stable network conditions even with an extremely varying offered load. Figure 4-66 shows end-to-end delay achieved when connection requests follow a Poisson process with different inter-arrival time. It can be seen that as the frequency or connection request arrival increase, the end-to-end delay remains stable with slight variation of about 50msec on an average delay of 0,1-0,2 sec. This proofs that WiFR is really able to offer QoS maintaining stable the network condition independently from external offered load. Finally Figure 4-67 shows comparison between WiFR and ad-hoc routing protocols implemented in Ns2, i.e. DSR, DSDV and AODV. Since the ad-hoc routing algorithms do not offer QoS admitting all offered load in the FRN and are not adapt to work with our new TDMA mac layer, simulations has been done with a low traffic level so that even WiFR admit it all in the FRN and using IEEE 802.11 as mac layer.

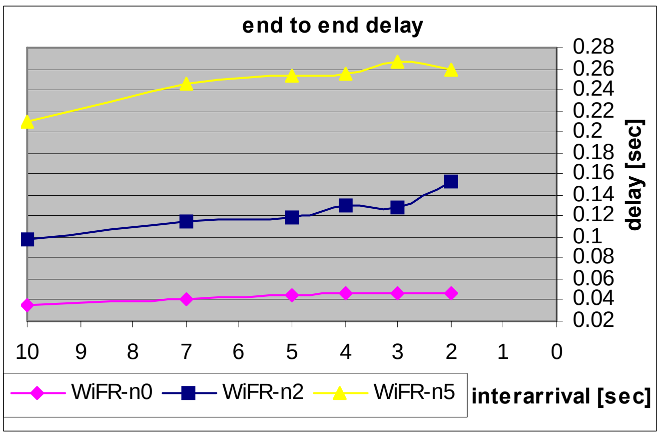


Figure 4-66 WiFR, end to end delay with Poisson arrival of connection request

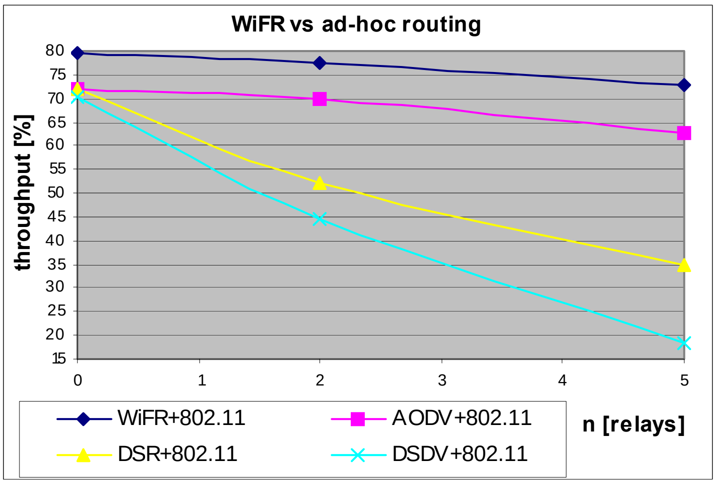


Figure 4-67 WiFR vs. ad-hoc routing protocols

WiFR outperforms all ad-hoc routing protocols of at least 10% with a maximum of 55% with respect to DSDV and Manhattan topology with five additional relays. This enforces our idea that FRN shouldn't be considered as a particular case of ad-hoc network, but new routing algorithms should be developed considering the new peculiarity of such scenario.

4.2.5.4.5 Conclusions

From the exhaustive set of simulations conduced to validate and test Wireless Fixed Relay routing algorithm, the following conclusions can be taken out.

About random generated topology, WiFR has shown to give advantages especially in network of medium, large dimensions, in terms of number of nodes, where is able to enhance overall network performances admitting a greater amount of traffic with low network connectivity with respect to a fully connected network. Further, a trade-off related to network connectivity between possibility of finding several possible feasible routes and number of relays influenced by a each single transmission has been highlighted and it has been shown that best results are achieved when network connectivity is low but allows anyway to different feasible routes for a given connection.

About Manhattan topology, simulations have shown that, especially for medium and high traffic matrixes, best results are achieved when two additional relays are inserted between two neighbour basic relays that act as source and/or destination of connections.

WiFR has proofed that working in pre-planning modality, is able to support very well both CBR traffic and exponential traffic sources generating packet bursts even when each connection requires a bandwidth equal to its average rate bit always offering Quality of Service in a high satisfactory way.

WiFR has shown to support very well also dynamic traffic where route requested are presented to the network according to a Poisson process and with randomly generated source and destination. QoS is always preserved also under very heavy traffic and end to end delay are very stable under different network conditions.

WiFR has always outperformed WSPF with respect to number of connections admitted in the network thus proofing validity of route searching routine and metric developed on purpose for WiFR.

Finally, it has been shown that IEEE 802.11 isn't a efficient choice since it loses a great part of advantages gained by WiFR with respect to Quality of Service even when it works in better conditions than our on purpose TDMA mac layer that instead has shown to be optimal for preserving Quality of Service offered by a Fixed Relay Network.

4.2.5.5 Routing algorithm

Traditional routing algorithms are designed to provide best-effort connections between end users, on the basis of a single metric only. Typically the hop count parameter is used for route selection, whereas no Quality of Service issue is taken into account. This approach is often not optimum and results in performance degradation when for example multimedia applications are involved. The reason is that on the one hand the selected path may not be able to meet the requirements of a specific flow and on the other hand there may still exist another path with sufficient resources, which was not chosen because the routing decision was made without any awareness of resource availability and requirements [147]. The necessity for the QoS routing seems to be obvious in this situation, but one should remember that finding a route subject to multiple constraints is in general an NP-complete problem [148]. Nevertheless there exist a class of heuristic algorithms which make the QoS routing computationally feasible.

4.2.5.5.1 Motivation for Quality of Service routing in WINNER

The radio network architecture of a future radio system, as in the scope of WINNER, may be based on fixed and mobile/movable nodes with different capabilities [147], which results in special requirements for the routing protocol. First of all, as it was mentioned in the introduction, the modern multimedia services demand strict Quality of Service levels to be guaranteed. The QoS may be regarded as the qualitatively or quantitatively defined performance contract between the service provider and the user applications [150] As a result the routing protocol should be based on an algorithm that is capable of choosing different paths connecting the source to the destination, so that distinct QoS levels could be offered to some specific applications. Secondly there should be special features embedded in that protocol to allow the node to decide whether it is taking part in relaying or not. This refers mostly to the movable nodes and is connected with the network topology management. For example when a battery powered MT node detects low power level, it may decide to no longer route packets coming from its neighbours. The last but not least issue is connected with the micro- and macromobility management, as it is indeed a crucial problem to manage the micromobility without significant control overhead [151]. In the sequel a solution will be presented, which meets all those requirements.

4.2.5.5.2 Multi-constrained Quality of Service routing algorithm

QoS routing is based on the resource reservation process that consists of the two following steps: finding resources and possibly making reservations [149][150]. Before making a reservation, the routing algorithm must select the appropriate path which meets the requirements given. To this end multiple metrics such as bandwidth, delay, loss probability, hop count, link load or other should be considered.

The network may be modeled as a weighted, directed graph $G = (V, E)$, where the vertices V represent the network nodes, and the edges E correspond to the communication links. Each of many possible weights of an edge is equivalent to one particular metric of a link. There may not exist a link either in one of or in both directions if there is no physical radio link available between any two radio interfaces. An example network graph is depicted in Figure 4-68.

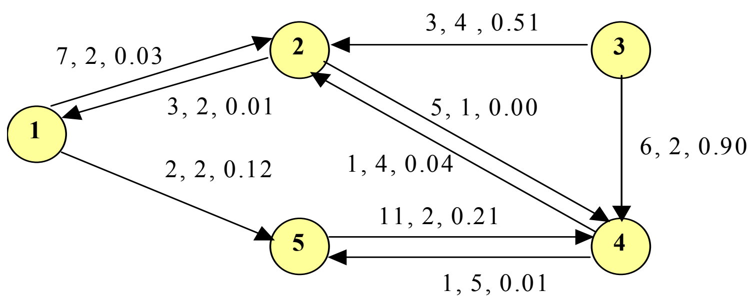


Figure 4-68: An example network graph

In this example situation there are three weights next to each edge corresponding to the bandwidth, delay and loss probability metrics respectively. For the purposes of QoS provisioning the routing algorithm must take all of them into account when selecting the best paths between any two distinct nodes in the network. However those metrics are of different types and cannot be mixed.

In general, if $w(i, j)$ is a metric for the link (i, j) then for any path $p = (i, j, k, \ldots, l, m)$ the metric $w(p)$ is [149]:

- additive if: w(p) = w(i, j) + w(j, k) + … + w(l, m),
- multiplicative if: w(p) = w(i, j)w(j, k)… w(l, m),
- concave if: w(p) = min[w(i, j), w(j, k), …,w(l, m)].

It means that metrics such as delay, delay jitter and cost flow are additive, loss probability is multiplicative and bandwidth is concave.

As it was mentioned in the introduction to this chapter, finding a route subject to multiple constraints is in general an NP-complete problem. The NP-completeness theorems for additive and multiplicative metrics are presented in [149], where the authors additionally explain why it is not recommended to define a single metric as a function of multiple parameters. Let us assume that the following metrics: bandwidth B, delay D and loss probability L are put together into the following formula (metric) (32):

$$f(p) = \frac{B(p)}{D(p)L(p)} \tag{32}$$

However, the metric (32) as is does not allow to assess whether the QoS requirements for a specific route are met or not. What is more if we assume a path $p = (a, b, c)$, then in case of delay (additive metric) we can write $f(ab + bc) = f(ab) + f(bc)$ and in case of bandwidth (concave metric) $f(ab + bc) = min[f(ab), f(bc)]$. In case of the metric (32) there may not exist a composition rule as it comprises of three different types of metrics (additive, multiplicative and concave).

Despite the NP-completeness problem, computationally feasible algorithms are known for bandwidth and hop count optimization. The following ones are mentioned in the literature [148]:

- *Widest-shortest path algorithm*: the path with the minimum number of hops is chosen. If several paths have the same number of hops, then the one with the maximum available bandwidth is selected.
- S*hortest-widest path algorithm*: the path with the maximum available bandwidth is chosen. If several paths have the same available bandwidth, then the one with the minimum number of hops is selected.
- *Shortest-distance algorithm*: the shortest path is chosen, where the distance of a *k-hop* path p is defined as (33):

$$dist(p) = \sum_{i=1}^{k} \frac{1}{r_i} \tag{33}$$

  and r_i is the bandwidth of link i.

Nevertheless those algorithms seem to be insufficient for the purposes of efficient QoS routing in the WINNER network. That is why a multi-constrained Quality of Service routing algorithm is proposed, which is based on the generalized Dijkstra's algorithm. In this approach k shortest paths are selected between the source and destination nodes. Hop count is proposed to be the main criterion for this selection, but it is still an open issue, which will be evaluated during the simulation research. Once the selection procedure is completed, every single application may select a path from the resulting set, which best meets any other criteria. So not necessarily the shortest path must be chosen, but for example the longer one with more bandwidth available. It is also important to provide some balancing procedures for the network traffic. To this end there may be the third criterion taken into account which guarantees that out of the paths meeting the first and second criteria the one is chosen, which additionally will allow to distribute the load between distinct routes more evenly. Of course, if there is a need, also other criteria may be applied to the initial set of paths, which was selected with the use of the main criterion.

The classical Dijkstra's solution [154] is applicable to the single shortest path problem, whereas the generalized version [153] allows to find k shortest paths leading from the source node s to the target node t. Those paths are the elements of the set P, which contains all the possible paths between the aforementioned nodes. Let the *i-th* shortest path in the graph G, which is leading from s to t, be denoted by p_i. The weight $w(p)$ associated with this path is defined as follows (34):

$$w(p) = \sum_{(i,j)\in p} w(i,j) \tag{34}$$

The generalized algorithm searches for the set of paths $P_k = \{p_1, p_2, ..., p_k\} \subseteq P$ under the following conditions:

1. The path p_i is selected prior to p_{i+1} for any $i = 1, ..., k-1$.
2. $w(p_i) \le w(p_{i+1})$ for any i = 1, …, k-1.
3. $w(p_K) \le w(p)$ for any $p \in P - P_K$.

In the classical Dijkstra's algorithm [154] there is a single label d_i associated with each node, which is equivalent to the distance between that particular node and the source node. The distance is expressed by the means of the sum of the weights corresponding to the individual edges. In case of the generalized version for selecting k shortest paths, there may be one or more labels associated with one node, depending on the number of paths passing the specific node. Each such label, connected with the node *i*, corresponds then to the distance between this node and the source node *s* on the *k-th* path.

It was noticed that associating many labels to one node is computationally ineffective. That is why a modified version of the generalized Dijkstra's algorithm was proposed [153]. The modification is concerned with applying the extended set of nodes. It means that each node may appear many times, depending on the number of the shortest paths that were built with the use of this particular node. As a result there is always single label assigned to each node. For the detailed description of the proposed version of the generalized Dijkstra's algorithm see 7.2.4.

4.2.5.5.3 The OLSR routing protocol

For the WINNER project purposes a modification of the Optimized Link State Routing protocol *OLSR* [152] is proposed. The classical *OLSR* routing protocol stems from the Open Shortest Path First protocol *OSPF*, which is one of the best solutions for wired networks. As a link-state class protocol it is very stable and due to its proactive nature it offers the instant availability of the routing information.

One of the main advantages of the *OLSR* routing protocol is that it uses the selected nodes only for transmission of the control messages. Those nodes are called *MPRs* (Multi Point Relays) and are chosen by a given node out of its all 1-hop neighbours, to which bidirectional links exist. In consequence all other neighbours in the range of the node *N*, which do not belong to the *MPR* selected by this node, also receive and process the broadcast messages sent by the node *N*, but do not retransmit them (Figure 4-69). Such an approach is aimed at minimizing the number of redundant retransmissions and therefore optimizing the global control traffic level.

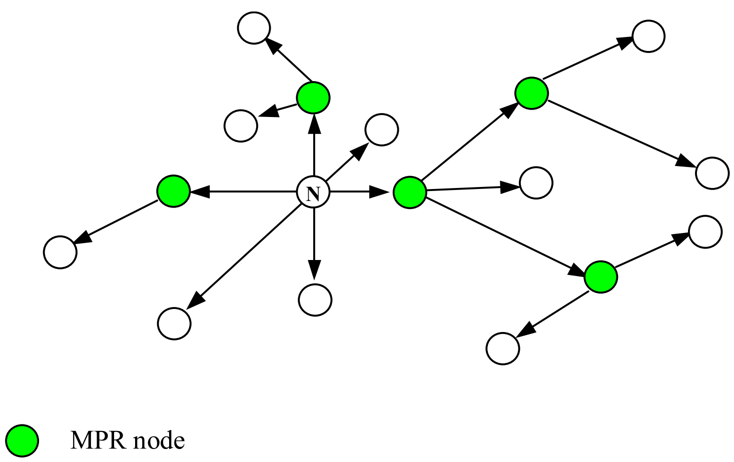


Figure 4-69: MPR selection in OLSR

There are two types of the control messages: *Hello* and *TC (Topology Control)* messages. The first type is used for neighbour discovery and the second one for topology dissemination [151].

Hello messages are broadcast by nodes on their all interfaces in a periodic manner for the purposes of link sensing, neighbour detection and *MPR* selection signaling [152]. They are generated on the basis of the information stored in the *Local Link Set*, *Neighbour Set* and *MPR Set*. Link sensing is necessary for detecting whether the radio link exists in both directions, just one or maybe none of them. There is a direct association between the existence of a link and a neighbour. *Hello* messages are received by all one-hop neighbours, but they must not be forwarded. Nevertheless they allow each node to discover its both one-hop as well as two-hop neighbours. What is more, by the means of *Hello* messages each node also selects a set of *MPRs* out of its one-hop neighbours, as it was described above. Each such selected node maintains a set containing those nodes, by which it was selected as the *MPR*. Those nodes are called the *MPR Selectors* and each selected node may broadcast messages on behalf of one of them only.

TC messages are broadcast by the *MPR* nodes only in order to build the topology information base. Those messages are flooded to all nodes and take advantage of *MPRs* [152]. As a result the information about the *MPR* selectors for all *MPR* nodes is advertised throughout the entire network in a way that minimizes the number of retransmissions. Routes are constructed on the basis of the advertised links and links to the neighbours. To this end a node must disseminate the information about the links between itself and all the nodes in its *MPR* selector set, so that there was sufficient information provided to other nodes. The *TC* message follows the pattern of the general message format, where the *Message Type* field is set to *TC_MESSAGE* [152].

The information concerning the neighbours and the topology is updated periodically, in a proactive way. Usually, due to the constraints of the battery powered mobile nodes, reactive approaches are rather recommended for ad hoc-like networks. Nevertheless if we assume, that in case of the WINNER project only no power constrained relays are chosen as the *MPRs* then there will be no penalty for the proactivity feature of the *OLSR* protocol.

The *OLSR* protocol has another enhancement, which matches the WINNER requirements. Namely, every single node may specify its willingness to forward traffic coming from its neighbours [152]. By default all the nodes *WILL_DEFAULT* forward packets. Later on some of them may decide that they *WILL_NEVER* do that and so on. In consequence a MT node may be willing to take part in routing as far as it does not detect too low battery power level. It is proposed that all the relay stations without power constraints were of the type *WILL_ALWAYS*.

Unfortunately, despite of many advantages, the *OLSR* protocol has also some drawbacks, which must be obviously modified. First of all it provides no *QoS* functionality and uses the classical Dijkstra's algorithm for the routing purposes. The *QOLSR* protocol proposition could be taken into account [154], which based on the *shortest-widest path algorithm*. However only two types of metrics (bandwidth and delay) are considered in this solution, which does not seem to fulfill the requirements of the WINNER project.

That is why the modified *OLSR* protocol is proposed as the optimum solution. The modification is connected with replacing the classical Dijkstra's algorithm with its generalized version, which is suitable for selecting *k* shortest paths. There will be obviously the necessity for collecting more information about different parameters of the network links, but in consequence the *QoS* routing will be feasible. What is more the inherent features of the *OLSR* protocol will be very useful for the purposes of the mobility management issue. It is also assumed that only nodes without power constraints (fixed or mounted on a vehicle) will be used for forwarding the traffic whereas other nodes will be not excluded form taking part in this process if they are willing.

4.2.5.6 Forwarding Mechanism in a HiperLAN/2 evolution

The focus of this section is the description of a routing and forwarding mechanism for a relay based systems as described in Section 4.2.4.1. In contrast to mesh networks only one route towards the UT is available and has to be established. As described in Section 4.2.4.1 the UT can be associated directly to the AP or via a FRS. In any case the AP has to be aware that the respective terminal is roaming in its multi-hop cell. In the HiperLAN/2 based approach as described in Section 4.2.4.1 the FRS appears towards the AP like a UT and towards the UT like an AP. The consequence is that each UT has to request an MAC ID from the AP.

For the case of signalling between the UT and the AP, two approaches are possible for the chronological signal flow:

- Synchronous signalling
- Asynchronous signalling

In the following the two approaches are explained on the example of a MAC ID assignment. Figure 4-70 shows the signal flow to assign the UT a MAC ID with synchronous signalling. By sending RLC_MAC_ASSIGN to the FRS the UT requests a MAC ID from the FRS and starts a control timer. In the following the FRS itself requests a MAC ID on the 1st hop from the AP. After the receipt of the RLC_MAC_ID_ASSIGN_ACK signal from the AP the FRS sends an RLC_MAC_ID_ASSIGN_ACK including the assigned MAC ID, not necessarily the same as between the FRS and the AP, to the UT. The advantage of the synchronous signalling is that the relay only transmits assigns the MAC ID after the AP has confirmed its availability on the first hop. The disadvantage are the timer constraints as the MAC ID assignment needs more time than in the single hop case. Therefore the respective timers have to be adapted.

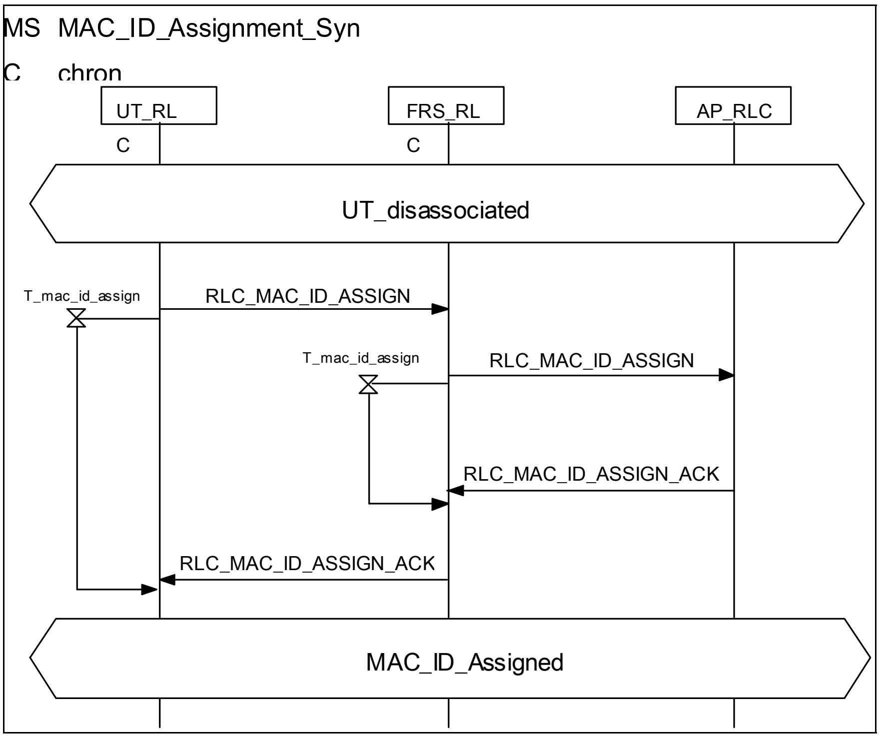


Figure 4-70: MAC ID assignment with synchronous signaling

Figure 4-71 shows the alternative case with asynchronous signalling. In contrast to the synchronous signalling the FRS assigns the UT a MAC ID after the receipt of the RLC_MAC_ID_ASSIGN signal without making an inquiry at the AP. After the MAC ID assignment procedure the FRS initiates the MAC ID assignment procedure with the AP. Both MAC ID assignment procedures are running autonomously. The advantage a faster signalling as the FRS does not wait for any response from the AP.

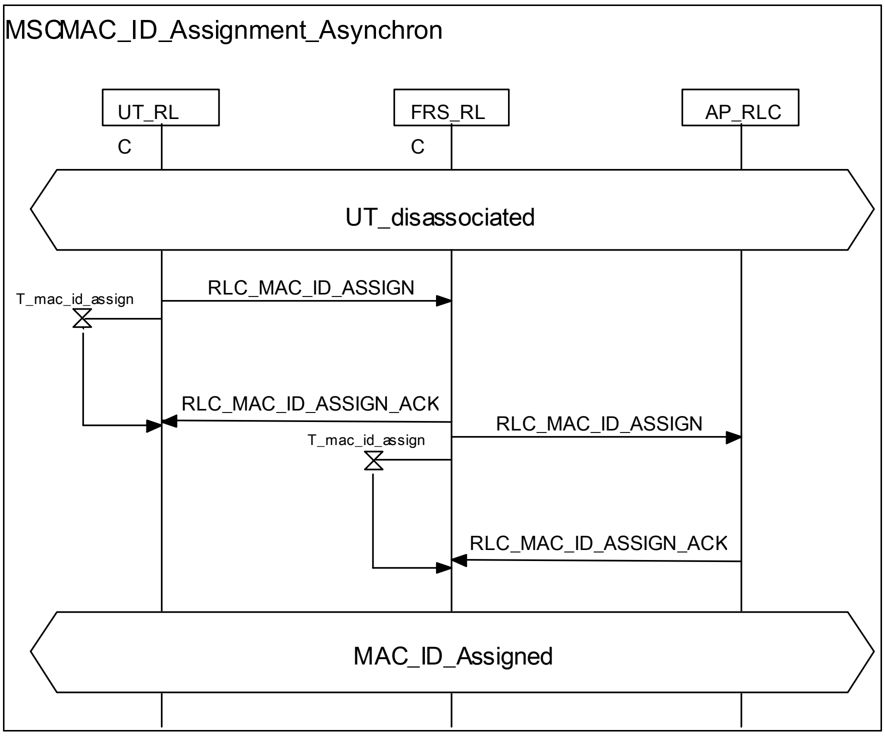


Figure 4-71: MAC ID assignment with asynchronous signaling

Asynchronous signalling has some crucial disadvantages in the case of long signal flows as unsolved dependencies between the second and the first hop might occur. Therefore the signal flows for the UT association and the connection set up will be described for the asynchronous case in the following.

4.2.5.6.1 UT association in a multi-hop environment

Following the MAC ID assignment procedure as shown in Figure 4-70, the RLC Link Capability procedure will be started. The RLC link capability procedure serves to align the protocol parameter between the UT – FRS and the FRS – AP connection. A forwarding manager within the FRS is in charge to map the parameter of the messages.

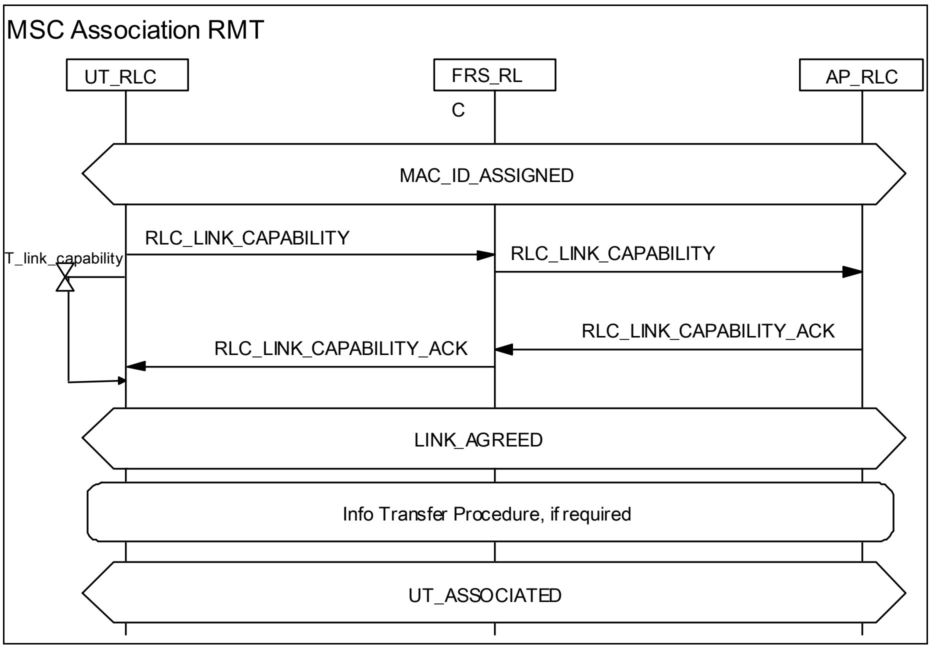


Figure 4-72: UT association over multi-hop link

4.2.5.6.2 UT multi-hop connection set up

Following the association the UT can initiate a connection set up by sending an RLC_SETUP to the FRS. The signal contains next to other information a list of Data Link Control Connection IDs (duc-descr-list), which includes the direction of the requested connection (Simplex-forward, Simplex-backward, duplex).The FRS forwards the request to the AP which response with an RLC_CONNECT and confirms the desired connection request. If the AP is not able to support the desired connection request the FRS will assess the modified parameter and forward them towards the UT. If the UT accepts the proposed connection parameter it answers with an RLC_CONNECT_ACK which is delivered via the FRS towards the AP.

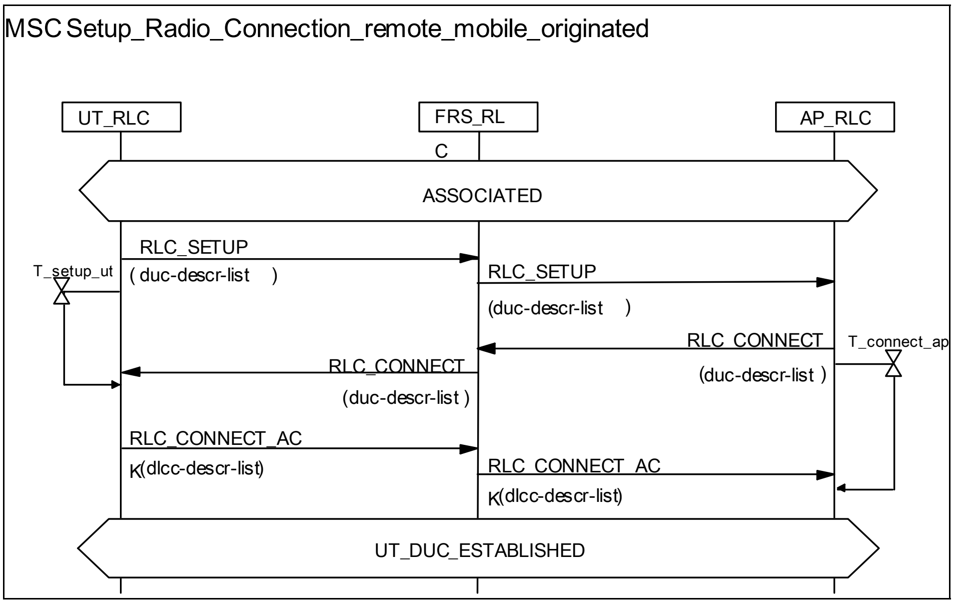


Figure 4-73: Multi-hop connection set up

4.2.5.6.3 Conclusion of HiperLAN/2 based forwarding mechanism and outlook for further WINNER studies

In this Section 4.2.5.6 an initial approach on how the forwarding connection in a multi-hop environment can be realized. This includes the addressing and route definition of for UTs directly connected to a FRS and thus with a two hop connection to the AP and backbone network. It has to be noted that this is an initial approach that has been approved by means of simulation results as discussed in Section 4.2.4.1.

For the described scenario further studies have to take the number of stations as well as the proposed ARQ mechanisms into account. It has to be investigated whether it is possible to establish a simple connection on the first hop without assigning an extra MAC ID. Furhter investigations will be ongoing for the channel based approach based on a TDMA system with fixed slot length as outlined in Section.

4.2.6 Cooperative Relaying

4.2.6.1 Cooperative Relaying – Motivation and Idea

Cooperative relaying brings together the worlds of multi-antenna systems and relaying. By allowing cooperation of mobile terminals, *distributed antenna arrays* can be built that overcome the drawbacks of correlation and spatial limitations. At the same time, placing a relay between communicating nodes is beneficial from the viewpoint of propagation losses.

In general, cooperative networks employ relay stations that receive signals from a source and resend a processed version of these signals to the intended destination. The destination node *combines* the signals from source (dashed-line arrow in Figure 4-74 and relays, thereby exploiting useful information that is unnecessarily discarded in conventional relaying or sometimes even regarded as interference. Similarly, combining from multiple relays is a valuable option.

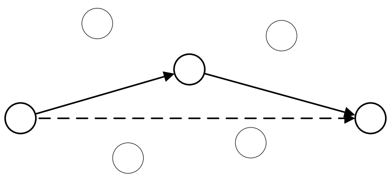


Figure 4-74: A possible two-hop relaying scenario. Conventional relaying occurs along the solid arrows. In case of cooperative relaying, the destination additionally takes the signals sent by the source into account (dashed-line arrow).

In particular, two advantages become apparent:

- First, cooperative systems are able to explicitly make use of the *broadcast nature* of the wireless medium as the signal transmitted by the source can in principle be received by the relays as well as by the destination.
- Second, the relay channel offers *spatial diversity*, as the signals from source and relays propagate to the destination through essentially uncorrelated channels.

In consideration of this, cooperative relaying can be understood as an *extension* of conventional relaying. Viewed even more generally, cooperative relaying represents a step on the way from *point-to-point coding* toward what is referred to as *network coding*. For an introductory example of a simple cooperative relay scheme, please refer to the appendix.

4.2.6.2 Principles and Existing Work

4.2.6.2.1 Advantage of Cooperative Relaying: Spatial Diversity

In addition to the relaying advantages discussed earlier, cooperative relaying exploits the spatial diversity that is offered by antennas that are distributed among different terminals rather than being located closely to each other at a single station. Essentially, one can assume that the different paths between the network elements are mutually uncorrelated, so that the strong information-theoretic gains are not reduced by correlated propagation.

The promise in spatial diversity is well known; however, it is hard to conceive that a large number of antennas can be deployed at a single network entity (e.g., relay). Virtual antenna array (VAA) structure is a pragmatic approach to communication network design in which communicating entities utilize each other's resources in a symbiotic relationship. In this context, relays (mobile or fixed) may cooperate to facilitate enhanced diversity architectures; such a distributed architecture may even emulate a MIMO system: MIMO-VAA [215]. Especially in situations where the number of antennas at the mobile terminals is the limiting factor, the MIMO-VAA architecture will be very helpful. In MIMO-VAA systems, since the antenna elements will be relatively further apart, the space constraint that leads to mutual correlation among the antenna elements in conventional MIMO systems (which may actually jeopardize the performance of MIMO systems in significant ways) will not be a concern anymore. The simple diversity architecture explained below lays the foundation to generalized MIMO virtual channels that we will investigate in the future phases of the WINNER project.

4.2.6.2.2 Classification of Cooperative Relaying

Relaying schemes and distributed multi-antenna systems can be categorized according to a variety of parameters. Based on the overview provided in Figure 4-75 we discuss the most crucial parameters, which will then allow us to examine previous and related work.

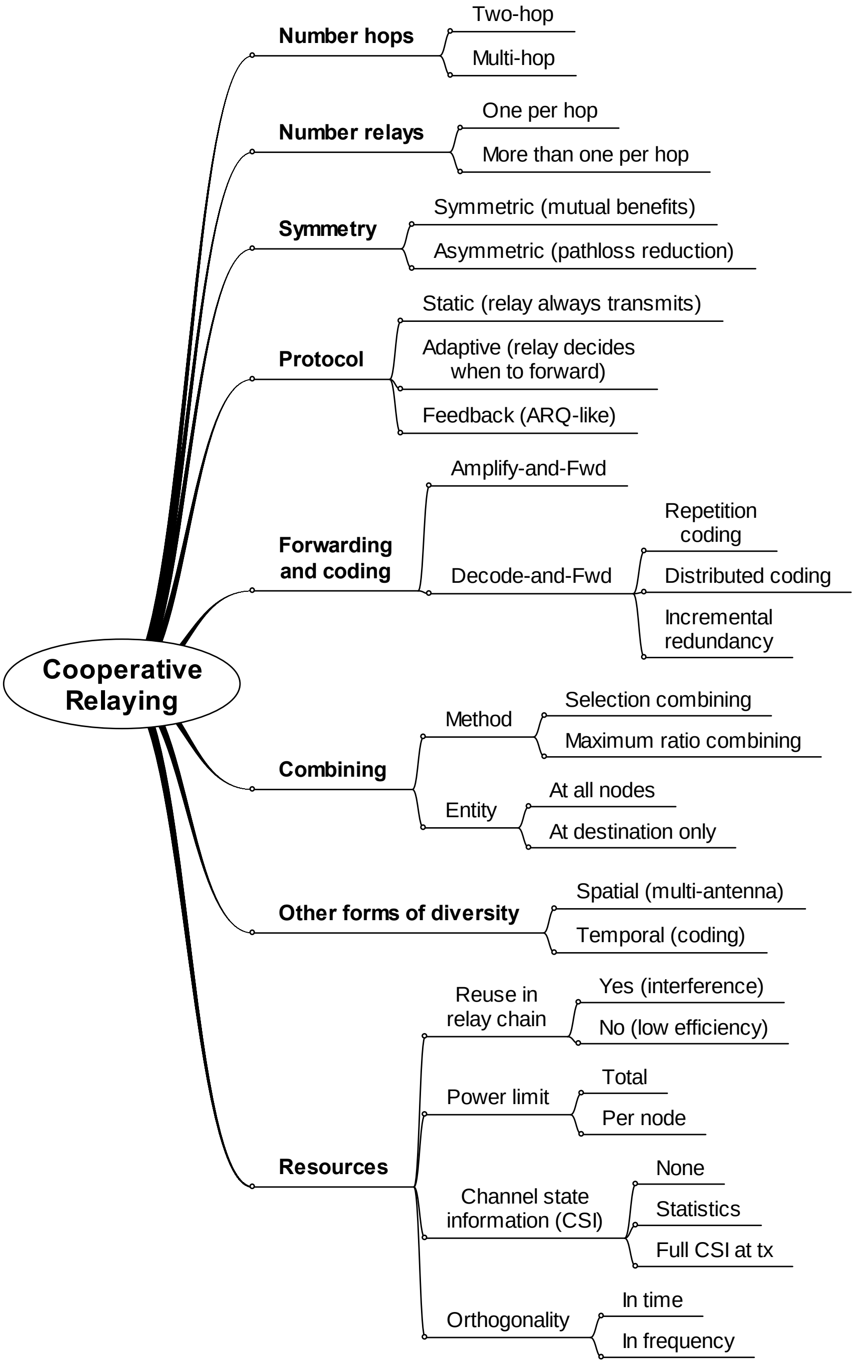


Figure 4-75: Classification of cooperative schemes.

Forwarding Strategy: Amplify- vs. Decode-and-Forward

Amplify-and-forward In amplify-and-forward schemes the relays amplify the signal in the analog domain, without decoding the received and re-encoding the newly transmitted message. Such retransmission is characterized by *noise amplification* as is already known from repeaters. While amplify-and-forward protocols are simple in nature, they may be hard to implement: the use of time-division schemes for relaying requires large amounts of samples to be stored. Similarly, frequency-division approaches require the relay to retransmit the received signal on a different frequency, thus calling for the hardware that performs such frequency conversion. Amplify-and-forward schemes are also known as "analog relaying" or "non-regenerative relaying"

.

Decode-and-forward In decode-and-forward schemes, also referred to as "digital relaying" or "regenerative relaying", require the relays to demodulate and decode the signal prior to re-encoding and retransmission. While these schemes do not suffer from noise amplification and complex implementation constraints, they pose the danger of *error propagation* that may occur if the relay incorrectly decodes a message and retransmits this wrong information "freshly" encoded. One can further subdivide decode-and-forward schemes according to the coding scheme:

- *Repetition coding* schemes have the relay repeat the source's message.
- Additional coding gains can be realized by advanced spatially *distributed coding* schemes, such as distributed Alamouti coding or distributed turbo coding. It is clear that these gains come at the cost of increased complexity.
- Schemes providing *incremental redundancy* have been proposed in connection with feedback. Protocols similar to ARQ can be designed, where the redundancy is delivered by relays instead of the original source node.

Protocol Class: Static, Adaptive, and Feedback Protocols

Almost independent of the above classification, relaying schemes can be of static or adaptive nature with respect to their transmission rate and protocol.

- **Static protocols** In static protocols, transmissions from source and relays follow a fix pattern. Usually, a two-phase protocol would be employed that makes the source broadcast in the first phase, and relays and destination receive this broadcast signal. In the second phase, the relays would then retransmit their signals, using either amplify-and-forward or decode-and-forward schemes. Such static schemes may suffer from the drawbacks of their repetition-coded nature: for some channel conditions, the destination may have successfully decoded the source message in the first phase, thus making the second phase a waste of resources. Similarly, when the relays were not able to receive the broadcasted message with a quality that allows for retransmission, then the second phase is useless as well. On the other hand, the static protocols are attractively simple.
- **Adaptive protocols** In adaptive protocols, the relays can prevent error propagation by deciding whether or not to retransmit signals. A variety of protocols can be designed.
- **Protocols based on feedback** can overcome the drawbacks of the fix two-phase nature by aiming at exploiting variable channel conditions. Retransmissions occur only if the destination indeed requires additional information. Such adaptive protocols generally yield a variable rate and may require some form of feedback, thereby complicating protocol design.

Symmetry: Symmetric vs. Asymmetric Networks

The term *symmetric* describes a scenario as in Figure 4-76(a), where two source terminals are alternatingly assigned the roles of source and relay. This truly *cooperative relaying*, which we discussed as the introductory example (Figure 7-12), is useful when both transmitting terminals experience similar pathlosses to the common destination. It is inherently clear that the *mutual* assistance precludes benefits from pathloss savings.

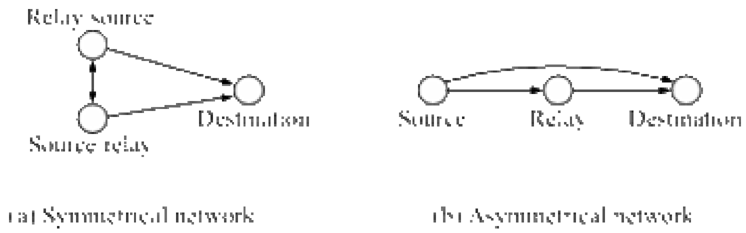


Figure 4-76: Symmetric vs. asymmetric networks. In symmetric networks, the average channel fading (or pathloss, respectively) is equal for all links from the cooperating source terminals to the destination. In asymmetric networks, the pathloss reduction comes at the cost of loss of symmetry, i.e., the roles of relay and source can no longer be exchanged for the mutual benefit of both stations.

By contrast, pathloss savings can be exploited in *asymmetric* scenarios as depicted in Figure 4-76(b). Having an intermediate relay node to help transmission from source to relay yields pathloss savings; yet, these advantages come at the cost of a loss of symmetry in the above sense. Instead, a different form of symmetry is introduced: source and destination can now act interchangeably, with the relay node offering help for the two-way communication from source to destination *and vice versa*.

Further classification can be made based on the parameters in Figure 4-75; we will do so when discussing related work in the following. Note that according to this classification, the conventional relaying schemes discussed in the previous sections can be categorized as static, asymmetric decode-and-forward schemes.

4.2.6.2.3 Concepts

We have argued that cooperative relaying has the potential of bringing various aspects of conventional relaying and multi-antenna systems together. We now discuss performance bounds and review related work.

Network Coding and Transmit Diversity Bound

In contrast to the point-to-point coding approach of Gupta and Kumar [169], Gastpar et al. [171] consider a complete network coding approach. In their study, a network consists of a source node, a destination node, and *M* relays that assist transmission from the source to its destination; hence, there are a total of *K=M+2* nodes in the network. Most importantly, the only communication is between the randomly selected but fix source and destination pair and the assisting relays only. Under strict information-theoretic assumptions, it is shown that the capacity increases as *log K*, which should be compared to the *constant throughput* per pair as discussed in for for conventional relaying in [47].

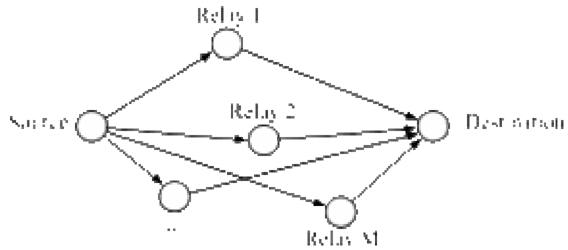


Figure 4-77: Network coding using M relays. Communication from source to destination is divided into two phases. In phase 1, the source node broadcasts information to the M relays, while in phase 2 the relays forward an amplified version to the destination node, which estimates the transmitted signal by combining all received versions.

The performance bound is simply the *receive diversity* case that arises if one assumes perfect channels between the relays and the destination. In that case, the network reduces to a system with one transmit and *M+1* receive antennas. The capacity of such systems and their fundamental tradeoffs are well-known.

Moreover, in order to allow for a fair comparison with a directly transmitting system, one needs to stay within the power limits. Specifically, the total power consumed by the *M+1* transmitting nodes, or antenna elements in this case, may not exceed the power *P* that the source, or single antenna element, uses for our baseline case of direct transmission (with one antenna). This consequently leads to a transmit diversity bound for cooperative schemes.

Previous Work

Corresponding work can loosely be classified as being related to information-theoretic fundamentals, and practical considerations on isolated soure/destination pair as well as cooperative networks.

A- Theoretical Fundamentals

For a discussion of the fundamental work on classical relay channel, which partially already include the first ideas of cooperation, please refer to 4.2.3.1.3. First interest in the relay channel dates back to the work of van der Meulen [202][203]. Cover et al. determined capacity for the physically degraded relay channel, where coding is performed in an incremental manner [163][164]. While these examples focus on the three-terminal case, a more general approach is taken by Gastpar et al., who establish performance bounds by examining the situation in which a single source-destination pair is assisted by a *network* of relay terminals [171][170].

Explicit *cooperation* for the mutual benefit of neighbouring nodes has first been considered by Sendonaris et al. [197][198][199]. For noisy inter-user channels and various degrees of knowledge of channel state information (CSI) in symmetric networks, they show that significant gains can be achieved over non-cooperative direct transmission. While the early discussion does not address the orthogonality constraint, the considerations are later extended to orthogonal transmit and receive resources at the relay. Moreover, alternative coding schemes that overcome the repetition-coded nature of relaying protocols [195][196] are proposed for symmetric networks.

More recently, Laneman et al. [188] propose various cooperative protocols for the three-terminal case, among them static and adaptive protocols for the amplify-and-forward and the decode-and-forward case, as well as a protocol based on feedback. The conducted analysis from the perspective of outage probabilities for limited bandwidth and constrained end-to-end delay shows that relaying - even in its cooperative form - may suffer from the necessity of providing orthogonal resources for reception and transmission at the relays. The analysis in [188] captures the significant parameters SNR (*i.e.*, energy), pathloss, spectral efficiency, and network geometry and focuses on *symmetric* networks.

A simple version of the adaptive decode-and-forward protocol has been suggested by Herhold et al. in [183]. The protocol operates as a simple extension of conventional store-and-forward; we will discuss this scheme in section 4.2.6.3

Zhao and Valenti [214] propose new strategies for the second phase in a two-hop protocol. Following a broadcast from the source in phase one, the option of having the source or relay should transmit in phase two is examined. In addition, repetition coding with maximal ratio combining (MRC) is compared with incremental redundancy (code combining, CC). The conducted theoretic analysis shows that code combining gives a gain of about one decibel over maximum ratio combining, thus suggesting that in the limit, incremental redundancy and distributed coding techniques do not provide significant gains compared to schemes based on repetition coding.

The work of Laneman is extended by Nabar et al. [190], who suggest an additional cooperative protocol based on *receive-collision*. The basic idea is to have source and relay transmit different signals simultaneously to the destination in phase two; thereby making more efficient use of resources. These benefits come at the cost of increased complexity as the receiver needs to resolve the collision based on orthogonal signal designs and multi-user detection.

Boyer et al. are to our knowledge the first to study *multi-hop* scenarios [160][161]. Four different channel models and simple static cooperative protocols are examined. It is argued that the feedforward and feedback interference in multi-hop chains is a form of intersymbol interference, and can therefore be eliminated by means of classical equalization techniques. Unfortunately, the required rate increase and the challenging problems of resources assignment in multi-hop chains are not considered, and feedforward and feedback interference are assumed to the negligible in the corresponding simulations.

B- Distributed Coding Strategies

Common to the contributions discussed so far is that they elaborate on the information-theoretic bounds - either in the sense of Shannon or ergodic capacity and the related concept of outage. We now turn to reviewing various *coding strategies* for relay systems.

These range from distributed Alamouti coding and the use of rate-compatible convolutional codes to the application of distributed turbo coding. Some of these concepts have the power of overcoming the repetition-coded nature of relay protocols, leading to benefits from additional coding gains at the cost of increased complexity.

Distributed Alamouti coding Anghel et al. [156][158] suggest *distributed Alamouti coding*, where two amplify-and-forward relays are assumed to perform simple operations on the analog signals, namely delaying and conjugating (Figure 4-78). This does not require decoding at the relay, and facilitates Alamouti-coding for a single source and two relays.

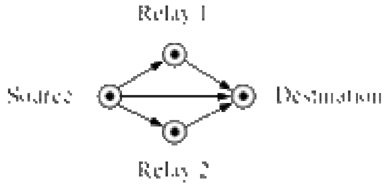


Figure 4-78: Distributed Alamouti coding from [156][158].

A slightly different approach for the deocode-and-forward case is proposed by Xue et al. [206] for a group of three cooperating terminals that use a *hierarchically modulated Alamouti code* in a cascased/windowed operation. The scheme explicitly exploits the stronger source-relay links by using the concept of multiclass information broadcasting [201] to convey information to cooperating relays while simultaneously transmitting different information to the destination. The demonstrated advantages clearly come at the cost of a strongly increased complexity. For details, refer to [206]. We will investigate this scheme in more detail in a later section.

Distributed block and convolutional coding

Hunter et al. [178][179][180] discuss cooperation for symmetric networks through code word partitioning. Codewords of N bits, containing K information bits ($R=K/N$), are partitioned into N_1 and $N_2 =N- N_1$ bits, where the N_2 bits can be determined from N_1 (parity). Two cooperating users (1,2) each broadcast their N_1 bits and a corresponding cyclic redundancy check (CRC) in a first phase, and the respective partners try to detect these (Figure 4-79). If successful, then the remaining N_2 parity bits are determined and sent by the assisting relay station in the second phase. Otherwise, *i.e.*, if decoding failed, the terminal send their "own" N_2 bits. In total, there are four different decoding cases. The scheme is depicted in Figure 4-79.

Figure 4-79: Distributed coding from [179]. Communication takes place in four phases. $N_1^{(k)}$ denote the systematic bits of user k, and $N_2^{(k)}$ are the corresponding redundancy bits. The drawing depicts the scenario where both terminals can decode their partner's systematic bits N_1, thus being able to compute the corresponding N_2 bits and sending them in phase 2.

The prosed scheme avoids error propagation by sending the assisting station's "own" parity bits if decoding failed. The authors suggest the use of *rate-compatible punctured convolutional codes (RCPC)* [172] and analyze the pairwise error probability and block error rates. A disadvantage of the scheme is the low overall code rate (e.g., *1/4* and *2/5*). Likewise, it is left unclear how the orthogonality constraint is taken care of. Results show a 1-2 dB gain over $R=1/4$ coded simple amplify-and-forward operation.

[186][177] extend the work by discussing space-time coding in phase two, where it is proposed that each terminal sends its own redundancy bits (N_2) and its partner's redundancy bits in an Alamouti fashion.

Distributed turbo coding

The first to discuss *distributed turbo coding* are Zhao and Valenti [204][213][212]. Janani and Hunter likewise consider this approach [186][177]. The source encodes using a rate *1/2* recursive systematic convolutional code; the relay decodes and re-encodes after *interleaving*. See Figure 4-80. The destination decodes using a standard turbo decoder. This scheme clearly avoids repetition coding, therefore achieving coding gains at the cost of additional complexity. Two asymmetric cases have been studied, and a comparison to the use of block codes is performed [180].

The results indicate that by using one relay in connection with a strong turbo code, one can outperform conventional repetition coding by 2 dB at a frame error rate of FER = 10^{-2}. These coding gains increase correspondingly with the use of multiple relays.

Similar work has been conducted by Oikonomidis [191], who investigates the case of turbo-coded transmission from source to relay in order to overcome the limiting effects of the bottleneck source-relay link.

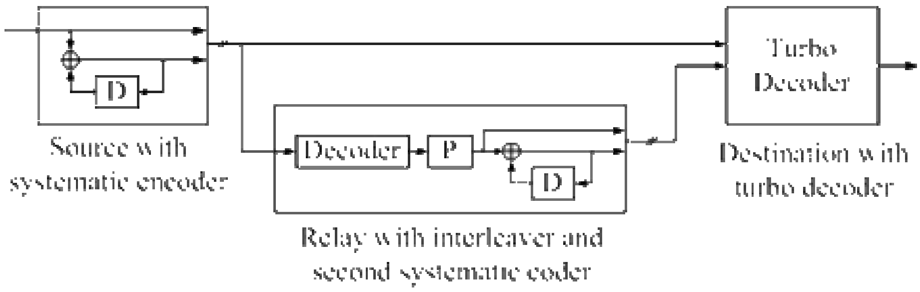


Figure 4-80: Distributed turbo coding from [213].

C- Incremental Redundancy

Zimmermann et al. [207][208][209][210] show that by allowing feedback, a whole class of distributed redundancy schemes can be created. For example, a simple protocol would have the destination decode after reception in phase one. The destination uses a single bit of feedback to indicate to source and relay whether or not decoding has been successful. If decoding was not successful, then redundancy is delivered by the relay. Otherwise, the source proceeds with transmitting the next block, thereby avoiding repetition.

We provide a discussion of additional concepts and aspects in section 7.3.3, among them the idea of virtual antenna arrays, large-scale cooperative relaying, network implications, and aspects related to mobility.

G- Summary and Conclusion

The above review highlights the following.

- Although fundamental work has been performed on cooperative relaying decades ago, it was only until recently that the ideas of explicit user cooperation and relaying have gained momentum. The past two years have seen an explosion of interest in cooperative relaying schemes.
- Many ideas have been proposed, ranging from simple repetition coding for the single-relay case to complex coding strategies such as distributed Alamouti coding and distributed space-time diversity for multi-relay networks.
- Despite the intensified efforts, work is still in its early stage. To our knowledge, no cooperative concept has been realized so far.
- Generally, stronger complexity leads to performance gains. Yet, the reported improvements, for example from using distributed coding techniques instead of repetiton coding, are most often at the order of a few decibels.

4.2.6.3 Adaptive Decode-and-Forward Schemes – Single-Antenna Case and Related Protocols

We study the basic building block of relaying scenarios - the two-hop scenario. We initially focus on transmission schemes for *decode-and-forward* terminals that are equipped with a *single antenna*. Such scenarios are analytically tractable, and still provide interesting insights into the related tradeoffs. For the relay protocols, we assume a certain relay has been selected by a higher layer routing protocol; based on our results that we obtain from considering the physical layer, we will later discuss requirements for effective routing protocols.

We start by outlining the protocols of interest; a summary is provided in Figure 4-82.

4.2.6.3.1 Protocols

Direct Transmission: Our baseline model for comparison is direct single-input single-output (SISO) transmission from a source node to its destination node.

Conventional Relaying: Recall that this form of relaying the basic means of service provisioning in ad hoc networks, and offerers a range-rate tradeoff for cellular networks. The protocol is designed to benefit from a reduction of the end-to-end pathloss between source and destination by exploiting the nonlinearity of attenuation as a function of distance. We will also refer to it as layer 3 decode-and-forward relaying (L3DF). The source transmits to the relay in phase one, and the relay re-transmits a newly encoded signal to the destination in a second phase. For decoding, the destination solely relies on the signal it receives from the relay.

Transmit Diversity: Spatial diversity gains can be exploited using multiple antennas at a terminal. Here, we refer to transmit diversity as employing two antennas at the source for direct transmission to the destination according to Alamouti's scheme [156]. While attractive due to its simplicity, this scheme requires integrating multiple antennas at a terminal and uncorrelated channels from each of the antenna elements. In many scenarios, these conditions cannot be presumed fulfilled. Cooperative schemes can overcome these limitations by “distributing” the two antennas among source and relay.

Cooperative Relaying - Adaptive Decode-and-Forward (AdDF)

In cooperative relaying, the destination combines the signals that have been transmitted from source and relay. The simplest solution would be to have the relay unconditionally forward to the destination. However, Laneman et al. [188] have shown that such fixed decode-and-forward protocols do not yield the desired diversity gains, as performance is limited by error propagation incurred by decoding errors at the relay. To circumvent this, they suggested an adaptive decode-and-forward protocol (AdDF), also referred to as selection relaying, in which the relay forwards only when it has reliably decoded. Part of our contribution is a proposal for a simple version of such a protocol, which is subsequently described together with the original, more complex scheme [188].

Simple Adaptive Decode-and-Forward

We suggest the following:

- *In phase one*, the source broadcasts its information. Both relay and destination receive faded noisy versions of this signal, and the destination stores the received signal for later processing. The relay measures the effective SNR of the received signal; if it allows for successful decoding, then it does so. Otherwise it refrains from decoding.
- *In phase two*, if the relay has decided to decode, then it re-sends a newly encoded version to the destination, thereby providing low-complexity redundancy in the form of repetition coding. The destination combines the received version of this signal with the stored samples it has previously received from the source. We assume maximum ratio combining (MRC). Otherwise, i.e., if the relay has decided not to decode, then it simply remains silent. The destination detects this case based on the lack of sufficient signal strength, and for decoding it needs to rely on the samples stored in phase one.

This protocol achieves spatial diversity gains, since for good channel conditions the relay frequently decodes and forwards copies of the original information over an uncorrelated channel to the destination. Simultaneously, we can benefit from pathloss savings: a relay station located between source and destination will receive the messages broadcasted by the source much more reliably than the destination, and in turn it needs to use a considerably smaller transmit power to “reach” the destination.

The key feature of this protocol is the relay's decision whether or not to decode and forward; to prevent error propagation, it forwards only if it can reliably decode.

Complex AdDF

This is the more complex “selection relaying” protocol suggested in [188]. It differs from the simple AdDF protocol only in the cases in which the relay has decided not to decode. Recall that in this case both relay and source remain silent in the second phase of the simple AdDF protocol. The complex AdDF protocol, however, prevents this “silence” by having the source repeat its message in this second phase. The destination then combines the two versions it has received in the two phases; the resulting gains from standard repetition coding come at the cost of increased complexity as the source must have information on the decoding status of the relay.

Cooperative Relaying - Receive Collision

A further extension of these protocols has been proposed by Nabar et al. (Protocol I in [190]). The basic idea is to have the relay as well as the source transmit in the second phase (Figure 4-81). This parallel transmission of data is beneficial from a capacity point of view; however, it calls for resolution of the resulting collision at the receiving destination. We will see that this receive collision protocol is a generalization of the previous protocols as it can fall back to direct transmission and the adaptive decode-and-forward protocol in certain scenarios.

It is worth noting that the term cooperative relaying is somewhat non-intuitively applied to our cooperative protocols. Source and relay do not truly cooperate in the sense of mutual benefits; rather, the relay solely serves the source without receiving any benefits. In fact, the key idea is to use combining as done in cooperative relaying, but to sacrifice the symmetry implied by interchangeable roles of source and relay so as to profit from the pathloss savings of an intermediate, i.e., asymmetrically placed, relay node.

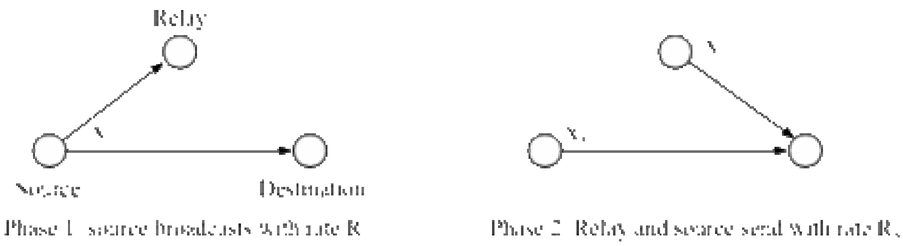


Figure 4-81: Operation of the receive collision protocol [190].

Cooperative Relaying – Distributed Hybrid ARQ

The protocols described so far are characterized by their need for two phases. At the cost of protocol complexity, further improvement can be achieved by avoiding the second phase if the destination has successfully decoded in the first phase. Using feedback from the destination, additional information is provided by the relay *only upon explicit request*. A protocol that minimizes the number of retransmissions can be defined as follows: the source transmits half of its message during the first time slot of duration T=2 in a broadcast manner to relay and destination. Relay and destination both try to decode this message and send feedback on their decoding status in a broadcast manner to the other nodes. Consequently, all nodes have all necessary information to act appropriately during the (optional) second time slot. In case the destination was not able to correctly decode the source message *and* the relay was able to decode the source message, the latter will send additional redundancy in the second time slot. The destination then assembles the complete codeword and retries decoding. In case of decoding failure, a block error is declared. The same occurs when destination and relay simultaneously fail to decode.

This protocol is currently under investigation; first results are promising promising. Due to their preliminary status, result are not included in this report; they will be discussed at a later stage.

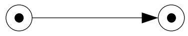
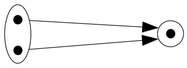
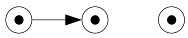
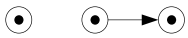
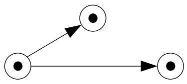
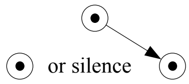
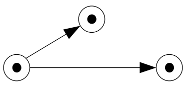
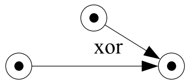
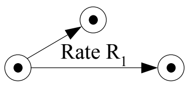
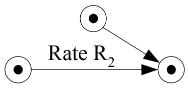
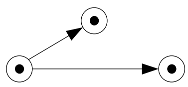
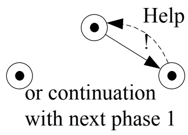
Protocol	**Phase 1**	**Phase 2**	**Pathloss savings**	**Diver-sity order**	**Remark**
Direct			No	1	Known SISO
Transmit diversity			No	2	Alamouti
Conventional relaying			Yes	1	Store-and-Forward
Simple Adaptive DF			Yes	2	Simple extension of conv. relaying
Complex Adaptive DF			Yes	2	Requires feedback
Receive collision			Yes	2	Requires multiuser detection and rate adaptation
Distributed hybrid ARQ			Yes	2	Requires feedback

Figure 4-82: Overview of studied decode-and-forward schemes.

4.2.6.3.2 Results and Performance

An analytical assessment of the protocols of interest is given in [173]. For reasons of compactness, we just provide selected results that are applicable to WINNER.

<u>Outage Probability</u>

We start our discussion by considering the case of a relay being located halfway between source and destination. Figure 4-83 depicts the outage probability as a function of the SNR for all protocols, for a moderate pathloss exponent of 3.0 and a low rate *R=1* bit/s/Hz.

Conventional relaying benefits from a reduction of the end-to-end pathlosses; it saves 1.3 dB over direct transmission. Direct transmission and conventional relaying achieve first-order diversity: the probability of outage decays by one order of magnitude for an increase of 10 dB in SNR.

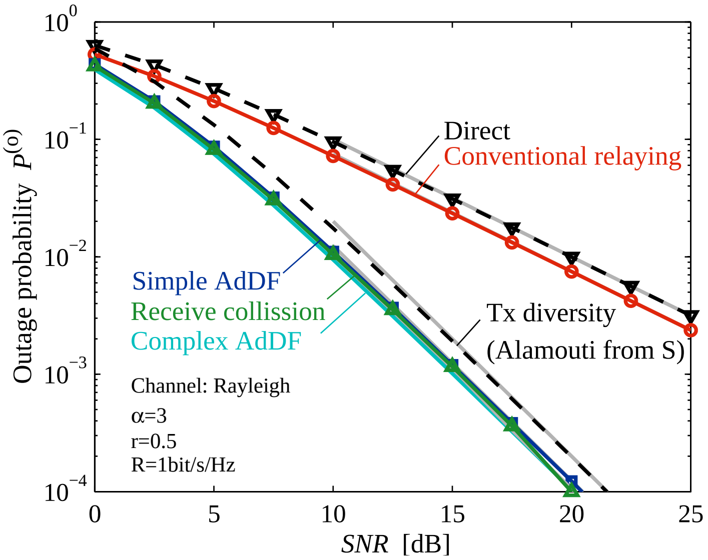


Figure 4-83: Outage probability vs. SNR for a relay located halfway between source and destination. Parameters: pathloss exponent 3.0, spectral efficiency R=1 bit/s/Hz.

By contrast, the transmit diversity scheme as well as the diversity-exploiting cooperative protocols realize second-order diversity, thereby outperforming direct transmission and conventional relaying. Note that both cooperative schemes perform better than the transmit diversity scheme; loosely speaking, these protocols jointly exploit pathloss savings and diversity gains by having one of the two transmit antennas (in the form of the relay) located between source and relay, while conventional transmit diversity has both antennas at the source and faces the full pathloss from source to destination. Next, the similar performance of simple and complex adaptive decode-and-forward scheme suggests that the gains from repetition coding that is employed by the source in the fallback case of the *complex* protocol are not substantial. As we have observed this outcome for all other studied scenarios, we focus on the performance of the proposed simple AdDF scheme in the following.

Both simple and adaptive cooperative protocols yield potential energy savings of approx. 10 dB over direct transmission and *8.7 dB over conventional relaying* under our assumptions, at an outage probability 10^{-2} that would correspond to uncoded scenarios. Finally, the receive collision scheme realizes comparable performance.

Influence of Relay Node Position

To study the impact of relay node location, we vary the position of the relay along the straight line connecting source and destination. The parameter of interest is therefore the relative distance from the source.

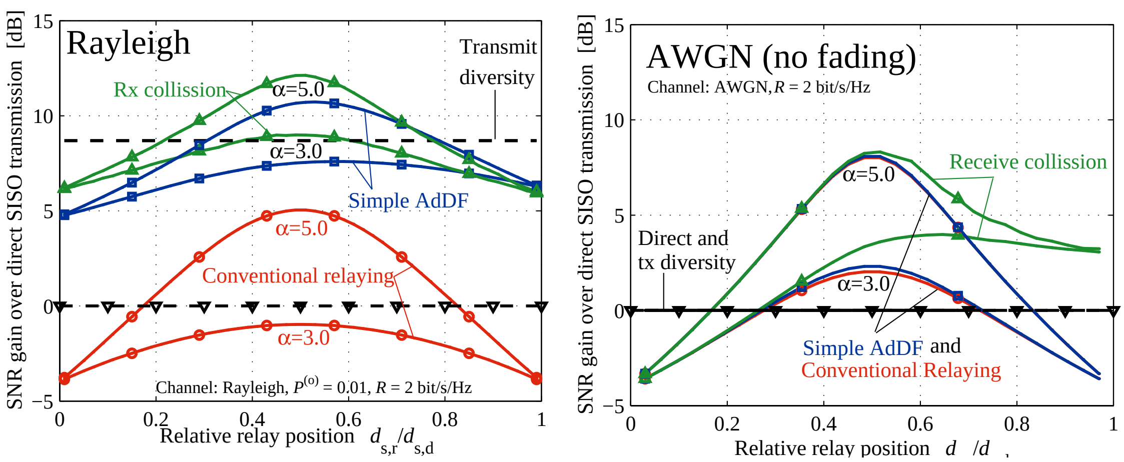


Figure 4-84: SNR gain, i.e., potential power savings, induced by a reduction of the required SNR, as a function of the position of a relay that is located on the straight line between source and destination. Left: Rayleigh fading, Right: AWGN (no fading).

Rayleigh fading: Figure 4-84 (left) depicts the SNR gains, *i.e.*, the potential power savings, as a function of this relative distance for two different pathloss exponents *(3.0 and 5.0)* for Rayleigh fading. As intuition suggests, relaying yields more benefits for stronger propagation losses. For the shown case of *R=2* bit/s/Hz, the simple cooperative protocol can achieve gains up to 10.5 dB over direct transmission; conventional relaying yields 5 dB. At the cost of additional complexity, the receive collision scheme performs best with gains of 12 dB.

Placing the relay halfway between source and destination (or, more precisely, having a routing protocol that chooses a relay close to this location) is well-known to be the best strategy for conventional relaying as this maximizes pathloss savings. The characteristics in Figure 4-84 indicate that this is likewise a good choice for cooperative relaying; in addition, we note that the cooperative schemes' performances are less susceptible to the relay position than that of their conventional counterpart.

AWGN (no fading): The results obtain for AWGN channels (Figure 4-84, right) confirm what can intuitively be expected: the cooperative AdDF protocol and the conventional protocol achieve almost the same performance. By combining the source's and the relay's signal, the cooperative protocol obtains a marginally better performance; yet, these broadcast advantages become apparent for low pathloss exponents (3.0) only. Interestingly, the receive collision protocol achieves a 3 dB gain over direct transmission when the relay is located close to the destination; this gain can be understood as a receive antenna gain that is provided by relay. The relay requires a negligible power for transmission to the destination, while the source can simultaneously transmit new data in the second phase.

These results imply cooperative relaying has more relaxed constraints on routing than it conventional counterpart. Even more interesting, the discussion of the usage region in section 7.3.4 reveals that routing algorithms with link metrics that are based on propagation losses work well for all of the studied relaying protocols. In other words, network protocols for routing, relay selection, and resource assignment that have been designed for conventional relaying are equally applicable for cooperative relaying.

Influence of Spectral Efficiency

Recall that due to the nodes' inability to receive and transmit simultaneously at the same frequency, orthogonal resources must be assigned for reception and transmission at the relay. For networks with *limited available bandwidth*, this calls in turn for a more spectrum-efficient use of the available resources, thereby counteracting against the intended energy-awareness. The results in [188] indicate that the incurred SNR loss is

$$\frac{2^{2R}-1}{2^{R}-1}.$$

For low rates, *i.e.*, $R \leq 1$, the SNR loss is negligible, but for high rates, relay protocols face a loss that increases by 3 dB for each additional bit/s/Hz over direct transmission. By allowing multiplexed transmission and hence a more efficient use of resources, the SNR resulting loss of the receive collision scheme is correspondingly lower.

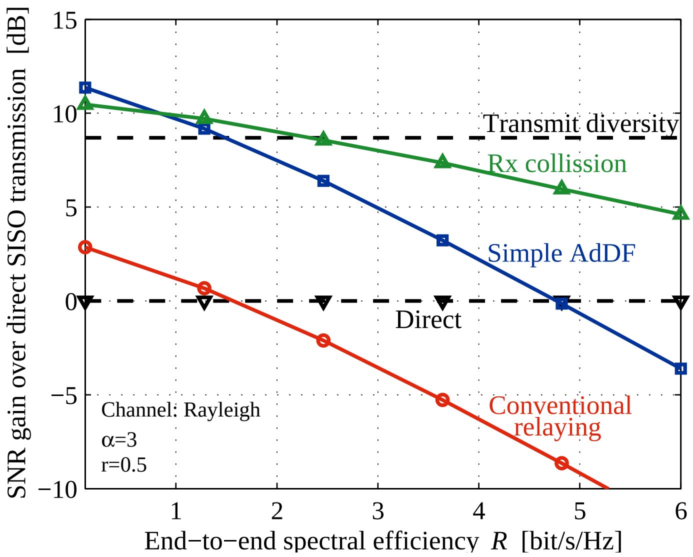


Figure 4-85: SNR gain, i.e., potential power savings, over direct transmission as a function of the end-to-end spectral efficiency R. For the considered bandwidth limited scenarios, the relay protocols suffer from the required doubled spectral efficiency. Parameters: outage probability 10^{-2}, pathloss exponent 3.0, relay halfwayeen source and destination.

More quantitatively, the resulting SNR advantages of relaying systems are depicted in Figure 4-85. Conventional relaying outperforms direct transmission only for $R<1.5$ bit/s/Hz. What is more interesting is that the *simple cooperative protocol achieves gains of approx. 7 dB over conventional relaying* for all spectral efficiencies; it is preferable over direct transmission up to 4.7 bit/s/Hz. Finally, the receive collision protocol advantageously exhibits an SNR loss of only 1 dB per bit/s/Hz compared to 3 dB per bit/s/Hz of the former relaying methods, thereby achieving gains over direct transmission for all considered spectral efficiencies.

4.2.6.3.3 Summary and Conclusion

We have considered various decode-and-forward relaying strategies, ranging from conventional store-and-forward relaying and simple cooperative extensions to complex receive-collision protocols. A simple version of such cooperative schemes was presented.

For purposes of tractable analysis, we have focused on the two-hop case. In general, conventional relaying achieves gains by reducing the end-to-end propagation losses; cooperative relaying makes *additional* profits from exploiting the spatial diversity of the relay channel and the broadcast nature of the wireless medium.

Under our strict assumptions of constrained energy, delay, and bandwidth, conventional relaying is attractive only for low spectral efficiencies and strong propagation losses. For example, for a spectral efficiency $R<2$ bit/s/Hz and a pathloss exponent of *5.0*, conventional relaying can save 5 dB over direct transmission, while cooperative relaying achieves *additional* gains of several decibels (5.5 dB at outage probability 10^{-2}). Due to the higher diversity order of the cooperative protocols, these additional gains increase by 5 dB if the target outage probability is decreased by one order of magnitude. However, in bandwidth-limited systems the gains of all relay protocols reduce by 3 dB per additional bit/s/Hz; if bandwidth-limits are of no concern, then the relay protocols can unconditionally benefit from the discussed advantages. If multi-user detection and flexible rate allocation can be implemented, then receive collision scheme provides gains that even exceed those of the simpler cooperative schemes.

Choosing the relay station such that it is located halfway between source and relay is appropriate for conventional relaying as well as for the cooperative protocols. The usage region of cooperative relaying is much larger than that of it conventional counterpart, hinting at more relaxed requirements on routing.

To summarize, we believe the lion's share of challenges to be related to conventional relaying; cooperative decode-and-forward relaying then comes as a low-complexity extension that promises attractive additional returns at low cost. Beyond that, a new class of receive collision protocols based on parallel transmissions from relay and source offer even more flexible tradeoffs - at the cost of complexity.

4.2.6.4 Adaptive Decode-and-Forward Schemes –Generalization to the Multi-Antenna Case

Fixed wireless relays may have the capability to carry multiple antennas contrary to the common assumption of relays with single antennas as they have been discussed so far. We now investigate the scenario where each relay is equipped with *M* antennas. Again, we consider the decode-and-forward relaying strategy; i.e., the relays are of the regenerative (digital) type. Since by design, the relay is expected to be cost-efficient and of low complexity, the relay may use a subset of these *M* antennas; in the limiting (worst) case scenario a relay uses only 1 antenna through selection combining of its *M* antennas. In this case, only one signal detection chain is required at each relay similar to the relays that do not use multiple antennas. Therefore, in terms of cost, this selection diversity-based relay system incurs no significant penalty (other than the cost of the extra antennas and selection mechanism). If the relays can utilize signals from more than one antennas, then the classical generalized selective combining (GSC) scheme can be employed. For forwarding, however, each relay uses only one antenna. Our earlier work includes the study of hybrid macro/micro-diversity in microcellular networks [216][217].

In this preliminary study, we consider *R* relays deployed in parallel (symmetric network mode) as shown in Figure 4-86. The realistic asymmetric network, where we will be able to exploit the path-loss savings, will be examined in a later phase. We assume homogeneous relaying. At the relays, the threshold decision to forward criterion introduced in [218] has been adopted (this scheme is referred to as “adaptive decode and forward”, AdDF, as well); in this scheme, the proportion of time the relays forward their signals is higher than that in the one antenna case. Hence, we conjecture (for the time being intuitively, without any diversity order attached) that the diversity order can be enhanced. This conclusion is actually derivable from the work in [219]. In that study two intermediate relays (say, *R1* and *R2*) are considered in amplify-and-relay (analog/non-regenerative) type multi-hop networks. It is assumed there that *R2* listens to *R1*, and combines the signal via *R1* with that from the source *S*; the resulting signal is denoted as *R2(S,R1)*. The source’s signal and the forwarded signals from *R1* and *R2* are subsequently utilized at the destination as *S* + *R1* + *R2(S,R1)*, where “+” could be MRC (maximal ratio combining). The diversity order of this scheme is calculated to be 3 in [219].

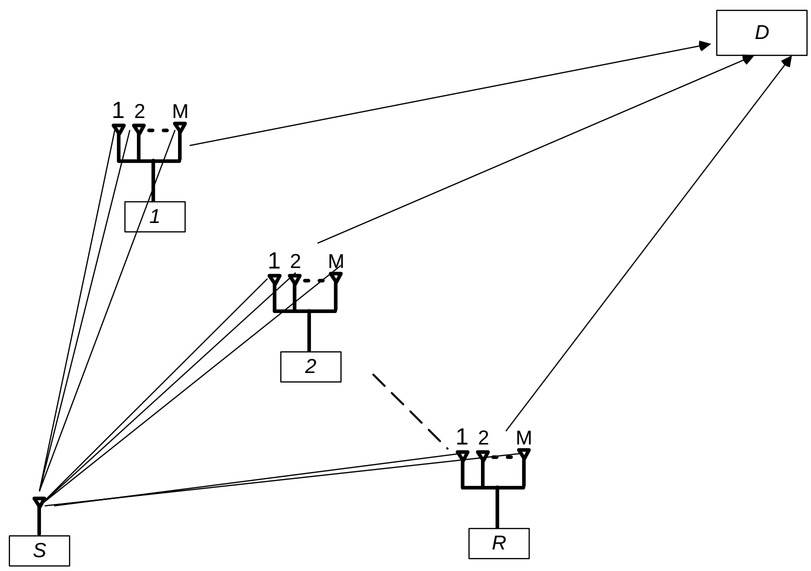


Figure 4-86 The relay layout; each relay has M diversity antennas.

Figure 4-87 shows the performance of the system described above. The modulation scheme considered is BPSK in all the links. The channels are modeled as Rayleigh fading type (shadowing is not considered). We also assumed that the relays are independent with their own power sources; so there is no power sharing among the relays (i.e., the total transmit power increases linearly with the increasing number of relays). However, in order to have accurate measure of the gain accruing from the deployed antenna systems and to have fair comparison with the reference systems, it will be fair to remove any power advantages; subsequent discussions will incorporate this factor.

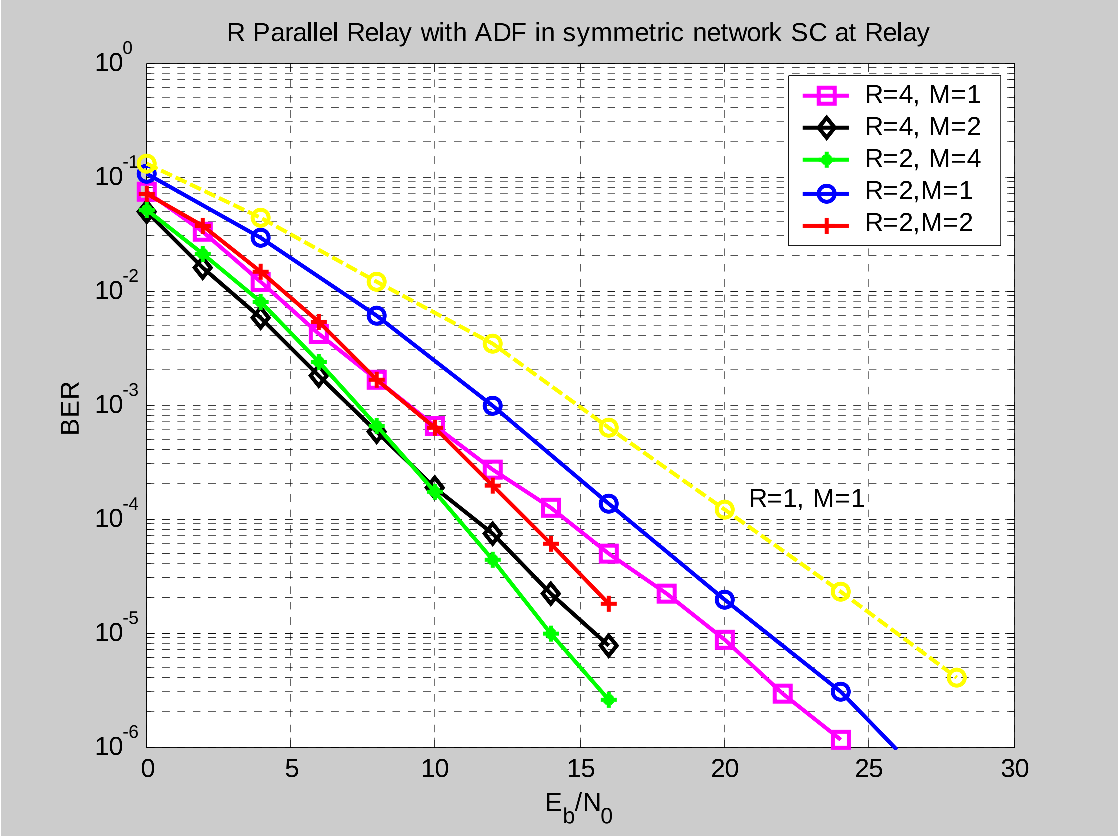


Figure 4-87 The BER vs SNR per bit (dB) performance of R parallel relays with each relay carrying M antennas. The relays operate in selection-combining mode to decode-and-forward based on the received SNR. BPSK is the modulation scheme in all the links which undergo Rayleigh fading.

Figure 4-87 reveals that deploying two relays each with two antennas (the [*R*=2, *M*=2] case) yields a better performance than deploying four relays each with one antenna (the conventional [*R*=4, *M*=1] case). It is worth mentioning that 2 detection resources are required in the [*R*=2, *M*=2] case whereas the [*R*=4, *M*=1] case necessitates 4 detection resources; besides deploying a relay is costlier than installing one microdiversity antenna. Considering the [*R*=4, *M*=2] and [*R*=2, *M*=4] cases as well, one can draw the conclusion that deploying microdiversity at relays may end up in considerable savings in the number of relays to be deployed in a given area.

4.2.6.5 Cooperative Relaying Based on Alamouti Diversity under Aggregate Relay Power Constraints

In this section, a cooperative relaying scheme based on Alamouti diversity [225] is investigated. Whereas this scheme has been studied in [220][221] by assuming that the powers of two relays are constrained individually, in the proposed scheme the power allocation optimization is done under aggregate relay power constraints. The rationality for a relay power constraint is that the aggregate relay transmit power should be kept at lowest level, for a given performance, as interference can be minimized. The optimization criteria of maximizing signal-to-noise-ratio (SNR) and/or average capacity are used.

4.2.6.5.1 System model

The communication link, between a transmitter and a receiver, deploying two cooperating relays is modeled as shown in Figure 4-88. The Alamouti diversity scheme is performed with respect to these two relays, which means that the two relays act as the two antennas in the traditional Alamouti scheme. The transmitter sends a zero mean complex Gaussian signal with power P_{TX} over the frequency flat channels $h_{1,k}$ towards the relays $k \in \{1,2\}$, where complex Gaussian noise (and interference) $z_{RS,k}$, with variance $\sigma^2_{RS,k}$ is added. Each relay first amplifies the received noisy signal or regenerates the signal, then applies the Alamouti scheme between two relays. Subsequently, each relay transmits the coded signal with a power P_k over channels $h_{2,k}$ (also frequency flat) towards the receiver, where complex Gaussian noise (and interference) $z_{RX,k}$ with variance $\sigma^2_{RX,k}$ is added. In Figure 4-88, a_1 and a_2 are the complex-valued amplifying factor of relay 1 and relay 2 respectively. The aggregate relay power, $\sum_k P_k$, is a constant and denoted P_{RX}. It is further assumed that power and phase parameters are adjustable in each relay and that antenna gains are included in the channels.

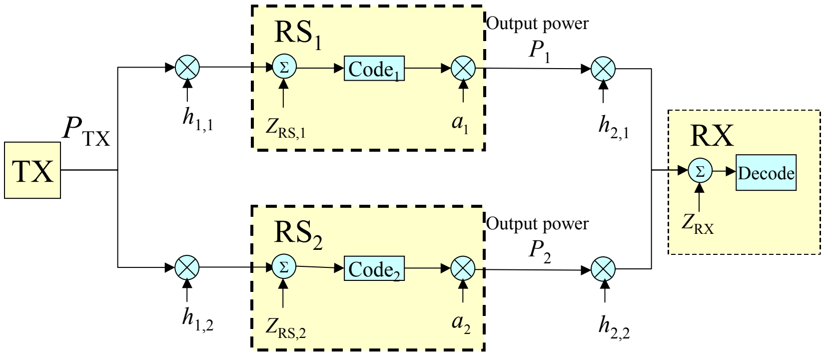


Figure 4-88 System model for Alamouti diversity relaying under aggregate relay power constraints.

4.2.6.5.2 Theoretical Analysis

It can be easily shown that the SNR expression in the non-regenerative case can be expressed as:

$$SNR = \frac{P_{TX} \cdot \left|h_{2,1} \cdot a_1 \cdot h_{1,1}\right|^2 + P_{TX} \cdot \left|h_{2,2} \cdot a_2 \cdot h_{1,2}\right|^2}{\left|h_{2,1} \cdot a_1\right|^2 \cdot \sigma^2_{RS,1} + \left|h_{2,2} \cdot a_2\right|^2 \cdot \sigma^2_{RS,2} + \sigma^2_{RX}} \tag{35}$$

In fact, the regenerative case can derived easily from Equation(35), since it can be viewed as a special case of non-regenerative relaying concept.

Assuming that the instantaneous channel state information (CSI) of two hops is available at the two relays. Then the power of every relay is constrained individually and a_1 and a_2 are set as follows [157][158]:

$$a_k = \frac{h^*_{1,k}}{\left|h_{1,k}\right|} \cdot \sqrt{\frac{P_k}{\left|h_{1,k}\right|^2 \cdot P_{TX} + \sigma^2_{RS,k}}}, k \in \{1,2\} \tag{36}$$

Substituting (36) into (35), the SNR can be expressed as

$$SNR = \frac{\dfrac{P_1 \cdot P_{TX} \cdot \left|h_{2,1} \cdot h_{1,1}\right|^2}{\left|h_{1,1}\right|^2 \cdot P_{TX} + \sigma^2_{RS,1}} + \dfrac{P_2 \cdot P_{TX} \cdot \left|h_{2,2} \cdot h_{1,2}\right|^2}{\left|h_{1,2}\right|^2 \cdot P_{TX} + \sigma^2_{RS,2}}}{\dfrac{P_1 \cdot \left|h_{2,1}\right|^2 \cdot \sigma^2_{RS,1}}{\left|h_{1,1}\right|^2 \cdot P_{TX} + \sigma^2_{RS,1}} + \dfrac{P_2 \cdot \left|h_{2,2}\right|^2 \cdot \sigma^2_{RS,2}}{\left|h_{1,2}\right|^2 \cdot P_{TX} + \sigma^2_{RS,2}} + \sigma^2_{RX}} \tag{37}$$

Since the aggregate relay powers are constrained as follows:

$$P_1 + P_2 = P_{RS} \tag{38}$$

the SNR in (37) has the following form:

$$SNR = \frac{P_1 \cdot a + P_2 \cdot b}{P_1 \cdot c + P_2 \cdot d + e} \tag{39}$$

Where a, b, c, d and e are the corresponding items in (37). By combining (38) and (39), the maximizing of the instantaneous SNR, can be simply done by taking the partial derivative of SNR with respect to P_1 and P_2.

- If $(a \cdot d - b \cdot c) \cdot P_{RS} + (a \cdot e - b \cdot e) > 0$, the total power should be assigned to relay 1;
- If $(a \cdot d - b \cdot c) \cdot P_{RS} + (a \cdot e - b \cdot e) < 0$, the total power should be assigned to relay 2;
- If $(a \cdot d - b \cdot c) \cdot P_{RS} + (a \cdot e - b \cdot e) = 0$, the total power could be arbitrarily assigned to one of the relays.

It can be seen that maximizing the instantaneous channel capacity is equivalent to maximize the instantaneous SNR. In fact the Shannon capacity is expressed as

$$C = \log_2 (1 + SNR) \tag{40}$$

The regenerative case

In this case, the two relays first regenerate the received signal, then amplify it and use the Alamouti scheme, and finally forward it to the receiver.

The sum of two relays power is constrained as follows.

$$|a_1|^2 \left(|h_{1,1}|^2 P_{TX} + \sigma^2_{RS,1}\right) + |a_2|^2 \left(|h_{1,2}|^2 P_{TX} + \sigma^2_{RS,2}\right) = P_{RS} \tag{41}$$

To simplify the analysis with help variables such as b_k, a_k and P_k might be chosen to be

$$a_k = \frac{\sqrt{P_{RS}} \cdot b_k}{\sqrt{\sum_{k=1}^{2} |b_k|^2}} \cdot \frac{1}{\sqrt{|h_{1,k}|^2 P_{TX} + \sigma^2_{RS,k}}} \tag{42}$$

$$P_k = \frac{P_{RS} \cdot |b_k|^2}{\sum_{k=1}^{2} |b_k|^2} \tag{43}$$

Then, (1) can be expressed as

$$SNR = \frac{\sum_{k=1}^{2} |b_k|^2 \cdot \frac{\Gamma_{RS,k} \cdot \Gamma_{RX,k}}{1 + \Gamma_{RS,k}}}{\sum_{k=1}^{2} |b_k|^2 \cdot \left(\frac{\Gamma_{RX,k}}{1 + \Gamma_{RS,k}} + 1\right)} \tag{44}$$

where $\Gamma_{RX,k} = |h_{2,k}|^2 \cdot P_{RS} / \sigma^2_{RX}$ and $\Gamma_{RS,k} = |h_{1,k}|^2 \cdot P_{TX} / \sigma^2_{RS,k}$.

In the regenerative case, assuming that the relays can perfectly regenerate the transmitted signal, i.e., $\sigma^2_{RS,k} = 0;\ |h_{1,k}|^2 = 1$, the resulting SNR in (44) is then

$$SNR = \frac{\sum_{k=1}^{2} |b_k|^2 \cdot \Gamma_{RX,k}}{\sum_{k=1}^{2} |b_k|^2} \tag{45}$$

To get the optimal power allocation between two relays is actually to solve b_k under a given optimization criterion.

By maximizing the instantaneous SNR in (45), it can be found that the total power should always be assigned to one relay from which to the receiver there is a relatively good channel condition in terms of $\left|h_{2,k}\right|^2$. By maximizing the instantaneous channel capacity, the same conclusion can be drawn.

<u>The instantaneous CSI of the 1st hop and the average CSI of the 2nd hop are available at two relays:</u>

The mean SNR and mean channel capacity can be derived similarly to the case where the instantaneous CSI in available at the 2nd hop.

4.2.6.5.3 Numerical results and analysis

For the regenerative case, the power allocation coefficients, i.e. $|b_i|^2/(|b_1|^2+|b_2|^2); i=1 \text{ or } 2$, may not always fall into (0, 1) with the maximization of the mean channel capacity. In Figure 4-89 and Figure 4-90, "Avg. path-gain to relay 1(2)" means the average path-gain between the receiver and the relay 1(2) (the 2nd hop).

Figure 4-89 shows the mean channel capacity for the Maximization of the mean SNR optimization scheme. Note that two P_{RS}/σ_{RX}^2 values are considered.

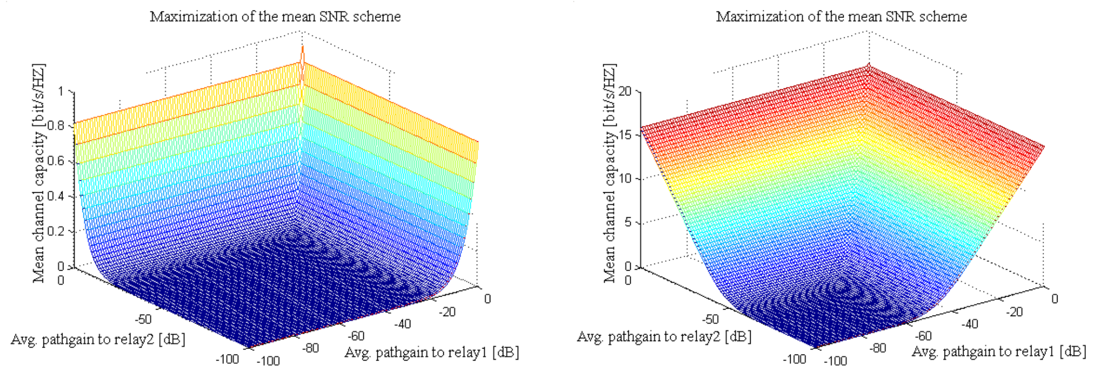


Figure 4-89 Mean channel capacity with the "Maximization of the mean SNR "scheme for different P_{RS}/σ_{RX}^2 values (left plot: 0dB, right plot: 50 dB).

Figure 4-90 visualizes the difference between the "Maximization of the mean channel capacity" scheme and the "Maximization of the mean SNR" scheme. In fact, the two schemes coincide with each other and result in almost identical mean channel capacity in most regions. The difference is mostly less than 0.15 dB for low P_{RS}/σ_{RX}^2 value (0 dB) and 0.25 dB for high P_{RS}/σ_{RX}^2 value (50 dB).

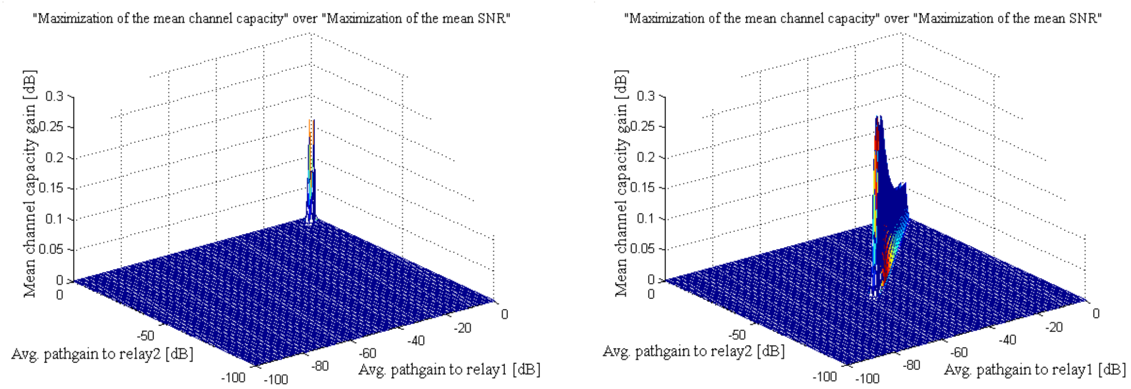


Figure 4-90 Mean channel capacity gain for different P_{RS}/σ_{RX}^2 values (left plot: 0dB, right plot: 50 dB), the "Maximization of the mean channel capacity" scheme over the "Maximization of the mean SNR" scheme.

4.2.6.5.4 Summary and Conclusion

When the relays have the instantaneous CSI of both the 1st and the 2nd hop, the maximization of the channel capacity is equivalent to the maximization of the SNR at the receiver. The analysis shows that the total power should be allocated to one of the two relays to achieve the maximum channel capacity, either for the non-regenerative or the regenerative relays.

When the relays have the instantaneous CSI of the 1st hop and the average CSI of the 2nd hop, the efforts focus on the regenerative relays while assuming the relays could perform decoding perfectly. Two optimization schemes, i.e. the maximization of the mean channel capacity, the maximization of the mean SNR are investigated. The "Maximization of the mean SNR" scheme still leads to the conclusion that the total power should be allocated to one of the two relays. The "Maximization of the mean channel capacity" is equivalent to the "Maximization of the mean SNR" scheme, and the two schemes give almost the same mean channel capacity.

4.2.6.6 Impact of System Resource Constraints on Wireless Relaying Channels

Each of the distributed spatial diversity techniques discussed so far places different requirements on system resources due to a reliance on mesh connectivity between terminals. Therefore, system resource constraints that limit the terminal connectivity constrain the distributed spatial diversity techniques that can be applied. The system resource constraints considered in this section are the available number of orthogonal relaying channels, the ability of relays to diversity combine incident signals, the ability of the destination to combine incident signals, the ability of receivers to combine signals on multiple orthogonal channels, and the ability of transmitters to transmit signals on multiple orthogonal channels.

Analysis of the impact of these system resource constraints on the connectivity and performance of relaying channels proceeds according to the following methodology. The relevant system resource constraints are introduced and motivated, all possible constraint combinations are analyzed and reduced where appropriate, the set of system connectivity models resulting from these constraint combinations are derived and associated with their minimum complexity constraint sets, and these system connectivity models are compared via simulation results.

4.2.6.6.1 System Resource Constraints

The system resource constraints are described in detail in this section. The motivation for each constraint is discussed in terms of system complexity and cost. Options for each constraint are introduced, along with their corresponding relative cost and connectivity impact. In all cases, constraint options with lower cost have higher connectivity impact. Connectivity impact is defined in comparison to a fully connected system with links between all terminals.

Number Channels Available (NCA): This constraint defines the number of orthogonal relaying channels available for the transmission of a signal between a single source-destination pair. The half-duplex nature of wireless terminal hardware requires that each relay transmit and receive with different channels, implying a minimum of two orthogonal channels. Use of more than two orthogonal channels increases the system cost since more bandwidth is necessary for each source-destination pair.

- N Channels Available (NCA): The source and each relay transmit using a separate orthogonal channel. There is no connectivity impact.
- 2 Channels Available (2CA): The source and all relays transmit using the same two orthogonal channels. The connectivity impact is that receivers may only be connected to transmitters on the opposite channel (an odd number of hops away).

Relay Combination (RC): This constraint defines the ability of relay terminals to diversity combine incident signals from multiple preceding terminals. Use of relay diversity combination increases the system cost since separate combination hardware is required for each relayed signal.

- Relay Combination (RC): Relays are able to diversity combine incident signals from multiple preceding terminals. There is no connectivity impact.
- No Relay Combination (NRC): Relays are not able to diversity combine incident signals from multiple preceding terminals. The connectivity impact is that relays may only be connected to one transmitter.

Destination Combination (DC): This constraint defines the ability of destination terminals to diversity combine incident signals from multiple preceding terminals. Use of destination diversity combination increases the system cost since combination hardware is required for received signals.

- Destination Combination (DC): Destinations are able to diversity combine incident signals from multiple preceding terminals. There is no connectivity impact.

- No Destination Combination (NDC): Destinations are not able to diversity combine incident signals from multiple preceding terminals. The connectivity impact is that destinations may only be connected to one transmitter.

Multiple Channel Reception (MCR): This constraint defines the ability of receivers to diversity combine signals that are on different orthogonal channels. Use of multiple channel reception increases the system cost since more complex multiple channel combination hardware is required.

- Multiple Channel Reception (MCR): Receivers are able to diversity combine signals on different orthogonal channels. There is no connectivity impact.
- Single Channel Reception (SCR): Receivers are not able to diversity combine signals on different orthogonal channels. The connectivity impact is that receivers may only be connected to a subset of transmitters that use one common channel.

Multiple Channel Transmission (MCT): This constraint defines the ability of transmitters to concurrently transmit on multiple orthogonal channels. Use of multiple channel transmission increases the system cost since more complex multiple channel transmission hardware is required.

- Multiple Channel Transmission (MCT): Transmitters are able to concurrently transmit on multiple orthogonal channels. There is no connectivity impact.
- Single Channel Transmission (SCT): Transmitters are not able to concurrently transmit on multiple orthogonal channels. The connectivity impact is that transmitters may only be connected to a subset of receivers that use one common channel.

4.2.6.6.2 Constraint Combinations

The possible system resource constraint combinations are analyzed in this section. The combinations are reduced where practical considerations deem appropriate. A set of resultant system connectivity models is derived from the combinations. Figure 4-91 summarizes the constraint combinations with the resultant system connectivity models. A detailed derivation of the connectivity models is not included due to space limitations. The acronyms used in Figure 4-91 are those introduced in the previous section. The following terminology is used when describing the resultant system connectivity models:

- Single Relay (1R): Relays connected to one transmitter.
- 2Chnl Relay (2R): Relays connected to the subset of transmitters on one channel.
- Full Relay (FR): Relays connected to all transmitters previous along the transmission path.
- Single Destination (1D): Destination connected to one transmitter.
- 2Chnl Destination (2D): Destination connected to the subset of transmitters on one channel.
- Full Destination (FD): Destination connected to all transmitters.
- 2Chnl Source (2S): Source connected to the subset of relays on one channel and the destination on the other.
- Full Source (FS): Source connected to all receivers.

NCA	RC	DC	MCR	MCT	Resultant Model
NCA	RC	DC	MCR	MCT	FRFD
NCA	RC	DC	MCR	SCT	FRFD
NCA	RC	DC	SCR	MCT	FRFD
NCA	RC	DC	SCR	SCT	1R1D[2]
NCA	RC	NDC	MCR	MCT	Impractical Model[1]
NCA	RC	NDC	MCR	SCT	Impractical Model[1]
NCA	RC	NDC	SCR	MCT	Impractical Model[1]
NCA	RC	NDC	SCR	SCT	Impractical Model[1]
NCA	NRC	DC	MCR	MCT	1RFD
NCA	NRC	DC	MCR	SCT	1RFD
NCA	NRC	DC	SCR	MCT	1RFD
NCA	NRC	DC	SCR	SCT	1R1D[2]

NCA	NRC	NDC	MCR	MCT	1R1D
NCA	NRC	NDC	MCR	SCT	1R1D
NCA	NRC	NDC	SCR	MCT	1R1D
NCA	NRC	NDC	SCR	SCT	1R1D
2CA	RC	DC	MCR	MCT	2RFDFS
2CA	RC	DC	MCR	SCT	2RFD
2CA	RC	DC	SCR	MCT	2R2DFS
2CA	RC	DC	SCR	SCT	2R2D
2CA	RC	NDC	MCR	MCT	Impractical Model[1]
2CA	RC	NDC	MCR	SCT	Impractical Model[1]
2CA	RC	NDC	SCR	MCT	Impractical Model[1]
2CA	RC	NDC	SCR	SCT	Impractical Model[1]
2CA	NRC	DC	MCR	MCT	1RFD
2CA	NRC	DC	MCR	SCT	1RFD
2CA	NRC	DC	SCR	MCT	1R2D2S
2CA	NRC	DC	SCR	SCT	1R2D
2CA	NRC	NDC	MCR	MCT	1R1D
2CA	NRC	NDC	MCR	SCT	1R1D
2CA	NRC	NDC	SCR	MCT	1R1D
2CA	NRC	NDC	SCR	SCT	1R1D

Figure 4-91: System Resource Constraint Combinations

[1] Constraint combinations with relay diversity combination but not destination diversity combination are unlikely to occur in practice since it is expected that diversity combination hardware will be used on signals where the terminal is the consuming destination before it is used on signals where the terminal is only a relay.

[2] Constraint combinations with destination (and optionally relay) diversity combination but single channel reception and single channel transmission actually achieve less connectivity when each relay transmits on a separate orthogonal channel than when the source and all relays transmit on the same two channels. Intelligent use of the available channels results in identical connectivity to a system constrained to two available channels.

4.2.6.6.3 Resultant System Connectivity Models

The set of system connectivity models resulting from the constraint combinations are described in this section. For each model the set of constraints that achieves the corresponding system connectivity while minimized the system cost (the minimum cost constraint set) and a graphical example illustrating the connectivity of the model are presented. Figure 4-92 shows the transition between the different system connectivity models for various constraint changes.

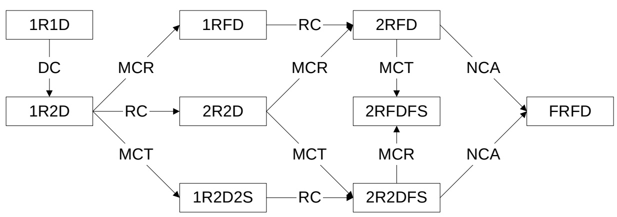


Figure 4-92: Transition between System Connectivity Models

Transitions are in the direction of decreased system resource constraints. Transitions that do not improve the system connectivity or that do not follow the minimum cost constraint sets are not shown. For example, adding NCA to the 1R1D model provides no improvement and adding MCR to the 1R2D2S model causes MCT to be redundant.

In section 7.3.5, we provide a detailed discussion of examples for each system connectivity model.

4.2.6.6.4 Simulation Results

The system connectivity models are applied in a series of simulations that provide a comparison with respect to probably of error. These simulations allow the performance impact of the individual system resource constraints to be isolated. Although a BPSK modulation scheme is used for the simulations, the qualitative statements derived from the simulation results generalize to other modulations schemes. The simulations use the example models shown in Figure 7-21 to Figure 7-29 and were generated using the equations derived in [228][229]. The topology is symmetric with normalized link distances:

$d_{S,D} = \frac{1}{3} d_{S,R1} = \frac{1}{2} d_{S,R3} = \frac{3}{4} d_{S,R4} = \frac{1}{2} d_{R1,R2} = \frac{1}{2} d_{R1,R4} = \frac{2}{3} d_{R1,R5}$.

Figure 4-94 to Figure 4-96, respectively compare the BER of the system connectivity models for amplified relaying, for decoded relaying where errors at relays propagate as a decoding error at the destination, and for decoded relaying where errors at relays do not propagate but are ignored by other receivers. Decoded relaying without error propagation upper bounds the performance of decoded relaying and corresponds to an error-free version of the adaptive decode-and-forward protocols presented in [231][233][234][235].

Comparison of the 1R1D model with the 1R2D model isolates the impact of destination combination. Comparison of the 2R2DFS, 2RFD, and 2RFDFS system connectivity models with the FRFD model isolates the impact of the number of channels available. Comparison of the 1R2D, 1RFD, and 1R2D2S system connectivity models with the 2R2D, 2RFD, and 2R2DFS models respectively isolates the impact of relay combination. Comparison of the 1R2D, 2R2D, and 2R2DFS system connectivity models with the 1RFD, 2RFD, and 2RFDFS models respectively isolates the impact of multiple channel reception. Comparison of the 1R2D, 2R2D, and 2RFD system connectivity models with the 1R2D2S, 2R2DFS, and 2RFDFS models respectively isolates the impact of multiple channel transmission. Figure 4-93 shows the impact of the constraints for each relaying method.

Relaying Method	NCA	RC	DC	MCR	MCT
Amplified	Small	Medium	Large	Large	Medium
Decoded w Prop	Small	Medium	Small	Small	Medium
Decoded w/o Prop	Small	Large	Medium	Large	Medium

Figure 4-93: Impact of System Resource Constraints

The impact of the constraints is different for amplified relaying, decoded relaying with error propagation, and decoded relaying without error propagation. The results indicate that for amplified relaying the priority is to maximize the connectivity of the destination terminal, while for decoded relaying (with and without error propagation) the priority is to equalize the connectivity of the destination and relay terminals. For amplified relaying the diversity order of the system is dependent on the connectivity of the destination. For decoded relaying with error propagation the diversity order of the system is constrained by the connectivity of the minimally connected relay and therefore limited to one. For decoded relaying without error propagation the diversity order of the system is dependent on the connectivity of both the destination and relays. Decoded relaying without error propagation is able to achieve diversity order equal to amplified relaying for system connectivity models that include relay combination.

The results provide guidance for the order in which the various constraints should be lifted to maximize connectivity and performance. For amplified relaying the constraints should be lifted in the following order: DC, MCR, RC, and MCT. For decoded relaying the constraints should be lifted in the following order: DC, RC, MCT, and MCR. For ideal decoded relaying the constraints should be lifted in the following order: DC, RC, MCR, and MCT. Since the impact of lifting the NCA constraint when all other constraints are lifted is small, but increases the system cost significantly, it is not expected that having the source and relays transmit using separate orthogonal channels will be implemented in practice. Although not shown due to space limitations, this constraint ordering holds for other topologies and channel allocations.

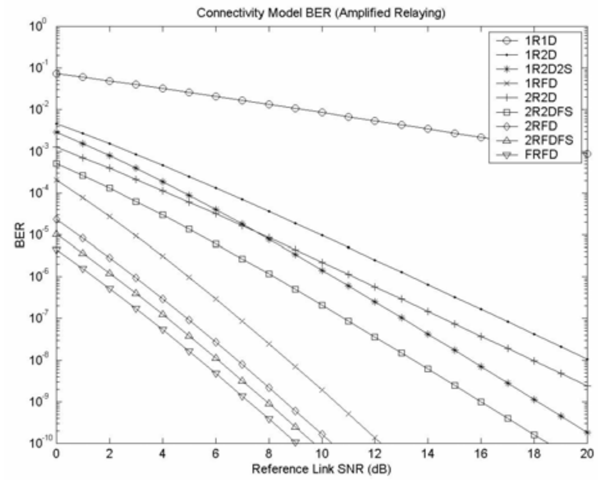


Figure 4-94: Amplified Relaying Connectivity Models

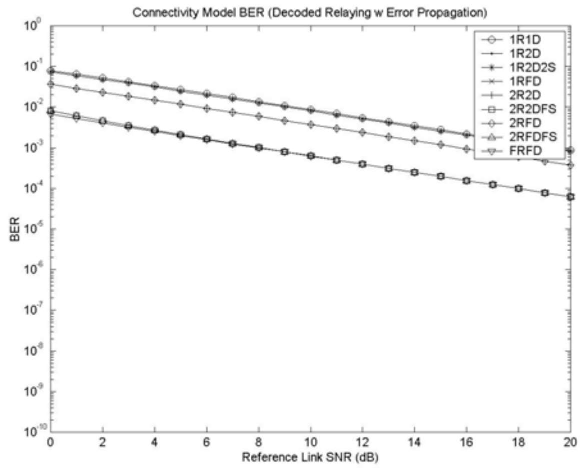


Figure 4-95: Decoded Relaying w Propagation Connectivity Models

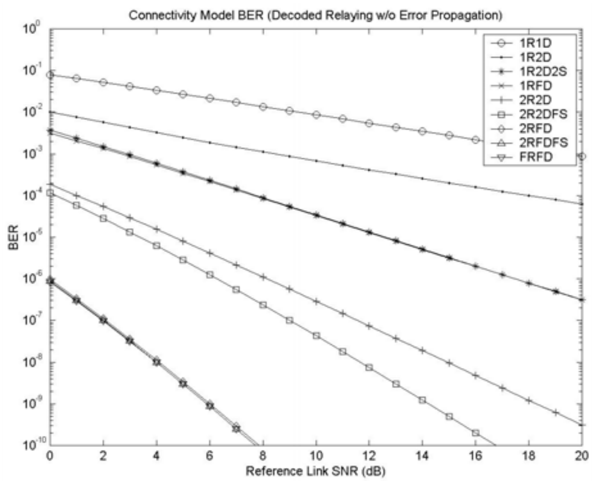


Figure 4-96: Decoded Relaying w/o Propagation Connectivity Models

4.2.6.6.5 Summary and Conclusion

The concept of cooperative relaying considers the impact of various system resource constraints on the connectivity and performance of relaying channels. System connectivity models are derived from the constraint combinations, and minimum cost constraint sets are given. The relationship of the models to the distributed spatial diversity techniques presented in the literature [227]-[229][232][235][237][238] is discussed. Simulations are presented that show the impact of the individual system resource constraints on amplified relaying, decoded relaying with error propagation, and decoded relaying without error propagation.

The impact is indicated to be different for the various relaying methods. The results indicate that for amplified relaying the priority is to maximize the connectivity of the destination terminal, while for decoded relaying (with and without error propagation) the priority is to equalize the connectivity of the destination and relay terminals. For amplified relaying the diversity order of the system is dependent on the connectivity of the destination. For decoded relaying with error propagation the diversity order of the system is constrained by the connectivity of the minimally connected relay and therefore limited to one. For decoded relaying without error propagation the diversity order of the system is dependent on the connectivity of both the destination and relays. Guidance is provided for the order in which the various constraints should be lifted to maximize connectivity and performance.

4.2.6.7 Cooperative Mobile Relaying

For mobile relays, additional aspects need to be considered, due to the mobility factor which complicates much more the whole concept of cooperative relaying. A simple example is that in fixed relays, the AP-Relay channel is not drastically changing which is not the case for mobile relays. Thus, it is imperative that we must implement those mechanisms which will fast and reliably evaluate the required channel if any metrics related to that are required. On the other hand, we might even gain from the mobility in the sense that the reception levels of the AP→Relay path might in different locations of the mobile relay be quite favourable

In general, cooperative relaying for mobile relays (cooperative mobile relaying) could be a quite complicated task, as long as we need to take into account a number of parameter of a dynamically changing environment.

Thus, as in the case of mobile relays, for cooperative relaying also we should first analyse what are the implications of mobility and what the possible scenarios that cooperative relaying could be used at and then start to evaluate simple approaches to see what the incremental gain is. Then we can “move” to more complex approaches. However, we expect that in the end the factors of mobility will pose limitations to the level of complexity we can introduce and in the end simple schemes might be preferable.

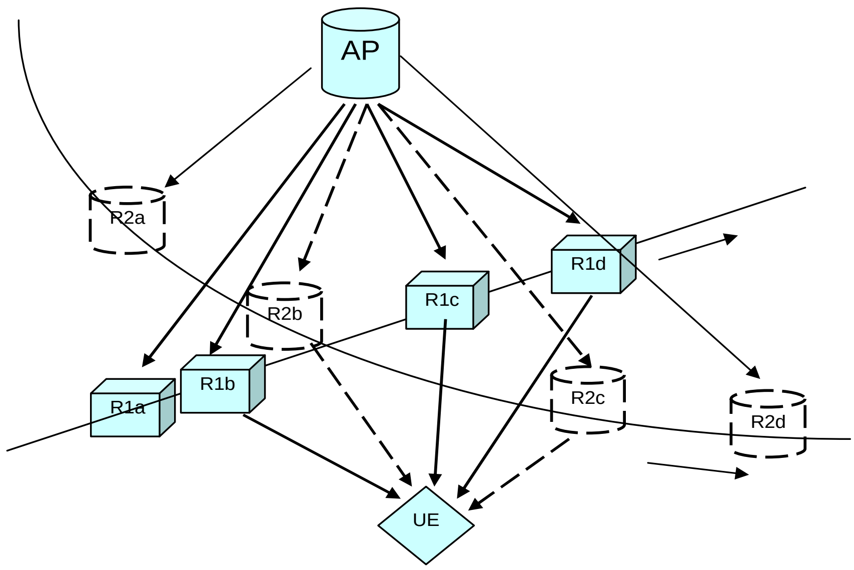


Figure 4-97 Cooperative relaying for mobile relays

What Figure 7-18 shows is a typical cooperative mobile relaying approach. It pictures one AP, two relays R1 and R2 and one UE. We assume that these relays follow a specific trajectory of a straight line and a curved line. We assume four positions for each relay and each of the four positions we name R1a/b/c/d and R2a/b/c/d. Straight lines are links through Relay1 and dotted lines are from Relay2.

What we show is that out of the four positions for each relay not all are used. For Relay1 and Relay2, three and two positions are used respectively. The other locations are far away from the UE to “add value” to the reception levels at the UE.

4.3 Heterogeneous Relays

Due to the few works regarding heterogeneous relay stations, an important objective of Task 3.4 before developing the new deployment concepts will be to study and analyse some aspects like its functionality (basic integrating parts), the scenarios and circumstances in what the utilization of heterogeneous relay station could be useful (comparing with other possible and current alternatives), the complexity of the interworking and congestion control mechanisms for different combination of radio interfaces (in particular for the different modes of the same radio access technology to develop inside WINNER), and the feasibility and compatibility between the selected radio interfaces implicated in the heterogeneous relay station.

So, the main objective of this section is to present a clear idea of heterogeneous relay concepts. It will review the definition of this new element in the context of a radio network as well as other important aspects as the different classes, the different uses and the identification of the scenarios where the use of these new elements could be useful, and the main integrating parts that this element should have in order to reach its performance. Further it will be introduced some of the particular protocol functions which concern to the heterogeneous relaying like the radio resources management and the routing/forwarding functions, and finally the possibility to use different air interfaces for cooperative relaying will be discussed, including the raisons which have imposed the abandon of future investigations for this topic inside the framework of WINNER.

For the definition of heterogeneous relays a good understanding of the terms RAN and RAT, as defined in Section 1.2, is necessary.

4.3.1 Basic Concept

From the technical annex of WINNER, the heterogeneous relay was defined as a network node with relaying capabilities that is wirelessly connected to an access point, other relay station and/or user terminal, using different radio interface technologies for its connections. Here an access point is understood like a network node terminating the physical layer, (parts of) the link layer, and possibly also parts of the network layer of the radio interface technology from the network side. Moreover, the AP is the node closest to the core (backbone) network that an user terminal may be (directly) connected to.

The radio interfaces involved in a heterogeneous relay station can be either two modes of the new WINNER radio interface as shows Figure 4-98, or comprise a legacy radio interface as one of these links. Further shown in the same figure is how relaying concepts can cover indoors from outdoors based on relay stations affixed to buildings, which would allow operating the same user terminals indoors and outdoors.

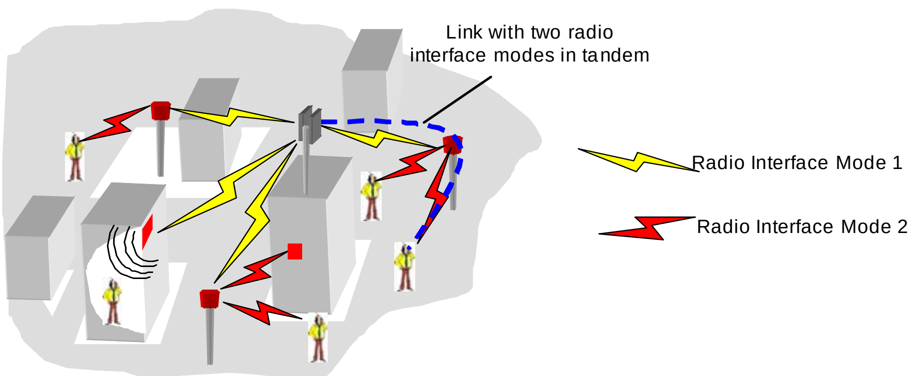


Figure 4-98: Example from TA with heterogeneous relay stations using two different radio interfaces modes in tandem

Taking the WWRF/WG4 Relaying Subgroup white paper [242] the following charactistics of a heterogenous relay have been defined in WINNER:

- The use of different spectrum for the relay to user links and the base to user links, not always involved heterogeneous relay. For instance, GSM 900 and GSM 1800 use different bands but have the same protocol, coding scheme and modulation method, so they use the same radio technology and the same mode. Therefore this example corresponds clearly to a homogeneous relay station.

- However, when talking about the possibility to communicate the users with the relay via the standard 802.11, while the communication between the relay and the base station uses a given cellular technology, that is, the relay has two different standards for its connections and so two different radio telecommunication technologies, we are in the heterogeneous concept. Moreover in this case it is impossible to implement a relay of amplify & forward type since it is necessary to handle the signals at level upper than physical layer and to do a protocol conversion in order to forward the signal in the other standard.
- The heterogenous relay is not possible as L1 relay or repeater because at least some mapping mechanisms between the two different L1s and therewith store and forwarding would be required. .
- It is not uncommon to use wireless links to connect the BSs to the backbone network (i.e. it is not unusual to use wireless links in the transport network).
- In some cases (e.g. for a very slimmed BS and distributed operation) the amount of functionality of BS and relay station may be very similar.

All these characteristics resuted in the heterogenous relay definintion as written in Section 1.2.

Nevertheless the fact to include in the heterogeneous relay definition the possibility that the relay might be fed by a Transport Node, can give rise to ambiguity since this wireless link corresponds really to the connection between an AP and a Transport Node, and therefore that is part of the transport network. In this way, the final users may not access this network using the wireless link, which is specifically designed for feeding wirelessly the AP. In any case, investigation of wireless links in the transport network are still a valid research issue inside WINNER and in particular inside WP3, since the wireless connection of APs with the transport network is other important aspect to take into account for the new network deployment concepts to develop in WINNER project. Later on this chapter (section called “Others”) is dealt this wireless link between AP and transport network, with the term “Wireless Feeder System”

Assuming the preliminary radio network architecture developed in AN project, we could separate the heterogeneous relay concept and the wireless feeder concept, as it is shown in Figure 4-99.

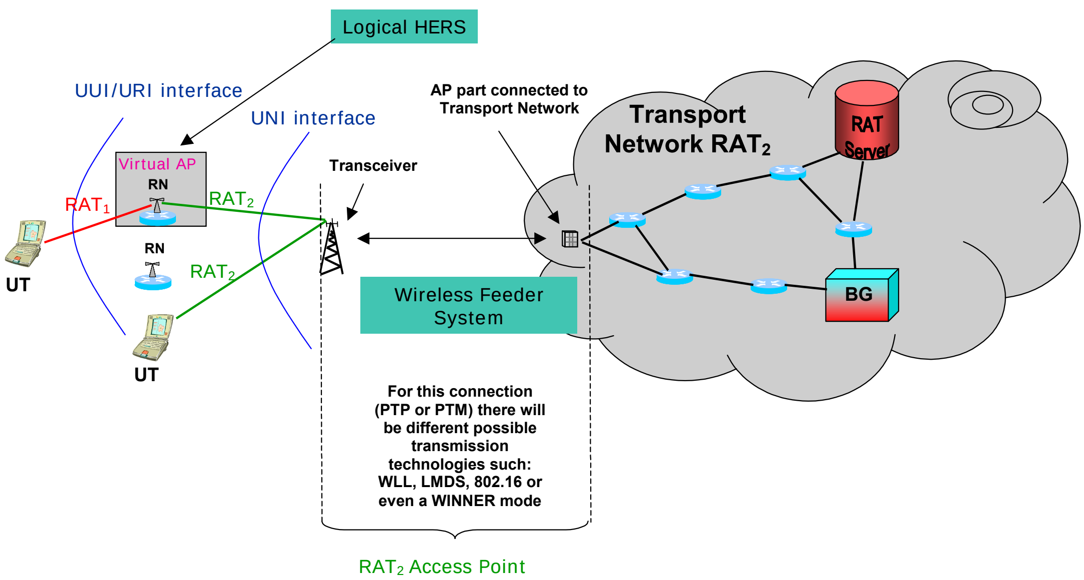


Figure 4-99: Scope of heterogeneous relay station (HERS) and wireless feeder concepts

Therefore and based on the previous discussion a heterogeneous relay station is defiend as a network element that is wirelessly fed using a given radio technology by means of another relay station or an access point (base station), and serves to another relay station or to a final user using a different radio interface technology. The radio interface technologies that a heterogeneous relay incorporate can be different modes of the same RAT or RATs completely different.

Although in this definition of the heterogeneous relay it is stated that the radio interface technologies involved might be different modes of the same RAT or entirely different RATs, that is, two different standards like for example HiperMAN and HiperLAN/2 (case investigated and developed inside STRIKE project [243]), WINNER is mainly focused on the case of two different modes of the only radio access technology that is being developed inside WINNER project.

Regarding heterogeneous relay concept, of course, the same physical node could act like a heterogeneous or homogeneous relay station. The example of Figure 4-99 shows a logical HERS since the link with the AP is done using a radio interface (RAT_2) different to the used for linking with the UT (RAT_1). In any case the same physical node could use the same RAT for the connection with the UT, like in the case of homogeneous relay station concept (e.g. in terms of distance from the UT to the relay node, this could decide to use RAT_1 or RAT_2, assuming of course an UT multi-mode). As aforementioned the RAT_1 and RAT_2 could be different standards, such WLAN and WINNER, or different modes of the same RAT (RAT_2 can be using 64QAM because this mode is thought for short distances and RAT_1 16QAM because this other mode is thought for large distances).

4.3.1.1 Simple HERS Description

Figure 4-100 shows a HERS, which is seen by the AP UT of using the same radio technology as the AP (RAT_{mode2}), and at the same time is seen by the UT working in the other radio technology (RAT_{mode1}) like an AP serving UTs using RAT_{mode1}. From a processing point of view a HERS receives data from network elements with a given radio technology mode and forwards the same data or equivalent to network elements with a different radio technology mode. In some cases could occur that exist messages in one of the two modes or technologies involved in the HERS without an equivalent message in the other one (e.g. in UMTS TDD and FDD there are RRM messages specific for each of the modes). In those cases, it will be necessary to implement some mechanism in order to generate or translate automatically such messages. This way a HERS would be a relay station that has to handle the incoming data and include some kind of conversion table (L2, L3 or both) and interworking mechanism in order to forward the data, reaching a correct performance of the element. Respecting control congestion in the two links (AP-RS & RS-UT or RS-RS) and in the two directions, both homogeneous and heterogeneous relays have to implement some kind of mechanism in order to avoid the possibility of congestion in some of the links

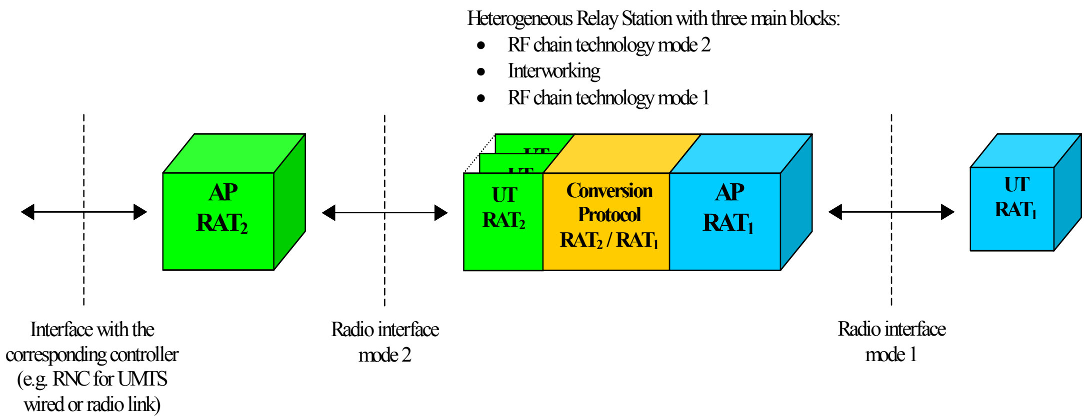


Figure 4-100: General description of Heterogeneous Relay Station

In WINNER the HERS will take advantage of the GLL and MCP concept as described in Chapter 2. Some ideas of these advanced concepts in the context of the HErS will be discussed in Section 4.3.4.2.

4.3.2 Types of Heterogeneous Relay Stations

As stated above the heterogeneous relay couples two RATs or in the case of WINNER RAT modes. The questions are what benefit do I have to couple two RAT modes instead of having one mode? What is the reason to have to RAT modes? In the following different use cases for relays comprising tow or more RATs or RAT modes will be introduced.

4.3.2.1 HERS Type A

One important use case is to get UTs that are less complex compared to APs by exploiting the fact that they need less functionality due to their usage scenario. If we have, e.g., a system that supports high mobility, mesh modes and seamless handover, which is needed to cover wide areas with mobile UTs in cars, busses or trains but want to use these system on the other side feed, e.g., households or offices with UTs that are temporarily stationary and thus do not require sophisticated mobility support like seamless handover. In the latter low mobility scenario it might be useful to allow less complex and therewith cheaper terminals which are based on an adapted RAT mode.

Figure 4-101shows a scenario as described above with the heterogeneous relays mounted on the roof top of the private housings. In the depicted example one wide area AP based on RAT mode 1 is feeding three heterogeneous relays with directed links as well as a UT in a car. Thus the RAT mode 1 allows high mobility with seamless handovers. In addition the heterogeneous relay serves UTs inside a house like a desktop PC, a laptop or a PDA. All of these devices are at least temporarily stationary and therewith don't need high mobility support.

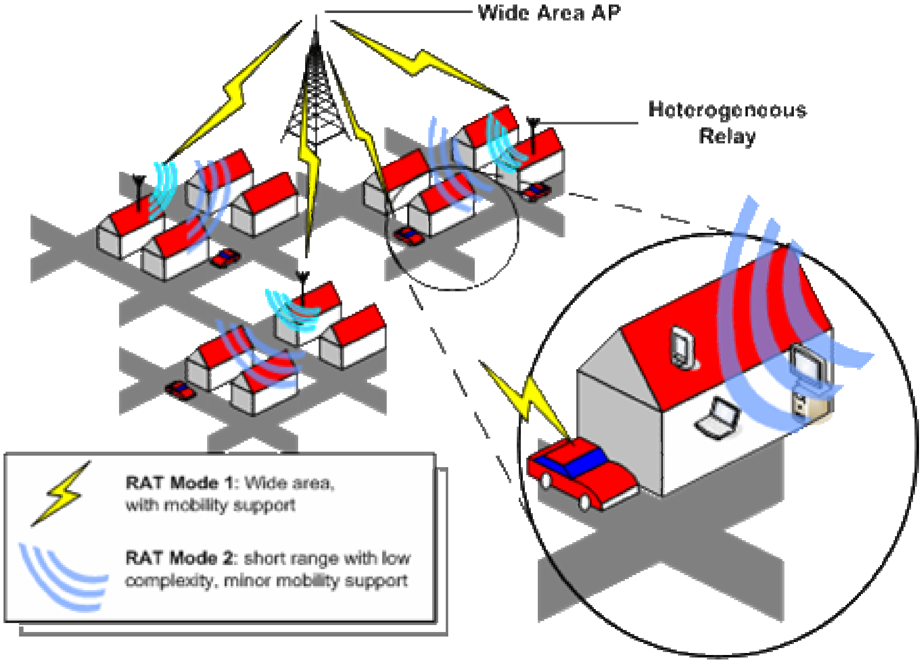


Figure 4-101: Exemplary use case for a heterogeneous relay

4.3.2.2 HERS Type B

The use case B is to have an heterogeneous relay in order to allow a best exploitation of the static link between the AP and the fixed relay by using a RAT mode tailored for the static link conditions and, e.g., the possibility to use more advanced antenna concepts as a fixed relay does not suffer from limited battery power or strict size constraints. This use case is shown the scenario depicted in Figure 4-101 if the car would be fed by the same RAT mode (RAT mode 2) as the other UTs, which could be provided either by the heterogeneous relay or by mean of a multi mode wide area AP.

4.3.2.3 HERS Type C: The wireless feeder

Use case C is the wireless feeder concept where the APs in the streets are fed wirelessly by feeder stations as part of the transport network. For the connection between nodes-B and its respective radio network controller (RNC) in current cellular networks (3G), it was contemplated the possibility to use radio links as an alternative to the wire line (E1). Following with this trend, WINNER project will investigate the characteristics and requirements of wireless feeder system focused in some of the scenarios described in the internal report IR3.1. Although at this moment the future architecture for the WINNER access network is not clear yet, the wireless feeder concept to develop in WINNER project will tackle with the feeding of WINNER APs with wireless point to multipoint (PtM) links.

Although there are a lot of proprietary solutions for implementing radio links, the radio transmission technology initially selected for the wireless feeder scenario has been IEEE 802.16 (WMAN) in the point to multipoint (PtM) mode, since the idea for the first phase of WINNER project was to use technologies based on existing systems supported for some important Standardization Organism. It is convenient to clarify that IEEE 802 LAN/MAN Standards Committee created in March 1999 the 802.16 working group on broadband wireless access and it focuses on standardizing US LMDS (Local Multipoint Distribution Service) systems.

An example for the feeder concept in the WINNER context is shown in Figure 4-102 where the APs are connected to the wireless point-to-multipoint feeder system (red arrows) on the one side and feed their FRS and terminals with radio access system on the other side. The feeder link has LOS conditions and could be served on a higher frequency band than the access link.

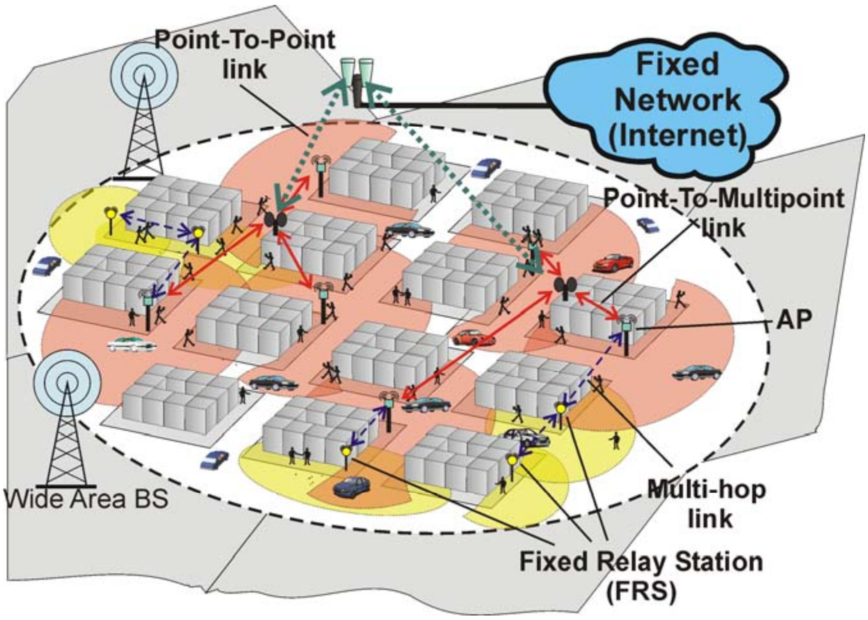


Figure 4-102: Wireless Media System with wireless point-to-multipoint feeder

It should be mentioned that the heterogeneous relay and especially the wireless feeder could combine RATs or RAT modes on different spectrum bands, according to the requirements of the different links.

4.3.2.4 HERS Type D: Coverage Extension

it is possible to enlarge the coverage of an access point of a given technology (RAT_{mode2}) beyond its own coverage area but using other radio technology (RAT_{mode1}). The number, distribution and location of those elements obviously would be aspects that we will have to contemplate in order to decide the most proper strategy for each of the scenarios and conditions where the use of HERSs could be interesting. For example, Figure 4-103 shows a possible configuration with an access point of RAT_{mode2} (proper to outdoors environment) and four HERSs fed with RAT_{mode2} technology by the AP and to serve to final users using RAT_{mode1} technology (proper to indoors environment). Regarding the way to connect the AP to the transport network, there are of course several possibilities, using a wire line or by means of a radio link depending on the available resources and the requirements of this connection. In other words, taking cost and reliability level into account, we should decide what class of link is the most proper for connecting the AP with transport network, but like we have explained before this aspect fits in the wireless feeder system.

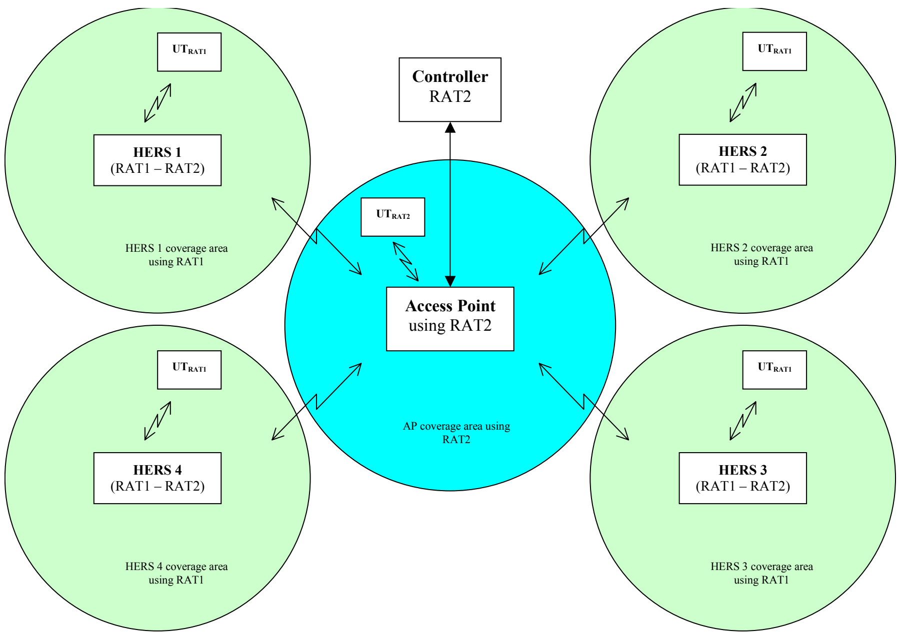


Figure 4-103: Example of coverage outdoor/indoor using HERSs

4.3.3 Main integrating parts of heterogeneous relays

4.3.3.1 Level of interworking

Taking the previous definition into account the HERS would be always a "Decode & Forward" relay type since the communication between the two elements that has to hold, involve the use of some kind of mapping table for the conversion of protocols and some interworking mechanisms (congestion control) between the two RATs or modes of the same RAT used by each of the elements. So before forwarding the data, it is necessary to decode the data of the incoming mode, to do the conversion and to encode the data in the other technology or mode. Assuming this operation and for the case of different physical layers, the simplest HERS would include only the analysis and handling of MAC frames in both directions as shown in Figure 4-104 which illustrates the simplest case of HERS connecting two different PHY layer modes (L1 and L1') via a common MAC. Even if at this moment it is not known yet the different modes of WINNER radio access, they will have likely some kind of divergence in MAC layer.

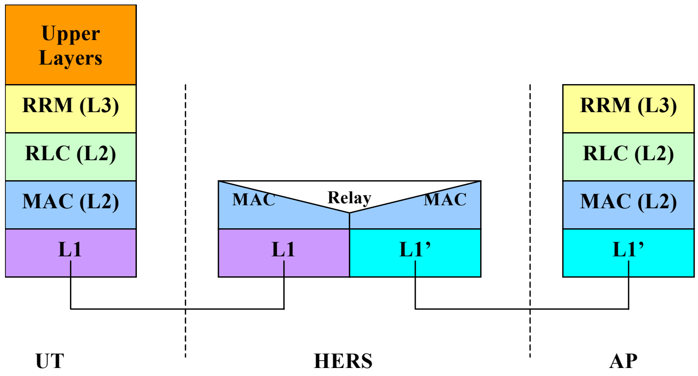


Figure 4-104: Protocol architecture for simplest HERS

The STRIKE project has investigated three different approaches for an interworking between HiperLAN/2 and HiperMAN as shown in Figure 4-105. The interworking has been performed on three different layers, namely by means of IP bridging, Ethernet bridging and DLC bridging. Thereby the DLC bridging was the only interworking mechanism where the two systems had a direct connection, without involving an independent higher layer protocol. In STRIKE it turned out that the DLC interworking was the most complex one with the highest problems in terms of QoS mapping and applicability to the standards. The main reasons identified by the IST STRIKE project [240] have been the different QoS schemes of both standards and the missing of an appropriate convergence layer (CL). However, the WINNER approach is different as the heterogeneous relay will be comprised of two RAT modes of the same air interface, which can be designed in order to allow a seamless interworking between the modes in a heterogeneous relay. Therefore the most promising approach will be to implement the interworking mechanism in the generic parts of the protocol stack as described in Chapter 2 .

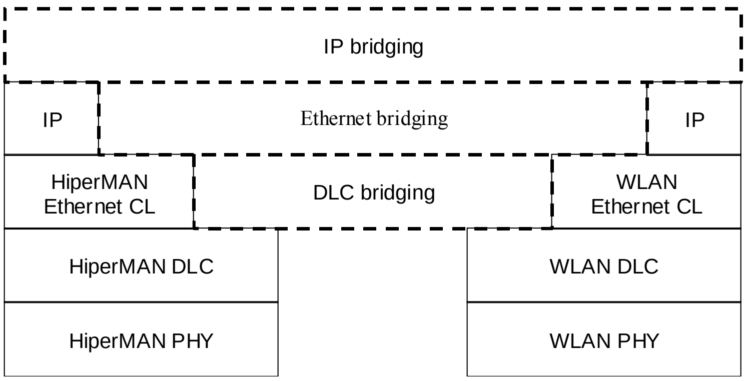


Figure 4-105: Possible strategies for HipeLAN/2 HiperMAN interworking [240]

Regardless the detailed design of HERS is not a priority objective of Task 3.4 at least in phase 1 of WINNER project, it is convenient in this first phase to identify from a functional point of view the main blocs that a HERS should incorporate in its design in order to cover all the functionally of this kind of element.

4.3.3.2 Basic functional Blocks of a HERS

In addition to RF modules for transmission and reception of both links (AP-HERS and HERS-UT), in general terms and pointing to modes conversion functionality, HERS would have to include for each direction (DL and UL) a block for decoding messages coming from a radio access technology mode, other block for transforming these messages to the other mode, and finally before forwarding messages to the other end a block for encoding the data in the other mode. Figure 4-106 illustrates the idea of this functional division. Of course the complexity of modes conversion block will depend on the differences of the two involved modes. For example if the only difference between the two modes is the modulation and coding scheme, this block would be simpler than for a case where the modes have also differences in upper layers of protocol. Besides there are other modules to implement for control and supervision functions such error detection, interferences control, congestion control, synchronization, power control and so on, which for each particular case there will be to include or not.

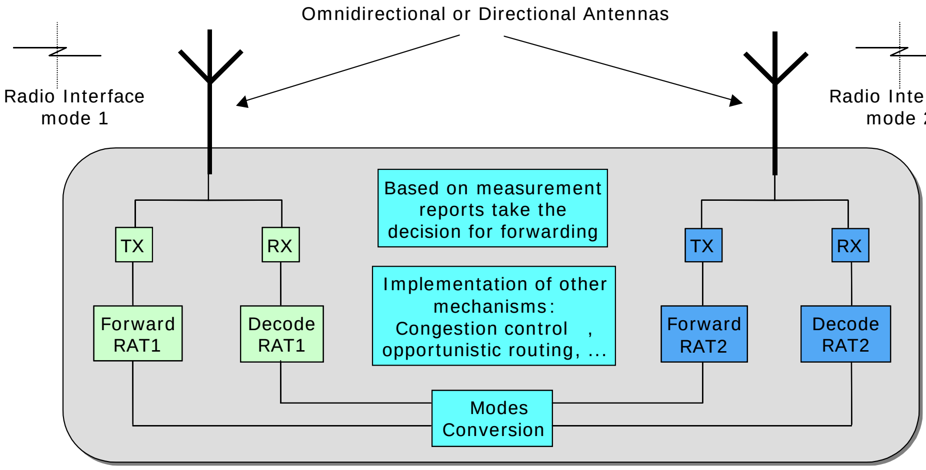


Figure 4-106: Main integrating parts of heterogeneous relay station

4.3.3.3 HERS example with different duplexing schemes

At this moment it is not clear if the different modes of the radio access technology to develop inside WINNER will employ the same duplexing scheme or if they will finally contemplate several schemes for using in terms of the particular scenario. In this way one more developed scheme for HERS working with different duplexing methods, for example TDD for AP-HERS link and FDD for HERS-UT link, is shown in Figure 4-107. Obviously assuming these two duplexing modes, the decision to use one of them for the link between the AP and the relay, and the other one for the link between the relay and the UTs, will depends on many factors and it will be necessary to carry out a deep analysis of pros and cons of each alternative from different viewpoints such feasibility, total throughput, complexity, interferences management and others. Anyway, notice that this block diagram is not so much detailed and it is only a tentative for an example where the two modes to interconnect through HERS, employ different duplexing methods. Even the selection of components (RF components, DSPs, FPGAs, microprocessors) for implementing each of the identified functions in the diagram, is an aspect to solve in the future and outside of WINNER scope.

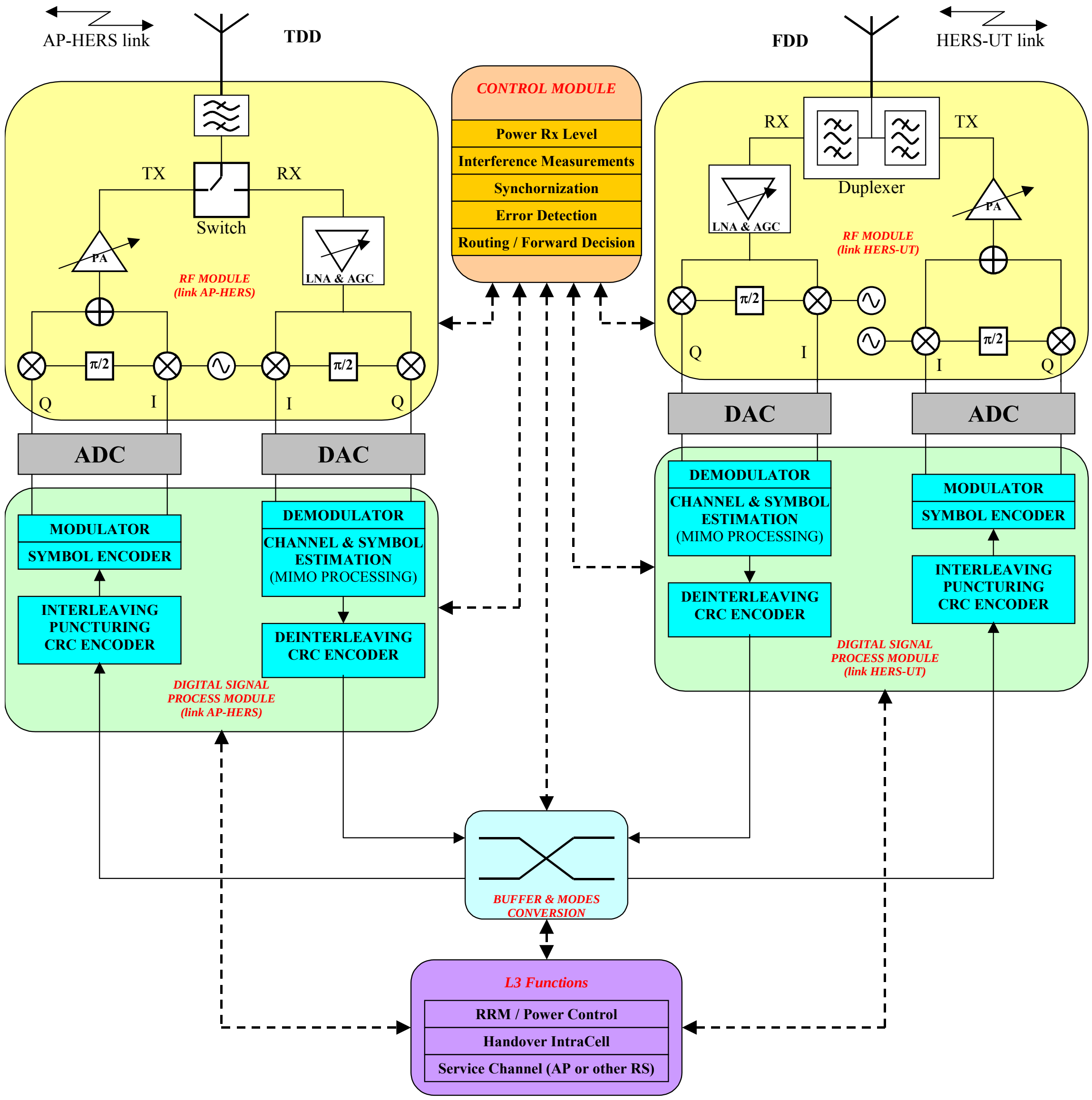


Figure 4-107: Block Diagram for HERS with two duplexing methods

4.3.3.4 Complexity of the HERS

Other important aspect it is the complexity of this kind of elements, since they would have to be cheaper than APs and commonly the location will be in lamp posts of flat roofs (perhaps also on facade of buildings). One possibility could be to include some functionality of the HERS in the own aAP (RAT_{mode1}) and with local and remote antenna units take the service (RAT_{mode2}) outside the coverage of the AP. This way the same AP would serve to users of different modes depending on its location as shwown in

Figure 4-108.

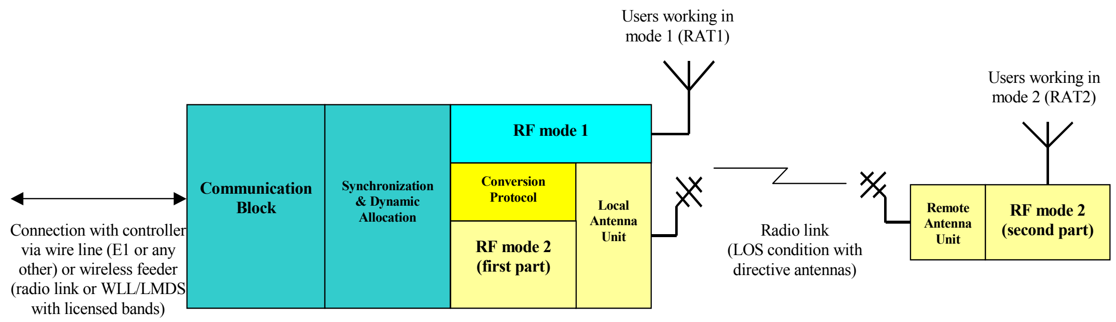


Figure 4-108: Possibility to divide the complexity of HERS

For example Figure 4-109 illustrates a tentative block diagram for a heterogeneous access point with the two modes of UMTS, FDD and TDD, allowing a dynamic distribution of radio resources available in a given site between the users of these two modes. In this case we have assumed that the same Communication & Signal Handling module can be used for both modes, although at this moment for UMTS at least this matter is not clear and in fact as far as is known there are not cases where an only Node-B merge FDD and TDD modes using the same module for communicating with its controller. Anyway the idea is that the AP could serve directly to final users (or relay stations) of one mode, for instance mode 1, and at the same time could establish, using mode 2, a link with a remote antenna unit (RAU), which would be at last the equipment in charge to give service at final users of mode 2. Therefore, please notice that in this example the remote unit is simply a homogeneous repeater, since the two involved radio interfaces in a given communication (AP-RAU link and RAU-UT link) are the same, that is, mode 2. Of course the complexity of this example is in the access point, and the problems for connecting the local unit with the remote unit is an aspect more or less studied and solved by means of different alternatives (wireless or wireline solutions).

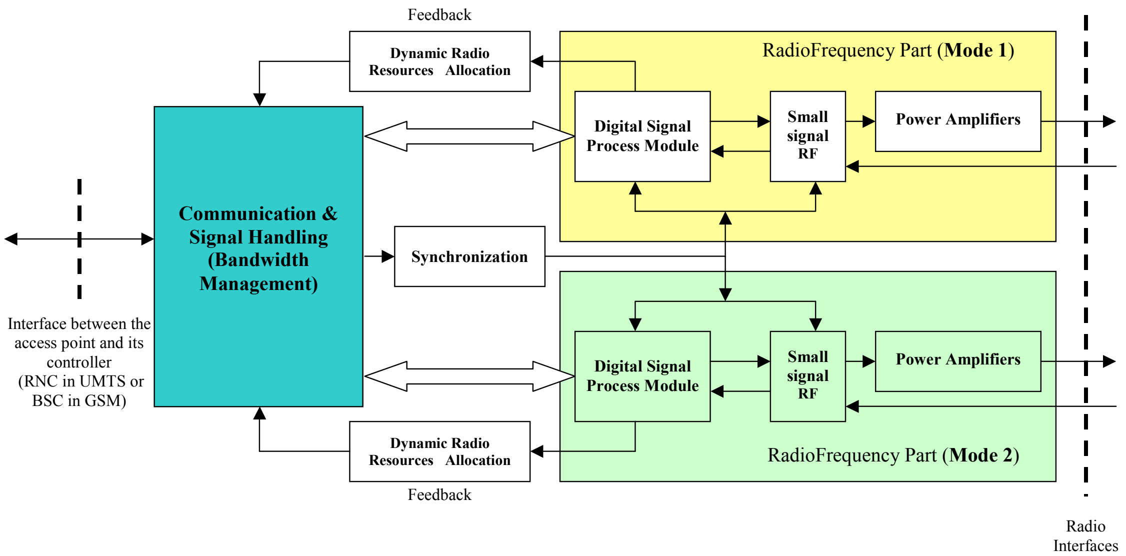


Figure 4-109: General block diagram of an AP merging two modes of the same RAT

4.3.4 Functionalities of heterogeneous relays

4.3.4.1 QoS parameter mapping

In the case of heterogeneous relays it is important that the QoS parameter will be maintained on the transition from one RAT or RAT mode to the other.

When, like in WINNER, a new radio interface will be designed the QoS classes are the same for the system concept as a whole and therewith for all WINNER RAT modes, which eases the design of the heterogeneous relay. Taking the approach of Chapter 2 into account where the existence of a GLL and a generic RRM are discussed it is obvious that the generic Layers will be the ones where the interworking will be performed.

4.3.4.2 RRM functionalities in the context of a heterogeneous relay

For the description of RRM functionalities in the context of heterogeneous relaying the scenario described as use case B in Section will be taken as basis in the following.

The interworking could look like shown in Figure 4-110 where it is performed by the GLL on the user plane to forward the user data and the RRM. The RRM in this case will have some common (generic) functions depicted as RRM-g in Figure 4-110 and some mode specific functions (RRM-rx). The RRM-g part is coordinating the resource demands between the two RAT modes.

The difference of the two links and therewith of the requirements for the two RAT modes on both sides of the HERS can be described as follows:

- **Hop 1 (AP-HERS):**
  - No change in link quality (static and well known link conditions)
  - Point to point connection from HERS's point of view
- **Hop 2 (HERS – Uts)**
  - Dynamic link conditions up to loss of connection
  - Resource has to be shared by one or more connections

This means for the HERS that on the one hand it has to distribute the resources between the UTs based on their demands and on the other hands it has to provide a mechanism to release resources on the first hop to achieve an efficient resource utilisation on both sides of the relay. Another possibility would be if the both hops of the relay share one radio resource. In this case a common MAC as introduced in Section 2.5 would be in charge of the shared medium access between both systems.

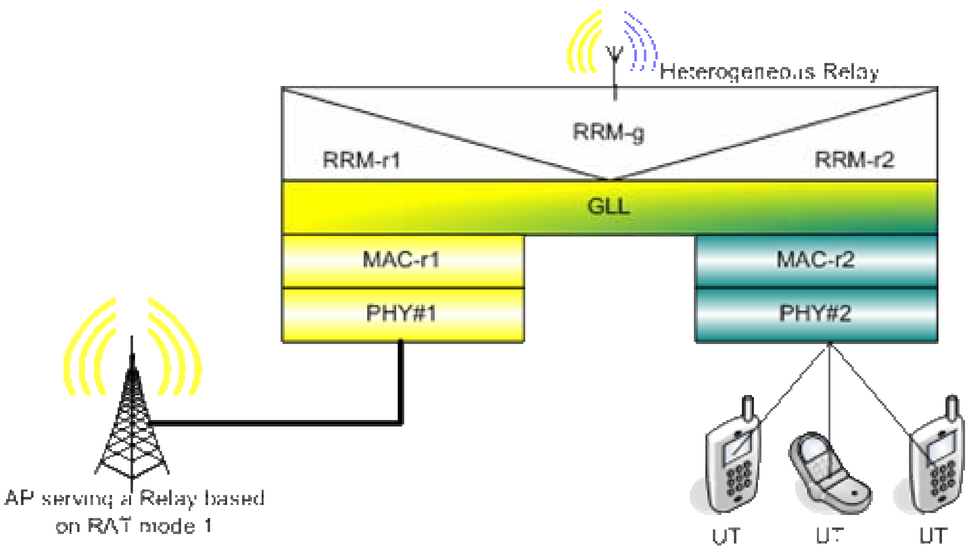


Figure 4-110: Possible protocol stack of heterogeneous relay station

In the following the impact of the two different radio links will be discussed in more detail. In Figure 4-111 a situation is shown where one HERS fed by one AP is serving two UTs. The link of UT1 to the HERS is now assumed to be interfered. It is assumed the UT1 will adapt ist PHY mode to the changed situation. In the following the scenario depicted in Figure 4-111 is discussed for separately for the UL and the DL:

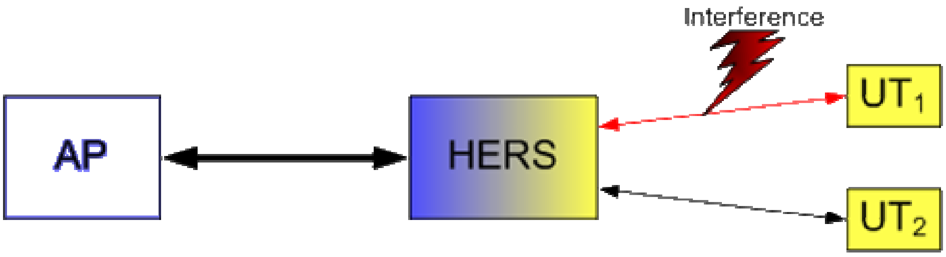


Figure 4-111: Interference situation in HERS

DL case. It is assumed that UT_1 changes to a PHY mode with less data rate, which means the HERS has to inform the AP to reduce the amount of data for $UT_{1\ in}$ order to avoid a buffer overflow. The consequence is that unused resources occur on the first hop. Theses resources could be used by the AP, e.g. to feed some other relays or UTs in the coverage area of the AP. On the other side the UT_1-HERS connection could be only temporarily interfered therefore the HERS would release the resources on the first hop also only temporarily. In case that the link between $UT_1$and the HERS is broken the HERS has two options one is to assign the free resources to UT_2 requesting more data for UT_2 from the AP and the other one is to release the resources on the first hop.

UL case. The situation on the uplink is quite similar with the only difference that the HERS is not endangered to run in a buffer overflow this means no end to end signalling is required. But also here the HERS has to release the radio resources on the first hop temporarily.

As conclusion it can be stated that an end-to-end signalling is required in order to allow an efficient utilisation of the radio resources for both the uplink and the downlink.

4.3.4.3 Routing/Forwarding in the context of heterogeneous relaying

Routing in multi-hop packet radio networks (MPRNs) and ad hoc networks have received considerable attention during the last couple of years. Nevertheless, there has not been much explicit research in the area of heterogeneous relaying.[11] Even though we will be very brief here, the goal for future work is to assess the benefits of performing forwarding at the GLL level as compared to more conventional routing procedures (i.e. IP-level routing) as well as to develop new novel schemes to be able to exploit the added diversity gain that one may use due to the presence of multiple modes. Hence, in this section we will give an overview of how forwarding in a heterogeneous relaying case may be performed at the GLL level (this was also briefly outlined in Section 2.5.

To illustrate the ideas we will consider a case with three nodes: one combined AR, BG and AP (which provides access to the internet) – henceforth referred to as the AP; one RN and one UT. Depending on how many modes the different nodes support, different scenarios may be envisioned (the figures below outline three such different cases with different possibilities to utilise multi-mode capable nodes). In the first case (Figure 4-112), the UT only incorporates one mode (m2) whereas the AP and RN incorporate two modes (m1 and m2). As shown in the figure, the downlink traffic from the AP to the UT may follow three different wireless paths[12]: one direct path using mode m2 and two two-hop paths (wherein the first path uses mode m2 for both hops and the second one uses mode m1 for the first hop and mode m2 for the second hop). In the same manner Figure 4-113 and Figure 4-114 show two scenarios were two different paths and four different paths may be used from source to destination respectively. From this we may conclude that in a scenario encompassing multi-mode capable nodes several different paths may be used for any flow between the AP and UT.

Moreover, by letting the transmitter to choose one of the available modes for data transmission (the mechanism of choosing which mode to use and denoted mode selection is further described in Chapter 2) the overall throughput may be enhanced and may be seen as an extension of transmission diversity techniques to the WINNER framework of being able to use multiple modes. Henceforth, this form of diversity will be denoted Multi-Mode Transmission Diversity (MMTD). In a relaying scenario another form of diversity may also be accounted for, namely the multiple paths outlined above. Henceforth, this

[11] Most developed MPRN routing protocols (which are IP-based) claim that they may be used over any underlying radio access technologies, however when one device contain more than one radio access technology it normally implies that for every interface a unique address is used for routing purposes.

[12] Usually a path is made up of the nodes traversed from source to destination, but here a path is made up of the nodes *and modes* traversed/used from source to destination.

form of combined diversity will be denoted Multi-Mode Path Diversity (MMPD) and may be defined as: *the dynamic selection of one (or more) mode(s) to be used on the one (or more) path(s) from source to destination*. One may further envision a case where several modes and routes are used in parallel. This case may be further refined to reflect whether the multiple parallel transmissions are used for increasing the robustness or data rate depending on if data is duplicated or not.

Applying MMPD in the AP for DL traffic would require (i) the knowledge of which RNs and which modes that the AP may utilise to reach the UT and (ii) some form of (mode dependent) routing metric so as to be able to make the route and mode selection (e.g. capacity estimations on the links between AP and UT).

The rate at which the nodes within the network may switch paths and modes will affect the possible gains of performing MMPD. The greatest gains (at least in theory) will be encountered when nodes are able to switch path(s) and mode(s) on a packet-by-packet basis. In this case, one may exploit the variations in radio channel quality over short time intervals to optimize the selection decisions. However, being able to select path(s) and mode(s) on a packet-by-packet basis is also accompanied by the greatest overhead since the nodes in the network will have to rely on up to date information on several paths and modes. Another possibility is to restrict the switch-rate and only allow GLL to switch mode and path once every x:th second or even not allow mode and path switching once for the whole duration of a communication session.

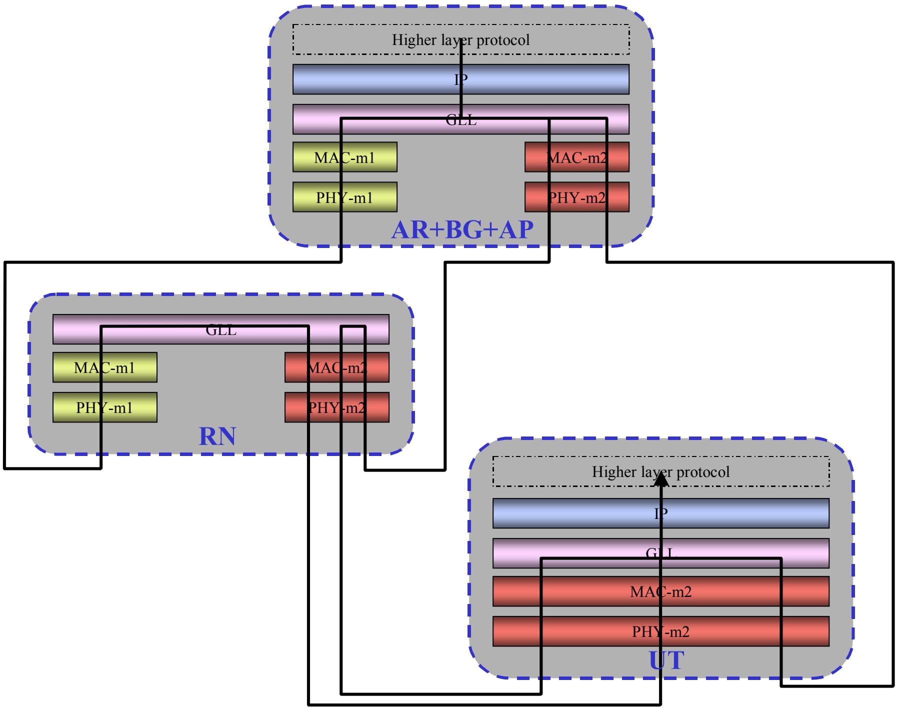


Figure 4-112 Possible relaying scenario involving two modes

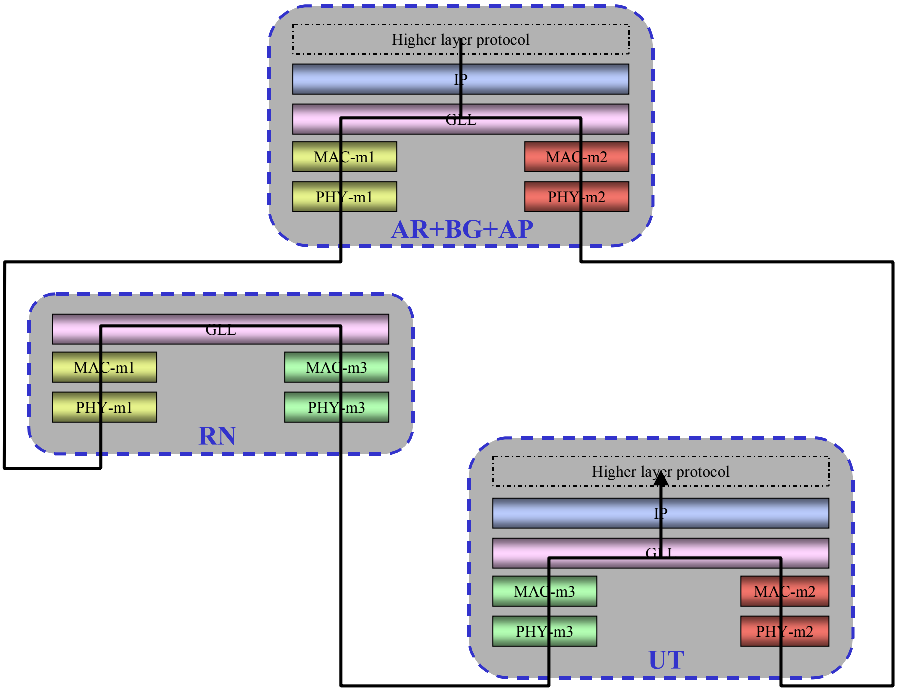


Figure 4-113 Possible relaying scenario involving three modes.

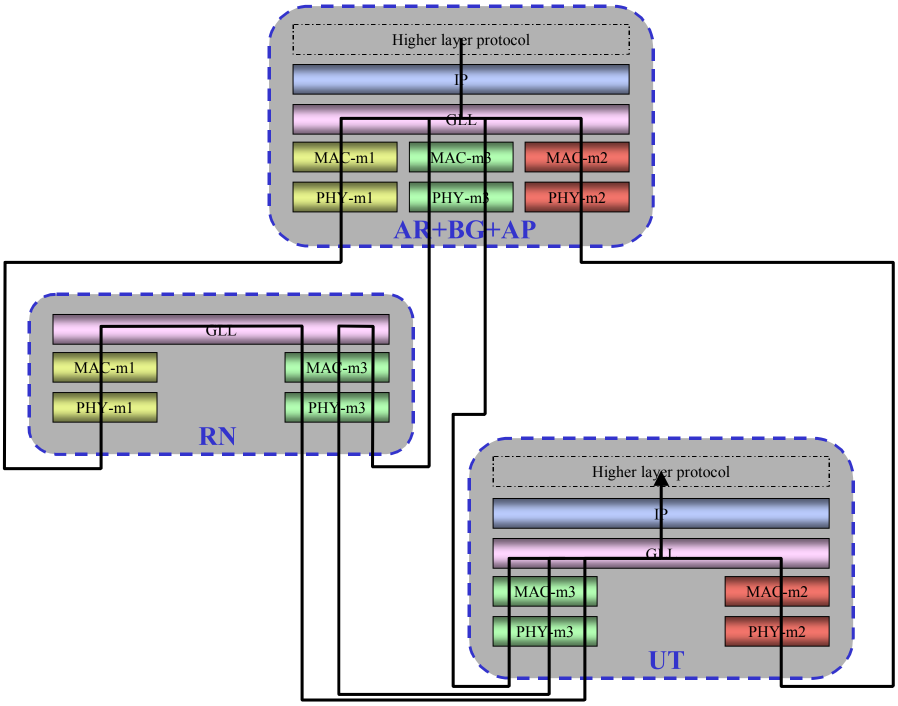


Figure 4-114 Possible relaying scenario involving three modes

4.3.5 Cooperative relaying using different air interfaces

In this section, we investigate the applicability of cooperative relaying as described in Section 4.2.6 to heterogeneous relaying. Recall that the general idea of cooperative relaying is to exploit two inherent characteristics offered by relaying systems: the spatial diversity and the broadcast nature of the wireless medium. This is achieved by combining the signals arriving from source(s) and relay(s) at the destination. We have described various protocols for the homogeneous relaying case, where the signals to be combined are carried via the same air interface.

The work plan defined for Task 3.4 identified a sub-task (T3.4#3) for investigating the cooperative relaying topic with different radio interfaces. The objective of this sub-task was to perform a preliminary analysis of the possibility to implement cooperative relaying schemes for heterogeneous relay stations. To this end, a feasibility study was carried out for relays with heterogeneous RIs. Following a review of related concepts, various concepts with a proper cooperation between the different radio interfaces were analyzed. It was clear from the beginning that complexity is the limiting factor; therefore, the evaluation of the concepts focused on this issue.

Based on the discussion of this feasibility study, it was decided to abandon this activity, and to concentrate all the effort for investigating the cooperative relaying using the same air interface. Implementation is much more promising for the homogeneous case, whereas the discussion about cooperative relaying using different air interfaces has shown that complexity does not seem to justify the benefits that we would achieve with heterogeneous cooperation.

4.3.5.1 Literature Review

Hereinafter the few works which at the present time are somehow related to heterogeneous relaying. Moreover only the first one contemplates the incorporation of user cooperation and diversity with different air interfaces.

4.3.5.1.1 Virtual Antenna Arrays

The only work on relaying using heterogeneous air interfaces was developed using the terminology Virtual Antenna Arrays (VAA) at the Center for Telecommunications Research, King's College, London, UK. Related publications can be downloaded from [244].To lend further insight into the advantages and disadvantages of the taken approach, we shall first describe how VAAs are constructed. Afterwards, the assumptions made in this context are evaluated since an in-depth understanding allows us to predict the potential benefits of this technology as well as discuss its technical feasibility.

System Model

A good bottom-down approach towards the entire topic of virtual antenna arrays is taken in [244] and especially in [245]. The concept of VAA is presented as to let spatially close mobile terminals communicate among each other in order to achieve gain from spatial diversity. The motivation for deploying VAA is drawn from the fact that future BS will very probably be equipped with more than one antenna and that highest MIMO capacity is achieved when the number of transmit at receive antennas coincide. The emulated MIMO system is promised to yield drastic increases in link quality of service and data rates. Application is said to be possible to a variety of current and future standards (2G, 2+G, 3G, B3G).

Various implementation aspects are shortly discussed, such as possible attach and detach scenarios, MS as well as BS requirements on system level, and physical layer problems that have to be tackled such distance estimation, synchronization, power control, and others. As every other system offering spatial diversity to handheld devices with single antennas, VAAs need relaying in order to enable inter-mobile communication.

Two different options are presented for providing this communication link. Either, we can use the same air interface for relaying as for the broadcast type transmission from the source, or we can use a different air interface to forward data from the relay(s) to the final destination. We shall use the terminology from Section 1.2 and refer to the wireless standard used at the BS as "main link technology" while "relaying link technology" describes the air interface used by the relay terminals. In the WINNER context, one could refer to air interfaces A and B, respectively.

Various problems associated with VAA deployment (increased system load and interference, orthogonality constraint) occur only when using the *same* air interface for main link and relaying link. Clearly, by employing a *different* interface for intra-VAA communication and transmission from the VAA to the final destination, these problems can be avoided. The traffic load created by the VAA is shifted from the main access network to the relaying network.

Three main link technologies (UMTS, GSM, HiperLAN/2) and five relaying link technologies (UMTS, GSM, HiperLAN/2 (H2), Bluetooth, Power Line Communications (PLC)) are proposed in [244] to construct the following example VAA scenarios:

- UMTS as main link; and UMTS, GSM, H2, Blue tooth or PLC for relaying links.
- GSM as main link; and GSM, H2, Bluetooth or PLC for relaying links.
- H2 as main link; and H2, Bluetooth or PLC for relaying links.

While this approach looks very promising at a first glance, various implications compromise the benefits which may be the reason why further research concentrated on the case of same air interfaces for main and relaying link. Subsequently, we shall discuss some of the challenges of the mixed air interface approach.

Signal Combining

In wireless systems, signal combining is most beneficial when done at the lowest possible level (physical layer) so as to be transparent to upper layer protocols. However, combining of radio wave signals from two different wireless standards at this layer will likely prove too complex and costly for implementation. The difficulties faced in hand-over from UMTS to GSM can be regarded as a reference, though this task is far less complex than the real-time signal combining over two different air interfaces probably using different frequencies and multiple access schemes (e.g. CDMA and OFDM), which would be required for VAA.

Signal combining at higher level (i.e. combining of demodulated/decoded signals with or without additional signal quality indicators) is unlikely to be advantageous, due to a variety of reasons:

- **CSI requirements.** Analog or only quantized channel state information is often required to efficiently combine signals from multiple antennas and/or efficiently decode convolutional and turbo codes. It remains to be evaluated if it is desirable that upper „convergence layers" (which would have to be developed) deal with these large amounts of data in handheld devices of low complexity and with harsh limits on processing power and available memory. Current VAA proposals also require the destination to have knowledge of the channel conditions between source and relay.[13]
- **Complexity**. Obviously, such a solution would require the relay as well as the destination to be able to deal with two different wireless standards. At least two different RF chains are required.
- **Delay**. Especially for real-time applications digital higher level combining would be no viable solution since the needed synchronization and handshake algorithms due to mapping between different air interfaces would surely introduce notable delay.

In that context, it would probably prove to be the most advantageous to simply select a MS in the vicinity of the final destination that experiences the best channel conditions and/or features multiple antennas and employ „conventional" relaying. A comparable technique for different air interfaces has been proposed in [246] ,which we discuss later.

Load of Relaying Network

As discussed above, the deployment of VAA may, depending on the implementation, impose an amount of traffic on the relaying network that may exceed the payload by degrees. While VAA is envisaged especially for high-data rate applications, it should be investigated how this "artificial" traffic increase affects the performance of the relaying network and whether the achieved benefits outnumber the incurred costs.

Wire-based Relaying Network

An additional option is to shift the relaying problem entirely from the wireless context to a wire-based relaying network. Consequently, bandwidth constraints and traffic are no longer an issue in forwarding data from relay to destination. What we lose on the other hand is the great advantage of wireless networks as the mobility. If we are bound to a cable either way, it will be technically far easier to use a local area network (LAN); data services are readily available and telephony services can be provided by voice-over-IP related protocols. Additionally, the devices most probable to connect to wire-based networks are notebooks, which probably feature multiple antennas anyway and do not require VAA help to gain spatial diversity.

[13] This information would have to be made available to the final destination receiver - a drastic increase in traffic since for each transmitted signal bit a considerable amount of soft information would have to be additionally transmitted in fast fading environments.

4.3.5.1.2 Conclusions

By using a different air interface for the relaying links, we can avoid some of the core drawbacks of VAA. However, due to the nature of this technique, it is no longer comparable to conventional cooperative diversity schemes since power normalization is impossible when using different wireless standards simultaneously. What we can achieve at best is obviously the performance of a receive diversity system with a number of antennas being equal to that of the VAA. It remains to be investigated how this performance can be attained while at the same time minimizing the traffic required to do so. The topic has not been further pursued in the following publications on VAA.

4.3.5.1.3 Integrated Cellular and Ad Hoc Relaying Systems

Wu et al. [246] have proposed an integrated cellular and ad hoc relaying system (iCAR). The basic idea of the iCAR is to place a number of ARSs in a cellular system to divert excess traffic from one (possibly congested) cell to another. Ad hoc relaying stations (ARS) are thereby placed at strategic locations, acting as repeaters that relay traffic between mobile hosts (MHs) and BTSs through a different air interface. The cellular interface **C** is used for conventional links to the BTS, while the relaying interface **R** enables communication between MHs and ARSs and between different ARSs. The **C** interface could be any cellular interface, and examples for the R interface are wireless LANs or ad hoc network interfaces. Using the Erlang-B model, the performance of the proposed system is compared to a conventional cellular system for an example scenario. However, the model assumes a fixed number of traffic channels available in each cell (FDMA, TDMA); mutual interferences, soft capacity etc. are not considered (CDMA). A review of related work is given. Clearly, this work does not incorporate user cooperation and diversity, the approach towards relaying is done in a conventional way.

4.3.5.2 Indentified concepts

Based on the above review, four different concepts have been identified. Before showing those cases, in the Figure 4-115 we describe the basic scheme for pure heterogeneous cooperative relaying. The basic scheme contemplates two phases and the heterogeneous relay includes two different modes (e.g. interface A or RAT mode 1 for linking with AP, and interface B or RAT mode 2 for linking with UT). During the first phase the source (AP) transmits the signal in mode 1 to relay and destination. In the second phase depending on the class of cooperative relaying (static, adaptive or feedback), the relay node always retransmits the signal in mode 2 or in terms of a given threshold parameter (e.g. fading coefficient), the relay retransmits or silences (with or without retransmission from source). Also, in this phase the destination has to combine the two signals of different or same modes (AP and relay signals or the two signals from AP) in order to get an only signal free of errors.

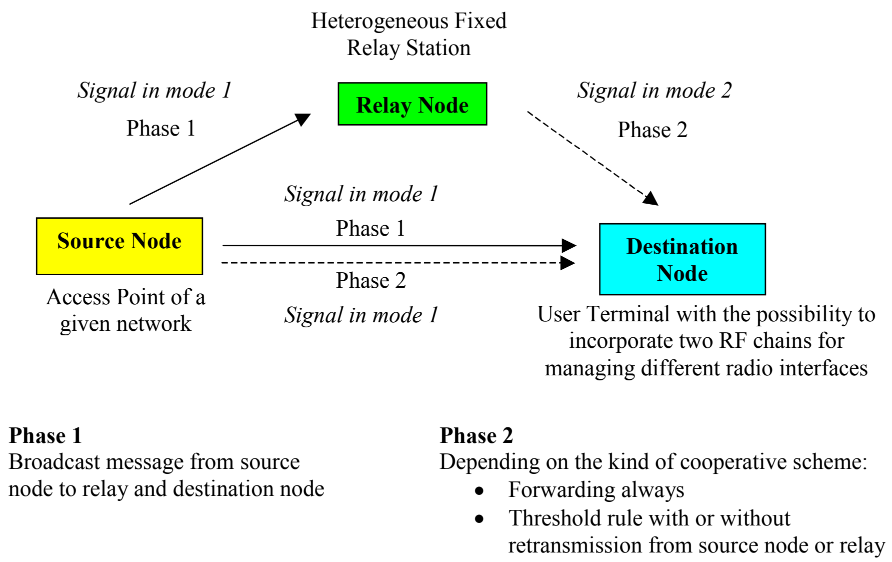


Figure 4-115: Basic Scheme for Heterogeneous Cooperative Relaying

4.3.5.2.1 True Cooperation on Separate Air Interfaces

This direct application of the idea of cooperative relaying is depicted in Figure 4-116. Considering a downlink scenario, an access point would send information for terminals via interface A. The terminals locally co-operate using an interface B to decode the information. Being a direct application of the virtual antenna array concept, this remains the most promising of the identified concepts. Local co-operation could range from simple selection combining to more complex methods such as distributed iterative decoding. Clearly, the main challenge is the coordination of the local interface B while simultaneously receiving on interface A.

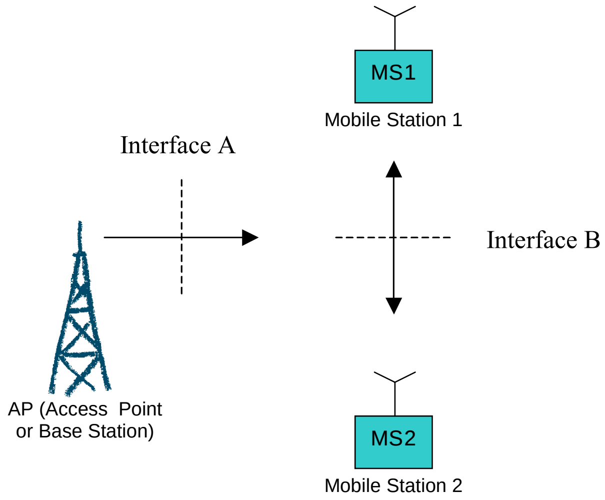


Figure 4-116: True Cooperation on Separate Air Interfaces

4.3.5.2.2 Diversity Provision by Relays

The second candidate concept is illustrated in Figure 4-117. In a two-hop scenario with multiple relays, the source sends signals via interface A to the relays; these forward the information on interface B to the destination.

Note the attractive separation of phases: the first phase solely uses interface A to convey information to the relays. The relays then independently perform conversion to interface B. The second phase is similar to the soft handover transmission known in 3GPP systems: two stations, which here are two relay stations, simultaneously transmit to the destination.

This systems' challenges are

- Coordination of transmissions, i.e., assigning relays, and conveying this information to the receiving mobile station,
- Orthogonal transmissions in phase two are required to achieve separation of the relays' signals. This can be done using different spreading codes as in 3GPP, or by using orthogonal space-time codes such as Alamouti's scheme. Note that the latter concept calls for symbol-level synchronization.

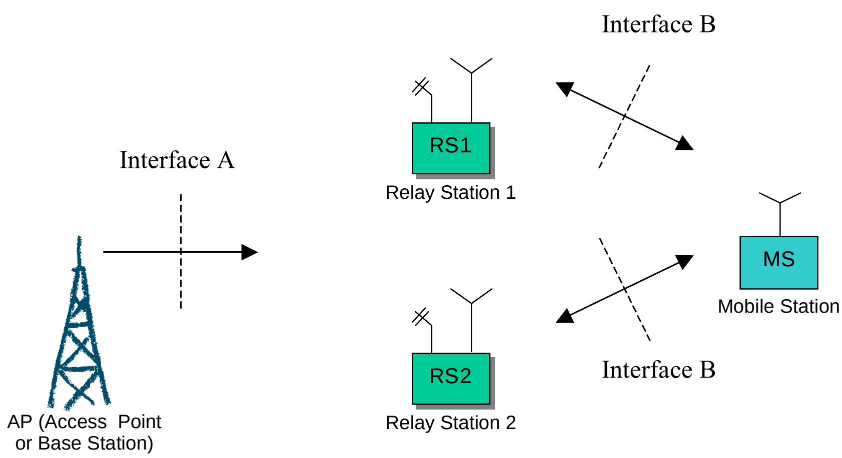


Figure 4-117: Diversity Provision by Heterogeneous Relays

4.3.5.2.3 Multi-Hoping

More complex scenarios are shown in Figure 4-118 and Figure 4-119. The common idea is to use two different interfaces in spanning multi-hop chains. Cooperation can be achieved by combining the various nodes' transmissions of the same interface.

The first option (Figure 4-118) targets at combining the reception from the same interface along the relay chain, while the second option (Figure 4-119) achieves diversity by using multiple parallel relays per hop.

Besides the general challenges of multi-hopping (e.g., routing, scheduling), these concepts face the same implications as the first two identified concepts.

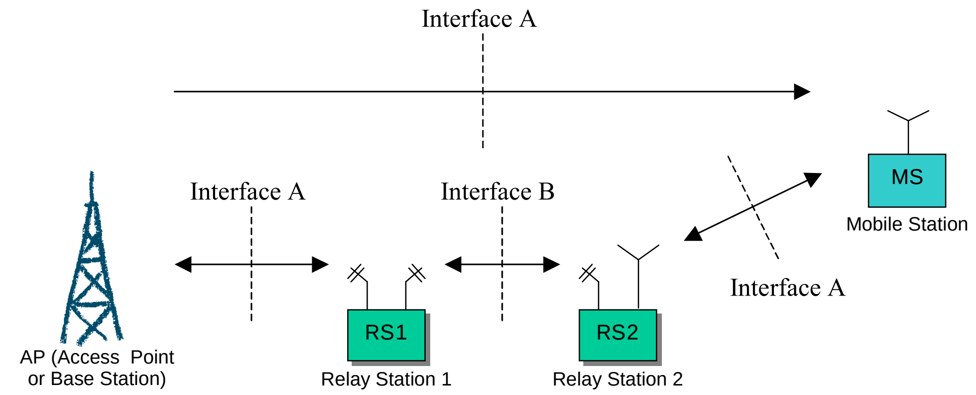


Figure 4-118: Multi-Hoping and Combining I

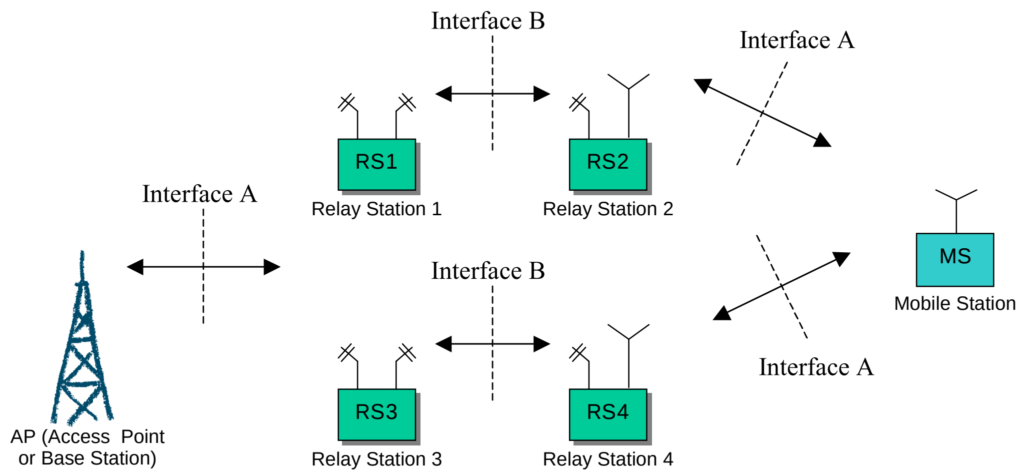


Figure 4-119: Multi-Hoping and Combining II

4.3.5.3 Abandoning the Investigation

Despite these theoretical possibilities, it was decided not to pursue cooperative relaying further in T3.4 (heterogeneous air interface). There are two main reasons, both related to complexity and difficulty of implementation.

- First, cooperative relaying calls for combining the transmissions from different nodes. In the context of heterogeneous air interfaces, this requires combining signals from different air interfaces or, more generally, combining signals of different physical layer modes (PHY modes). The arising complexity seems to be intolerably high as the receiving terminal must be able to simultaneously or consecutively synchronize, demodulate and detect two possibly fundamentally different air interfaces.
- Second, at the network level, the corresponding transmissions and receptions must be coordinated. This coordination in terms of routing, scheduling, and resource assignment already constitutes challenges for conventional store-and-forward relaying and homogeneous air interfaces. These should be tackled first, thereby eventually paving the way for cooperative relaying over heterogeneous interfaces.

To summarize, while it is clear that the theoretical benefits of cooperative relaying are applicable to heterogeneous air interfaces, it is for implementation issues that cooperative relaying focuses on the case of homogeneous air interfaces in WINNER. In fact our work has so far focused on the homogeneous case, where the applicability of cooperative relaying is indeed more straightforward.

4.3.5.4 Summary

Various options for cooperative relaying using heterogeneous air interfaces have been discussed, among them the idea of virtual antenna arrays (VAA), and two multi-stage concepts. We have outlined that applying cooperative ideas to relaying over heterogeneous air interfaces implicates various challenges. It was argued that it is strongly questionable whether the increased complexity justifies the achieved improvements; therefore, it was decided within the work package that effort on cooperative relaying should focus on the homogeneous case. There, application is simpler and concepts are more attractive with respect to their feasibility.

4.4 Others

4.4.1 Wireless Feeder Systems

For the connection between nodes-B and its respective radio network controller (RNC) in current cellular networks (3G), it was contemplated the possibility to use radio links as an alternative to the wire line (E1). Following with this trend, WINNER project will investigate the characteristics and requirements of wireless feeder system focused in some of the scenarios described in the internal report IR3.1. Although at this moment the future architecture for the WINNER access network is not clear yet, the wireless feeder concept to develop in WINNER project will tackle with the feeding of WINNER APs with wireless point to multipoint (PtM) links.

4.4.2 Radio over Fibre

The scope of WINNER project is to develop a new universal radio interface for future mobile communications of B3G. Further in this project will be investigated aspects regarding new network deployment concepts based on wireless relay stations, conveniently distributed in the radio access network. Although WINNER is focused in developing the wireless links between these new elements, user terminals and APs (or base stations) of a given cell, in some particular cases, where optical fiber is available, it could be interesting to use the radio over fiber transmission technology like a possible alternative for the connections between the access point and relays, as well as between relays. Anyway the choice of a specific option in each individual case is of course a network planner responsibility.

On the other hand it exist at the moment an evident trend from the operators point of view for using the radio over fibre technology in the network deployment. In fact, this technology is well known mainly in connections point to point and in mesh topologies. For instance, the European project COST 273 has recently presented a paper [247] proposing, for future WPAN and WLAN applications, a 3D pyramid network for millimeter-wave radio, constitutes by a single access point (mounted on the ceiling) and four active relays (mounted on the walls) operating in parallel, in order to mitigate the impact of shadowing at these high frequencies. From the two possibilities for linking the relays and the access point, that is, by means of radio over fibre method (wired relaying) or employing directional antennas with narrow beams for wireless relaying, it was utilized the RoF method since it provides the most reliable links with minimal cable losses and path delays, as well as having maximum control since the AP may then act as the central control node in this way. It is important to note that in reality a relay connected on the one hand to the AP by optical fibre using RoF technology, and on the other hand to final users by means of a wireless connection using a given radio access technology, may be considered like another case of heterogeneous relay station.

Also for broadband fixed wireless access it is possible to find important references like the recent publication about an experimental demonstration of 622 Mbps millimeter-wave over fiber link for this kind of systems [248]. In this work the transferred RF signal format was based on four multi-carriers in a 37 GHz band with a 64 QAM modulation. The experiments showed that they can successfully transfer a 622 Mbps millimeter-wave signal over a 10 km fibre with less than 5 dB power penalty for a downlink and less than 4 dB CNR penalty for an uplink.

In any case although there is a clear interest for RoF technology investigations and in fact there are currently a lot of works devoted to this topic, this chapter pretends simply to give a general vision of the current state of this technology, but without including new investigations about this topic, since it is not a main issue in WINNER project.

4.4.2.1 Motivation of the use of radio over fibre techniques (pros and cons)

It is globally admitted that in the future generation of cellular networks due to both unavailability of lower microwave frequencies and insufficient bandwidth of lower frequency ranges, next generation of wireless access systems will operate in the upper microwave/millimeter wave frequency band, and so the cells will be much smaller than present generation networks. It is also foreseen a high increase of demand for broadband services, in both fixed and mobile wireless access. Of course for mobile access the only possibility to achieve broadband services is to use some kind of air interface. However for fixed access, that is, communication between fixed elements of access network (e.g. access point and fixed relay stations), there are several possibilities (wireless or wired).

For the wired solution an important emerging technology applicable in high capacity, broadband millimeter-wave access systems is Radio over Fiber (RoF). In this system the signal to and from the involved elements is transmitted in the optical band, via an optical fibre network. This architecture makes design of link between fixed elements very simple, comprising, in general terms, the electrical-to-optical (E/O) and optical-to-electrical (O/E) converters, an antenna and some microwave circuitry (a duplexer and two amplifiers which sometimes can be excluded). In the last decade important research work was done in this field with significant results; and although the number of publications is numerous the most important results are summarized in a recent monograph [249].

Concerning the transport means of RoF technology, that is the optical fibre, there is to remark that basically two types of fibre exist in the market, single-mode and multi-mode. Usually the decision to employ either is based on the desired length to cover. Therefore in terms of the distance between the elements to interconnect it is selected the mode more adequate. In general single-mode fibre is used for distances from above 1 Km (standard single-mode fibre G.652 has an attenuation around 0.3 dB/Km at 1550 nm), high rate requirements and multi-carrier transmission. Multi-mode fibre is more used for smaller areas, underneath of 1 Km, such local area network environment, low rates and with only one carrier. The fibre core is about 9.5 μm for the case of single-mode, and around 50 μm for multi-mode fibre.

On the other hand the RoF technology has advantages and disadvantages as any other technology. Following we summarize the main pros and cons of this technology without analyzing its in depth since as we have mentioned before, the goal of this section is merely informative.

4.4.2.1.1 Advantages

The main pros of this technology derived from its own characteristics are:

- Low attenuation loss. It is clear that electrical distribution of high frequency radio signals in free space (losses due to absorption and reflection increase with frequency) or through transmission lines (impedance increases with frequency) is always problematic and costly. Although in order to avoid this problem it would be possible to use base-band or intermediate frequencies for distribution, this solution on the other hand has severe requirements on repeater amplifier and equalisers, such as linearity, as well on high performance local oscillators. So, an alternative solution for reducing losses would be to use optical fibres as transport means that distribution system. For example, at present there is available standard Single Mode Fibres (SMFs) made of silica with attenuation losses lower than 0.2 dB/km and 0.5 dB/km in the 1.5 μm and the 1.3 μm windows, respectively. Likewise a more recent kind of optical fibres as Polymer Optical Fibres (POFs), exhibit higher attenuation ranging from 10 to 40 dB/km in the 500 to 1300 nm regions. In any case these losses are much lower than those encountered both in free space propagation and copper wire transmission.
- Very high capacity. The available bandwidth of only one fibre is higher than terahertz (e.g. for a single SMF optical fibre, the combined bandwidth of the 850 nm, 1300 nm and 1550 nm windows is around 50 THz, although the current state of the art commercial systems utilize only a fraction of this capacity). So in the case of UMTS it is able to transmit simultaneously several carriers. Further using multiplexing techniques is possible to exploit better the capacity of the fibre. Between the most common multiplexing techniques we can mention SCM (Sub-Carrier Multiplexing) and DWDM. (Dense Wavelength Division Multiplexing). Other important benefit coming from huge bandwidth offered by optical fibres is the possibility to exploit high speed signal processing that in electronic systems may be more difficult or impossible to do.
- Possibility to share infrastructures and operational flexibility. Thanks to multiplexing techniques mentioned before, it is possible to share the fibre with other services even with other operators, adding flexibility to the system. Moreover including some functions at the central node (switching, modulation and so on), it is possible to allocate dynamically the total capacity, distributing optical wavelengths as need arises.

- Immunity against interferences in the optical fibre link. The radio over fibre systems are basically repeaters with frequency change with the advantage of do not undergo any problem of electromagnetic interferences in the link between elements. Thanks to this quality the deployment of radio over fibre repeaters is simpler than the radio repeaters.
- Easy deployment without oscillation and precision frequency problems. Due to the input to repeater is an optical interface it is not possible the necessary feedback for generating oscillations. The requirement of precision frequency neither is a problem since it is possible to send through the optical fibre a clock reference in order to synchronize all the local oscillators of the system.

4.4.2.1.2 Disadvantages

From a general point of view the RoF technology has some disadvantages which can be summarized in the following three points.

- Extra costs on fibre cable installation. Different to wireless technologies, RoF transmission needs optical fibre as transport medium. So, planning RoF system should take into account the availability of optical fibre for interconnecting the network elements, which constitute our system. At the present, in some particular cases such urban populated zones, there are many optical fibres available. However this is not always the case and it is necessary to install optical fibre in order to connect the equipments, falling into extra costs of the total price of our system. Moreover although in office or similar indoor environments is relatively easy to add at general canalisation the optical fibres, most of the times some civil works are mandatory, provoking in this way the increasing of total costs of our system. For example, at the moment the price of an installation of optical fibre in a typical inter-metropolitan area may be around 21000 € per Km, including cables, register boxes and joints in the vain of installation.
- Extra costs on RF-optical front ends. The special optical components needed for RoF system increase the total costs of the complete system. For instance a laser diode selected in wavelength is more expensive than other one of general purpose and in general, optical multiplexer is also more expensive than electrical one. Besides regarding the fibre mode used it is convenient to remark that for fibre single-mode the optical components are more expensive than in the case of fibre multi-mode since the fibre core for single-mode is smaller than multi-mode and so it is required lasers, photodiodes and connectors more precise.
- Mobility and flexibility of location. Obviously the use of RoF technology constraints to dispose the remote nodes (relays) in fixed places, and therefore it has less flexibility in the placement of remote sites than the use of wireless technology, which may achieve similar reliability levels, most specially, when it utilizes directional antennas with narrow beams as long as they maintain LoS conditions with the AP at all times. Moreover, the wireless relays comparatively with relays based on RoF technology, have the advantage of movement in terms of necessity for each situation. In other words, may have isolated cases with high traffic demands wherein the use of wireless mobile relays be cheaper and more effective than the use of relays based on RoF.
- Exposure to aggressive agents. Fibre installation is generally designed for operating correctly during at least 10 years, and so in any RoF system it is contemplated a given loss margin, taking optical fibre degradation into account, in order to ensure a minimum quality level for the transmission through all its lifetime. Nevertheless besides the own ageing of optical fibre it is important to mention other problems which may shorten the lifetime of optical fibre, such attacks of rodent animals (rats), mechanical strains and contaminant agents. Comparing to other wired technologies, the use of optical fibre is very sensitive to mechanical stresses (torsion, elongation, bending, …), causing important attenuations if the installer doesn't take care the laying of optical fibre. Otherwise optical fibre is subjected also to rodent's attacks and polluting elements (the most important contaminant reported has been OH^- ions), which provoke serious degradation of fibre and therefore reducing its lifetime. In any case these problems have been studied and analysed in depth and nowadays they are partial or totally solved thanks to the use of special covers that minimize the effects of these degradation factors.

Relating to particular optical techniques used for generating, distributing and detecting microwave signals in a RoF system, other disadvantages and drawbacks may be mentioned. A detailed analysis of these cons may be found in [250].

- RF generation by direct intensity modulation (IMDD technique). With this method only low RF frequency signals can be generated because to generate signals of higher frequency, the modulating signal must be at the same high frequency. This is not possible for direct laser modulation due to lack of bandwidth and laser non-linearity (inter-modulation product terms that cause distortions). Even using external modulator such as the MZM (Mach Zehnder Modulator), which supports high frequency radio signals, lead to very costly drive amplifiers since they require high drive voltages.

- Photodetector-based optical heterodyning. Most RoF techniques are based on the principle of coherent mixing in the photodiode (Remote Heterodyning Detection, RHD). The photodiode in addition to optical-electrical conversion, also acts as a mixer. The major drawback of this method is the requirement fro the two optical carriers to be phased correlated. That is, the phase noise in the laser directly translates into phase noise of the generated RF signal and therefore, the generated signal is very sensitive to phase noise occurring in the optical link. Due to semiconductor lasers are prone to phase noise, it is necessary to take extra measures in order to minimize the noise, giving rise to more complex systems.
- Optical FM-filter system. This is a single-laser technique for modulating the optical frequency applying an electrical signal to one of the laser's terminals. This generates a series of optical spectral lines or sidebands, all spaced by the drive frequency. Two sidebands are then selected and separated by the required RF frequency. Afterwards, the selected sidebands run into the surface of the photodiode and mix coherently to generate the desired RF signal. There are different methods for filtering the required sidebands, but the major problem is the high accuracy necessary that sometimes lead to very complex circuitry.

Of course there are more techniques for generating microwave and millimeter-wave signals in RoF system with their particular advantages and disadvantages. For instance dual laser methods use two laser diodes to emit light at frequencies separated by the required microwave frequency and due to the laser emission frequency is highly sensitive to temperature variations, phase noise and other effects, it is necessary to employ sophisticated methods, such Optical Frequency-Locked Loop (OFLL), Optical Phase-Locked Loop (OPLL), Optical Injection Locking (OIL) or Optical Injection Phased-Locked Loop (OIPLL), to maintain the required frequency offset in reception (photodiode). Each of them has advantages and disadvantages, but in general terms they are costly and difficult to implement. In reference [252] there is a chapter devoted to the analysis in depth of pros and cons of these techniques and others not mentioned here.

4.4.2.2 Future trends

Although there are many thinks done in RoF technology certainly there are problems, which are yet not solved and so it would be necessary to analyze and investigate in order to achieve more benefits with the use of this technology. Most of the new initiatives and proposals in this sense are focused on the coordination of these systems with the layers of radio management. For example from operators point of view there is an emergent interest in the following particular areas:

- Requirements of a new protocol architecture adapted for heterogeneous networks, based on wireless and wired networks.
- The special case of re-configurable radio over fibre feeders networks and its interaction with a wireless network. This could be a promising solution for an effective use of radio resources available in a particular zone.

Based on predictions, registration of user behaviour and statistical calculations, telecommunication operators have for decades provided the necessary resources to keep blocking probability of the voice service at the level accepted by users. In wireless networks, such as GSM and UMTS, data services are expected to play a more and more important role as the bit rates starts to become acceptable for current Internet application and new multimedia services. Compared to fixed access networks the location of users in these networks are hard to predict and as such the provisioning of resources becomes even more difficult and at the same time resources are typically more expensive and high degree of utilisation is needed. Allocating the maximum possible capacity for every single user in any location and time period would be unrealistic and very expensive.

In a fibre-fed re-configurable network, as Figure 4-120. illustrates, the head-end that is in charge of all resource administration should be able to move the available capacity between remote antenna sites in order to make an efficient use of the network resources, that is, the finite capacity is dynamically distributed to different remote antenna sites as its is required. In general terms not all the base stations will be supporting full traffic capacity all the time; most of them will be only supporting a heavy load during specific periods of the day or the week. So it would be a good idea to have some pools of resources (big central locations with many base stations) and many antennas connected to each pool by means of a re-configurable RoF network. In this way less resources will be necessary to support the traffic load, as the capacity can be moved to where it is needed.

On the other hand in current implementations, the base stations or the RoF systems are not directly involved in any resource administration, and only very limited re-configuration capabilities are available. Also, it has not been foreseen any co-operation between the RoF management system and the wireless system that is transported through the fibre, and great improvements in network efficiency could be expected if a co-operation would exist.

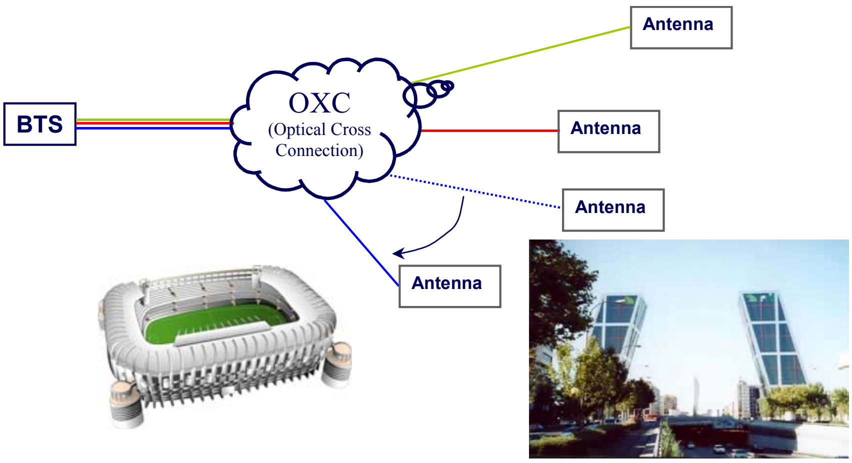


Figure 4-120: Re-configurable RoF system proposal

Finally and trying to summarize the main aspects to investigate in the future of RoF systems, following we list some of the open problems in this area:

- Long-term and short-term traffic modelling to be used in RoF networks reconfiguration:
    - Intra-day predictions.
    - Week-end and labour days predictions.
    - Seasonal predictions.
    - Traffic location predictions.
    - Short-term predictions based on current network traffic data.
- Predictive (*a priori*) and responsive (*a posteriori*) RoF network reconfiguration strategies and protocols:
    - This knowledge gap could be between “long term traffic modelling” and “Radio over Fibre interaction with other layers”. A predictive RoF reconfiguration is more related with long term traffic modelling, and a responsive RoF reconfigurations has key relation wireless protocols interaction.
- RoF-network protocols interaction with wireless protocols.
    - Wireless data is useful for traffic analysis and RoF network reconfiguration, but not available up to now.
    - Who decides RoF system switching?, should be defined a new layer below the physical wireless layer?[14]
    - Wireless connections affected when the RoF system is reconfigured.
    - Use some kind of soft handover before RoF reconfiguration?, how is it implemented?.
    - How is updated the mobile active set?
    - Planning issues as pilot power assignment in a re-configurable network.

[14] There was a similar situation in the 90’s, when the new optical functionalities (Optical Amplifiers, Optical Add Drop Multiplexers) forced the definition of an Optical Layer below the Physical Layer in SDH (Synchronous Digital Hierarchy).

- RoF OA&M (operation, administration and management) and QoS data supported by its own network layer?. Any interaction regarding this data with the upper wireless layers?. Define a set of relevant RoF OA&M and QoS data.

5. Conclusions

The deliverable D3.1 has shown some very promising concepts regarding the WINNER radio network deployment. The WINNER radio network deployment differs from the today's systems as it has to cope with some new features introduced in WINNER which are

- Multi RAT mode capability of the new WINNER air interface for optimal solutions in different scenarios
- The integration of relays into the radio network architecture

Therefore one important feature of the proposed architecture is that it can make use of multiple radio interface operating modes. Consequently, different sets of multiple access protocols, duplex techniques and radio link control protocols can be used for different operational scenarios. Hence, the proposed protocol architecture can support a large variety of completely different deployment scenarios such as wide-range and short-range communication as well as integration of conventional cellular system infrastructures with multi-hopping and relaying. These radio interface modes may be viewed as co-designed radio interface technologies optimized for different deployment and user scenarios that can interact seamlessly

The concept to integrate multiple radio interface operating modes is that the radio interface functions are divided into generic user plane and control plane functions such that independent scalability, dimensioning and evolution of the user and control planes is achieved. Therefore generic and specific part of the RRM and Link Layer have been identified. The concept of generic link layer (GLL) provides a common interface towards upper layer functions/protocols by generalizing all common link layer functions for different modes. The GLL may be viewed as a toolbox of link layer functions that may be adapted to the characteristics of the different supported modes. The same is performed for the RRM functionalities. In addition the concept of the Mode convergence Protocol has been proposed to enable a well structured design of B3G protocols of layer 2 and 3 according to the various operating modes. This MCP allows by means of cross layer optimisation and cross plane management a context transfer for an efficient handover between two radio modes. The MCP allows e.g. the transfer of status information of the ARQ mechanism. Compared to today's approaches to integrate two existing systems by identifying common functional blocks the WINNER approach can deal with this issue already during the design phase of the air interface which allows to harmonise these functionalities already from the beginning.

The concepts of GLL and MCP were taken up for the development of heterogeneous relays based on two WINNER modes as the GLL and the MCP would allow an efficient integration of such a network element.

The integration of the relays into the radio network architecture has been taken up the proposed logical nodes architecture representing the relay either connected to the AP or another relay station. The proposed flexible logical node architecture model is also able to reduce the inter node interface definition, standardization and maintenance effort. It is envisioned that the logical architecture model can result in a lower number of distinct interfaces than conventional (restrictive) deployment models. It is desirable that the node architecture can support all envisioned deployment scenarios for WINNER without introducing too many logical nodes and/or interfaces.

Further some relays concepts have been presented which are based on a MAC frame based approach as modern wireless broadband air interfaces are based on MAC frames, the only exemptions being IEEE802.11a/b/g but 802.11e uses a periodical Medium Access, too. MAC framed air interfaces have been established to be useful for relaying in the time domain by just using the functions available from the existing standards. Deployment concepts using fixed relay stations have been shown to be of high benefit to substantially reduce the effort of interfacing APs to the fixed network (owing to a substantial reduction of APs needed). Relays have been proven to substantially extend the radio coverage of an AP, especially in highly obstructed service areas. Gain antennas at FRSs have been established to substantially contribute to increase the throughput at cell areas far away from an AP.

As an additional approach the concept of cooperative relaying considers the impact of various system resource constraints on the connectivity and performance of relaying channels. System connectivity models are derived from the constraint combinations, and minimum cost constraint sets are given. Simulations are presented that show the impact of the individual system resource constraints on amplified relaying, decoded relaying with error propagation, and decoded relaying without error propagation.

The impact is indicated to be different for the various relaying methods. The results indicate that for amplified relaying the priority is to maximize the connectivity of the destination terminal, while for decoded relaying (with and without error propagation) the priority is to equalize the connectivity of the destination and relay terminals. For amplified relaying the diversity order of the system is dependent on the connectivity of the destination. For decoded relaying with error propagation the diversity order of the system is constrained by the connectivity of the minimally connected relay and therefore limited to one. For decoded relaying without error propagation the diversity order of the system is dependent on the connectivity of both the destination and relays. Guidance is provided for the order in which the various constraints should be lifted to maximize connectivity and performance. Thereby the concept of cooperative relaying might also be adapted as complementation of "conventional" relaying concept.

The overall goal of the routing and forwarding mechanisms is to facilitate two-hop or multi-hop communications through efficient and intelligent forwarding and routing algorithms, respectively, in relay-based cellular, WiFi, and WiMax networks. Pure ad hoc networks which do not rely on any infrastructure (Mobile Ad-hoc NETworks [MANETs]) are not a big issue in the scope of WINNER.

Thereby it was shown that for high signal to noise values, a single hop has better spectral efficiency compared with a *n*-hop replacement; for small SNR values, a more efficient *n*-hop link is possible, with the optimal locations of multi-hop regenerative relays being at equal intervals along the straight line between the source and destination. A novel quantitative criterion is developed, which offers threshold mean SNR values below which a n-hop replacement should be considered over a single-hop link. Additional research into the statistical distribution of spectral efficiencies for multi-hop links may bring further clarifications on the properties of cellular systems using fixed relays for multi-hop communications.

About random generated topology, Wireless Fixed Routing (WiFR) has shown to give advantages especially in networks of medium and large dimensions, in terms of number of nodes, where is able to enhance overall network performances admitting a greater amount of traffic with low network connectivity with respect to a fully connected network. Further, a trade-off related to network connectivity between possibility of finding several possible feasible routes and number of relays influenced by a each single transmission has been highlighted and it has been shown that best results are achieved when network connectivity is low but allows anyway to different feasible routes for a given connection. Simulations have shown that, especially for medium and high traffic matrixes, best results are achieved when two additional relays are inserted between two neighbour basic relays that act as source and/or destination of connections. WiFR has proven that working in pre-planning modality, is able to support very well both CBR traffic and exponential traffic sources generating packet bursts even when each connection requires a bandwidth equal to its average rate bit always offering Quality of Service in a high satisfactory way.

After the introduction of new concepts for a B3G radio network deployment and related architecture the next steps will focus on the further development of the concepts. Thereby the integration of the newly identified radio interface technologies of WP2 will take an important role. First promising concept for the integration of OFDMA have been presented in D3.1 already. Also some ideas on future step like the exploitation of the spatial domain, where smart antenna concept can play a leading role.

The future work for the radio network and protocol architecture is highly dependent on the development of the WINNER radio interface and its different operating modes. As already mentioned, the architecture proposal is still preliminary and therefore some effort should be concentrated to clarify the radio interface specificities, its architecture, its functional elements and functional distribution.

6. Annex I: Benchmarks and Link metrics

6.1 Detailed common set of system level parameters

Benchmarking simulations are carried out based on the existing conventional single-hop radio interfaces in order to compare and assess the system performance in different deployment scenarios taking into account the characteristics of enhanced radio interfaces comprising advanced radio transmission techniques, new strategies and algorithms.

Since the performance evaluation of the radio interfaces depends highly on the level of the definition of the proposed scenarios and system parameters to ensure the "comparability" of the results, a well-defined common set of system parameters is required. In this way the proposed common set of system level simulation parameters should be detailed enough to be used across different radio interfaces technologies for system benchmark. The parameters are defined in terms of environment type, user location, mobility modelling, equipment (MT, BS/AP) characteristics, and channel models. The defined set comprises the input and output parameters (as figures of merit - FoM) aligned with the recommendations of the D7.2 [18].

6.1.1 Input parameters

The set of input parameters is aligned with the definition of parameters on the D7.2 [18]. The parameters are defined in different aspects taking into account system deployment characteristics, users mobility, traffic types, link level parameters, channel characteristics, power channel control and RRM aspects.

Table 6-1: Common set of input parameters

Input parameter group	*Input parameter*	*Comments/description*
Cell deployment and configuration	Cell type (pico, micro, macro)	Defines the cell coverage area
	Network topology	PMP topology, mesh topology, etc
	Antenna type	Defines the characteristics of the antenna and its radiation pattern (omni, sectorised, beamformed)
	AP/FRS position	Defines the position of BS/AP (roof top, below roof, etc)
	Site-to-site distance	Inter-distance between adjacent AP/FRS
Link level parameters	BS/AP Power	BS/AP total maximum output power
	Terminal Power	Terminal maximum output power
	BS/AP downlink power	BS/AP maximum downlink channel power
	Terminal uplink power	Terminal maximum uplink channel power
	BLER vs SNIR curves	BLER vs. Eb/No curves over different types of propagation channels, velocities, modulation schemes and code rates
	PER vs SNIR curves	BLER vs. SNIR curves over different types of propagation channels, velocities, modulation schemes and code rates
Channel type and control parameters	Channel bandwidth	Minimum channel bandwidth required to transmit at certain data rate
	Channel data rate	Maximum transmission bit rate over the channel bandwidth
	Path loss	Defines whenever the path is under LOS or NLOS conditions (see D5.2)
	Link delay	Maximum link delay allowed
	Radio propagation model	Propagation model used to estimate the path loss
Traffic characterization	Traffic type	"Conversational Voice (VoIP)", "Video Streaming", "Audio Streaming", "Web Browser", File Transfer (FTP))
	Delivery requirement	Real-time (RT), non-RT (NRT), etc
	Traffic directionality	Unidirectional (UNI) or bi-directional (BID)
	Connection symmetry	Symmetric or Asymmetric and the respective asymmetry factor
	Mean call duration	
	Mean call arrival rate	
Mobility	Mobility type and speed	Stationary; Pedestrian & Low Mobility Speed; Medium Mobility Speed (urban/suburban vehicular); High Mobility Speed (highway) vehicular
	Terminal density	Uniform, hotspot other distributions, typical urban, typical suburban, etc)
RRM mechanisms	Power control	
	Link adaptation	
	Packet scheduling	
	Handover	

6.1.2 Output parameters

The following results should be used to assess the system performance and mostly are commonly used and independent of the radio technology. The out put parameters shall be aligned with D7.2 as figures of merit (FoM).

Table 6-2: Common set of output parameters

Output parameter group	*Output parameter*	*Comments/description*
Coverage	% area for a minimum sustained user throughput	Defined as % area to guarantee a minimum user throughput (conversional, 5Mbps, 50Mbps, etc)
Throughput	Packet call throughput	Defined as (good bit received)/ (packet call delay)
	Average user packet call throughput	Defined as (sum of the packets call throughputs for user)/ number of packets calls for user
	Packet call throughput at cell edge	Defined as CDF of packet call throughput at cell edge
Fairness	Packet call throughput as function of distance from site	Mean distance from the site for each packet call throughput
Delay	Packet call delay	Defined as CDF of user's packet call delay experienced for 95% of the user's packet calls

6.2 System benchmarking results

6.2.1 3G (HSDPA)

In an Enhanced UMTS network the total available bit rate must be well above 2 Mbps and even higher in hotspots for packet switched services. The design of wireless techniques that effectively achieve 10 Mbps or higher, as total bit rate without additional complexity is one of the objectives of this technology.

The introduction of HSDPA (High Speed Downlink Packet Access) enhances 3G mobile systems by offering higher data rates in the downlink direction.

To support an evolution towards more sophisticated network and multimedia services, the main target of HSDPA is to increase user peak data rates, quality of service, and to generally improve spectral efficiency for downlink asymmetrical and bursty packet data services. This is accomplished by introducing a fast and complex channel control mechanism based on a short and fixed packet TTI (Transmission Time Interval), hybrid ARQ and AMC (Adaptive Modulation and Coding). Spectrally efficient modulation schemes, e.g. 16QAM and 64QAM, turbo codes and other suitable codes for packet transmission are possible enhancements. To facilitate fast scheduling with a per TTI resolution in coherence with the instantaneous air interface load, the HSDPA-related MAC functionality is moved to the Node-B. It relies on a new type of transport channel, the HS-DSCH, which is terminated in the Node B. HS-DSCH is applicable only to PS domain RABs. The High Speed Downlink Shared Channel is a downlink transport channel shared by several UEs. The HS-DSCH is associated with one downlink DPCH, and one or several Shared Control Channels (HS-SCCH). The HS-DSCH is transmitted over the entire cell or over only part of the cell using e.g. beam-forming antennas.

The HSDPA concept facilitates peak data rates exceeding 2 Mbps and the cell throughput gain over previous UTRA-FDD releases has been evaluated to be in the order of 50-100% or even more, highly dependent on factors such as the radio environment and the service provision strategy of the network operator. Summarizing, HSDPA is a packet-based data service in WCDMA downlink with data transmission up to 8-10 Mbps (and 20 Mbps for MIMO systems) over a 5MHz bandwidth in WCDMA downlink. HSDPA implementations includes Adaptive Modulation and Coding, MIMO, Hybrid ARQ, fast cell search, and advanced receiver design.

Compared to the Release'99 architecture (see Figure 6-1), HSPDA introduces a short 2ms transmission time interval (TTI), adaptive modulation and coding, multicode transmission, fast physical layer (L1) hybrid ARQ (H-ARQ), and moves the packet scheduler from the RNC (Radio Network Controller) to the Node-B where it has easy access to air interface measurements. The latter facilitates advanced packet scheduling techniques, meaning that the user data rate can be adjusted to match the instantaneous radio channel conditions.

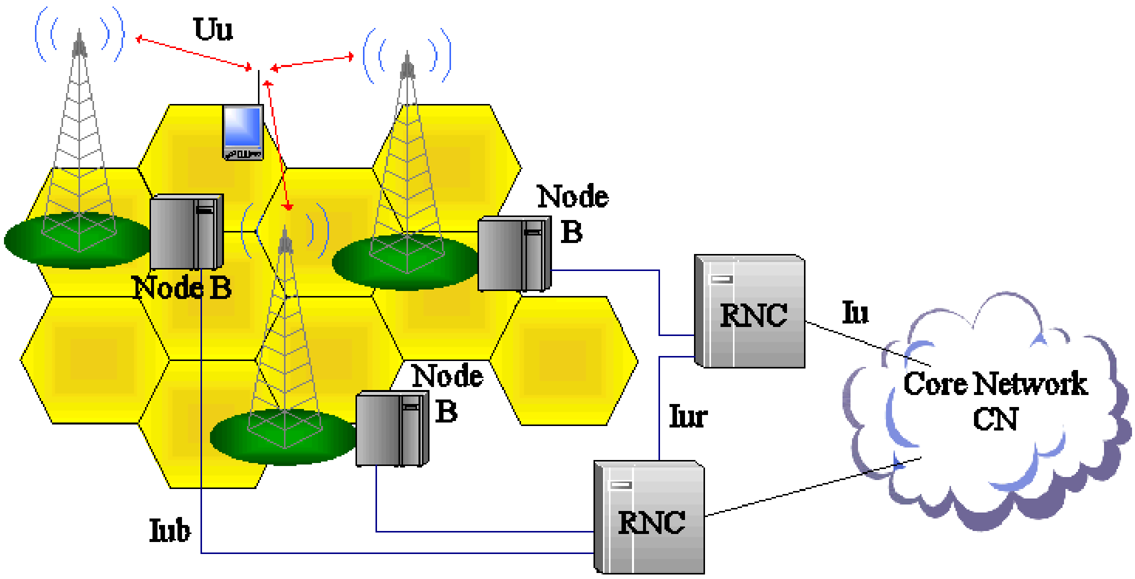


Figure 6-1: UMTS Release'99 architecture [12]

The basic technologies being considered for HSDPA rely on the rapid adaptation of transmission parameters to match channel conditions. Accordingly, the corresponding functions, such as fast link adaptation and fast scheduling, should be placed close to the air interface, preferably in Node B. In the present-day WCDMA architecture (Figure 6-1), the scheduling and transport-format selections are performed in the RNC. Thus, for HSDPA, it is advantageous to move parts of the functionality from the RNC to Node B, forming a new Node B entity, MAC-HSDPA.

However, some RNC entities, such as RLC (Radio Link Control) and MAC, should remain in their current location. The functions of the RNC include ciphering and in-order delivery of data (these are functions provided by the RLC). In soft-handover scenarios between two separate Node Bs, the RNC can guarantee that no data will be lost even if the hybrid ARQ mechanism in Node B fails. The extended features of Node B should thus not be seen as a replacement of the RNC, but rather as a complement, which (in terms of the RNC) provides a highly reliable channel that supports high data rates.

In what concerns the general operating principles of HSDPA, a basic example is depicted in Figure 6-2. While connected, HSDPA user equipment periodically sends a CQI (Channel Quality Indicator) to the Node-B indicating what data rate (and using what coding and modulation schemes and number of multicodes) the user equipment can support under its current radio conditions.

The user equipment also sends an acknowledgement (Ack/Nack) for each packet such that the Node-B knows when to initiate retransmissions. With the channel quality measurements available for each UE in the cell, the packet scheduler may optimize its scheduling among the users. For example, in Figure 6-2 the channel scheduler bases its selection on the highest available channel quality.

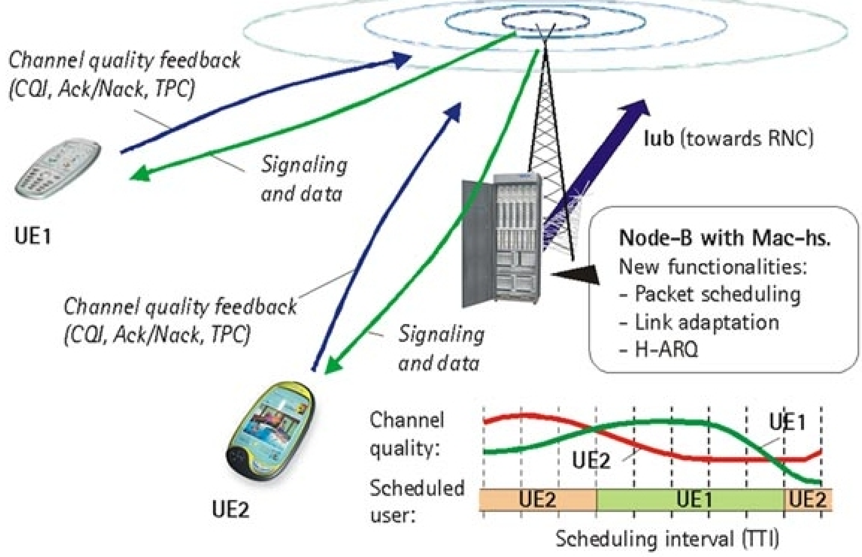


Figure 6-2: General operating principles of HSDPA [13]

A possible deployment scenario for a typical urban environment is depicted below.

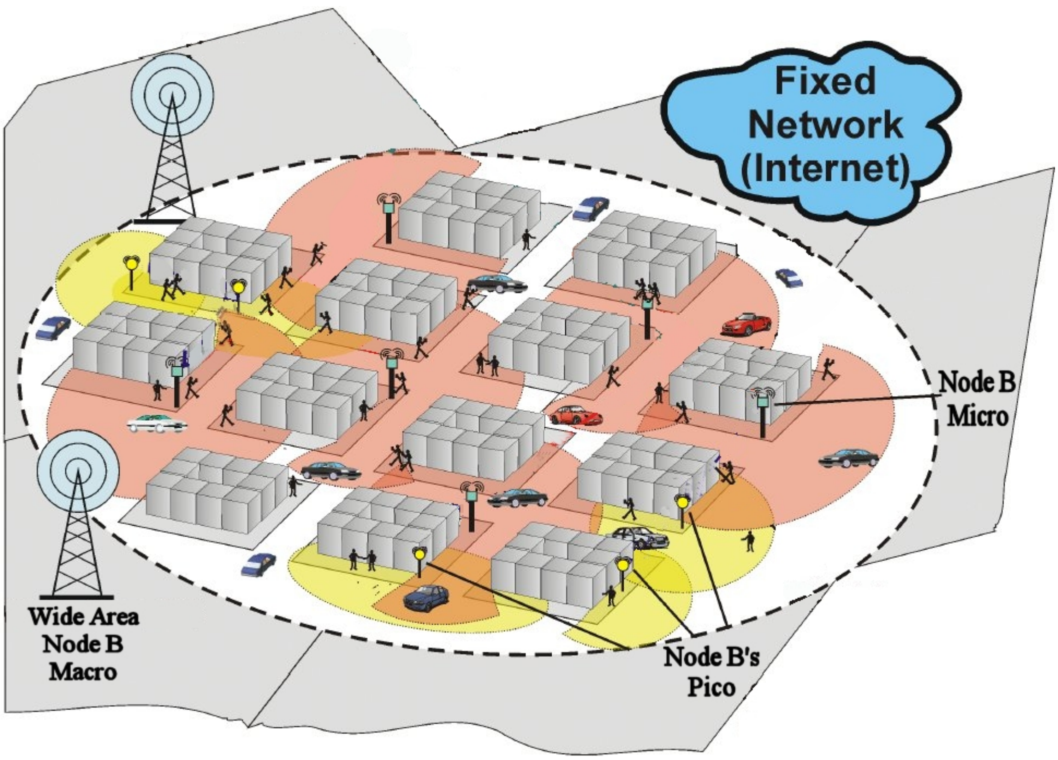


Figure 6-3: Deployment scenario for UMTS-HSDPA [20]

A deployment scenario like this would involve some HSDPA deployment parameters like those listed in Table 6-3 below.

Table 6-3: HSDPA deployment parameters

FEATURE	**PARAMETER**
Aggregated BS throughput	10Mbps
Channel data rate	960kbps (16-QAM) 480kbps (QPSK)
Medium access	FDD – CDMA: uses the remaining code power on the BS
Transmission/way	Multi-code capability. Unidirectional – downlink only Possibility to broadcast in the entire cell
Channel bandwidth	5 MHz
Channel type	High Speed Downlink Shared Channel - HS-DSCH – many users sharing the same physical channel HS-PDSCH - High Speed Physical Downlink Shared Channel
Modulation	Dynamic for first transmission and retransmission. There shall be mandatory support for QPSK and 16QAM
Channel coding	Type of error protection: turbo code rate 1/3.
Mobility	Very Low <5km/h Low 5km/h
Coverage/Range	Hotspot – short range
Spectrum usage	Dedicated
Networking Infrastructure capabilities	Dedicated, point-to-point BS - MS
Environment deployment	Pico and micro cellular ***WINNER scenarios: 1 (In building); 2 (Public Hot spot area); 3 (Urban)***
Traffic type	Suitable for High speed data traffic services
OTHER FEATURES	H-ARQ – Hybrid ARQ: combining schemes; AMC – Adaptive modulation and coding; FCS – based on CPICH Ec/Io measurements in the downlink. The UE continuously measures CPICH Ec/Io for each base station (or sector) added to the active set (in DCH mode) MIMO – associated to adaptive antenna to minimize the interference; Smart antennas and beamforming – to increase capacity and reducing overall system interference; STTD – Space-time transmit diversity increases the number of resolvable multipath components. In the reception paths remain orthogonal,

6.2.1.1 Simulation Overview

This section describes the performance of UMTS HSDPA as benchmark for comparison. The simulations were carried out using a system level dynamic simulation tool developed by Vodafone Group R&D. Further details of the tool can be found in [PIMRC reference]. Simulations are carried out for both single and dual antenna Rake receivers at the terminal.

6.2.1.2 Simulation Parameters

An area of 20km2 was simulated comprising of 12 hexagonal sites with 3 sectors each with a site-to-site spacing of 1..5km. 50 users per sector were uniformly distributed in the centre site and all surrounding sites transmitted on full power. Results were recorded for the centre site only.

Other simulation parameters are given in Table 6-4 below. The users were stationary in location but the lognormal and fast fading were updated according to the parameters in Table 6-4.

Table 6-4 General Simulation Parameters

Parameter	Value
Simulation length	30 seconds
Pathloss	128.1 + 37.6 log10 (d)
Carrier frequency	2GHz
Lognormal fading Sector correlation Site correlation Standard deviation Correlation distance	 1.0 0 10dB 50m
Fast fading	Veh A (50km/h)
HSDPA resource Maximum power Codes	 80% node B power 10
Serving cell selection	CPICH Ec/N0
Node B power	43 dBm
Control channels CPICH SCH Others	 33dBm 30dBm 33dBm

All the users in the central site are in active session with the service as web-browsing, modelled as in below.

Table 6-5 WWW Model Parameters

Parameter	Random Variable	Values
Packet Call Size	Pareto with cutoff	A = 1.1, k = 4.5Kbytes M = 2 Mbytes μ = 25 Kbytes
Time between packet calls	Geometric	μ = 3 seconds
Packet Size	Segmented based on MTU size	1500 octets
Packets per Packet Call	Deterministic	Based on packet call size and packet MTU
Packet Inter-arrival Time	Geometric	μ = 6 ms

6.2.1.3 Alignment with D7.2 configurations/parameters

The situation modelled is based on test scenario C2 – typical urban metropolitan ubiquitous coverage using the typical urban test environment described in [18].

6.2.1.4 Results

Table 6-6 shows the headline benchmark figures for HSDPA.

Table 6-6 Site benchmarks for HSDPA

UE Receiver	Antennas (Transmit, Receive)	Scheduler	Site Throughput (Mb/s)	Site Spectral Efficiency (b/s/Hz/site)
Rake	(1,1)	Round robin	3.16	0.63
Rake	(1,1)	Max C/I	6.55	1.31
Dual Rake	(1,2)	Round robin	3.66	0.73
Dual Rake	(1,2)	Max C/I	7.84	1.57

As expected the max C/I scheduler provides an increase in the site throughput as users in good radio conditions are scheduled in preference to those in poor conditions. The additional of a second antenna at the UE brings some gain in performance (~20% increase in site throughput) but is limited due to fact that there is already significant time diversity from the Veh A channel.

6.2.1.5 Conclusions

Benchmark results for HSDPA have been presented for a wide-area coverage deployment scenario for single and dual antenna Rake receivers at the UE. The performance of other multiple antenna schemes including transmit diversity and VBLAST is reported in [PIMRC reference].

6.2.2 WLAN 802.11

The design and evaluation of a new system like WINNER must go with the study of the already existing technologies in order to allow an effective comparison of the performance provided by new solutions. With respect to the coverage of small areas, two main scenarios have been taken into account: indoor environments (e.g. office, home, etc.) and public hotspots (e.g. airports, hotel lobbies, coffee shops, etc.). These two scenarios are today usually addressed by the WLAN systems. The main attraction of WLANs is their flexibility. They can extend access to local area networks, such as corporate intranets, as well as support broadband access to the Internet particularly at hot spots, public venues where people tend to gather. WLANs can provide quick, easy wireless connectivity to computers, machinery, or systems in a local environment where a fixed communications infrastructure does not exist or where such access is not permitted. These hosts can be stationary, handheld, or even mounted on a moving vehicle.

The most common WLAN standard today is IEEE 802.11 (http://grouper.ieee.org/groups/802/11/index.html). Originally released in 1997, many task groups have since been created to enhance alternately the MAC layer or the PHY layer in order to achieve better performance in the unlicensed 5 GHz UNII band as well as in the original 2.4 GHz ISM band. Although the existence today of many standards spun off by the originally IEEE 802.11 release, this contribution will be exclusively devoted to the benchmarking of the 802.11a version. Nevertheless, in the following paragraphs, along with some technical highlights related to the MAC protocol of that system and some possible enhancements, a brief overview of the 802.11 standard evolutions will be provided.

6.2.2.1 The 802.11 standard evolution

The past 5 years have witnessed the emergence of Wireless Local and Personal Area Networks (WLAN/WPAN) in the home, enterprise, and public access environments. In the enterprise, high speed WLAN represents a flexible alternative or complement to wired Ethernet. This provides motivation for continuing to increase the available data rate beyond 100 Mbit/s, along with stringent security requirements. In public access scenarios, WLANs have the capability to provide high-speed Internet access, requiring an optimal trade-off between bit rates and range. Further, the home environment represents a number of significant challenges, namely the simultaneous distribution of High Definition video, high speed Internet, and telephony inside the house. In this contest, the IEEE 802.11 protocol operating in both the 2.4 GHz ISM and 5.2 GHz UNII bands has enjoyed spectacular market success under the Wi-Fi name.

Besides the United States based IEEE committee, the ETSI BRAN in Europe and the ARIB MMAC in Japan have also defined two standards for WLAN technology: the ETSI BRAN HIPERLAN/2 and ARIB MMAC HiSWANa.

The original 1997 standard defined a MAC based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), two PHY layers in the 2.4 GHz band, and a third PHY based on Infrared (IR). The 2.4 GHz PHY employs Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) respectively to achieve between 1 and 2 Mbit/s. Two years later the task group 802.11b ratified a new standard for the PHY layer operating at 2.4 GHz. This specification introduces Complementary Code Keying (CCK) capable of data rates of 5.5 and 11 Mbit/s. This is clearly the most successful WLAN standard to date.

Following 802.11b was 802.11a that specified a new PHY for the 5GHz UNII band. The system was initially designed to operate in the 5.15GHz-5.35GHz and 5.725GHz-5.825GHz bands allocated in the United States, in channels spaced by 20 MHz. The transmission technique employed is Orthogonal Frequency Division Multiplexing (OFDM), which when combined with convolutional coding and bit interleaving is very robust against multi-path propagation, especially in Non Line Of Sight (NLOS) conditions.

Another 2.4 GHz PHY layer specification was developed by Task Group G. The 802.11g PHY layer includes a mandatory OFDM mode in order to achieve 802.11a data rates. Optional 8PSK and CCK-OFDM modes are also included. Claimed advantages of 802.11g are the backward compatibility with the widespread 802.11b products, and a larger cell range compared to 802.11a. Nevertheless, the 2.4 GHz ISM band provides only 3 non-overlapping channels and is already crowded by systems such as Bluetooth.

The 802.11 Working Group has recently determined the technical requirements to enhance the current 802.11 standard. This work, performed by the 802.11 High Throughput Study Group, led to the creation of the 802.11n Task Group. Within this group some proposals are currently being considered to considerably improve the performance of the system.

6.2.2.2 Technical description

The IEEE 802.11 standard defines both an infrastructure mode, with at least one central access point connected to a wired network (Basic Service Set – BSS), and an ad hoc or peer-to-peer mode, in which a set of wireless stations communicates directly with one another without needing a central access point or wired network connection (Independent Basic Service Set – IBSS). Since the consortium decided to neglect in the first phase of the project the pure ad hoc systems, we here roughly describe the BSS only.

The basic access method in the IEEE 802.11 MAC protocol is the distributed coordination function (DCF) which is a carrier-sense multiple access with collision avoidance (CSMA/CA) MAC protocol. In addition to the DCF, the IEEE 802.11 also incorporates an alternative access method known as the point coordination function (PCF) — an access method that is similar to a polling system and uses a point coordinator (PC) to determine which station has the right to transmit. The DCF access method, hereafter referred to as basic access, is summarized in Figure 6-4. When using the DCF, before a station initiates a transmission, it senses the channel to determine whether another station is transmitting. If the medium is found to be idle for an interval that exceeds the distributed inter-frame space (DIFS), the station continues with its transmission. On the other hand (i.e. the medium is busy), the transmission is deferred until the end of the ongoing transmission. A random interval, henceforth referred to as the backoff interval, is then selected, which is used to initialize the backoff timer. The backoff timer is decreased for as long as the channel is sensed as idle, stopped when a transmission is detected on the channel, and reactivated when the channel is sensed as idle again for more than a DIFS. The station transmits when the backoff timer reaches zero. The DCF adopts a slotted binary exponential backoff technique. In particular, the time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each slot time, which is equal to the time needed at any station to detect the transmission of a packet from any other station. The backoff time is uniformly chosen in the interval (0, CW-1) defined as the backoff window (contention window). At the first transmission attempt, $CW=CW_{min}$, and it is doubled at each retransmission up to CW_{max}. In the current standard version, $CW_{min}=16$ and $CW_{max}=1024$. Immediate positive acknowledgments are employed to ascertain the successful reception of each packet transmission (note that CSMA/CA does not rely on the capability of the stations to detect a collision by hearing their own transmission). This is accomplished by the receiver (immediately following the reception of the data frame) which initiates the transmission of an acknowledgment frame after a time interval, short interframe space (SIFS), which is less than the DIFS. If an acknowledgment is not received, the data frame is presumed to have been lost, and a retransmission is scheduled.

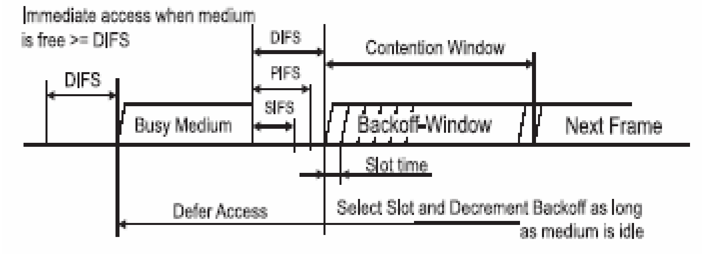


Figure 6-4: Basic Access Method

Generally, the PCF operates by stations requesting that the PC register them on a polling list, and the PC then regularly polls stations for traffic while also delivering traffic to the stations. The PCF makes use of the PIFS that is shorter than the DIFS, to seize and maintain control of the medium. The PC begins a period of operation called the contention-free period (CFP), during which the PCF is operating. This period is so called because access to the medium is completely controlled by the PC. The CFP begins when the PC gains access to the medium, using the normal DCF procedures, and transmits a Beacon frame. The CFP also alternates with a contention period (CP) where the normal DCF rules operate (Figure 6-5).

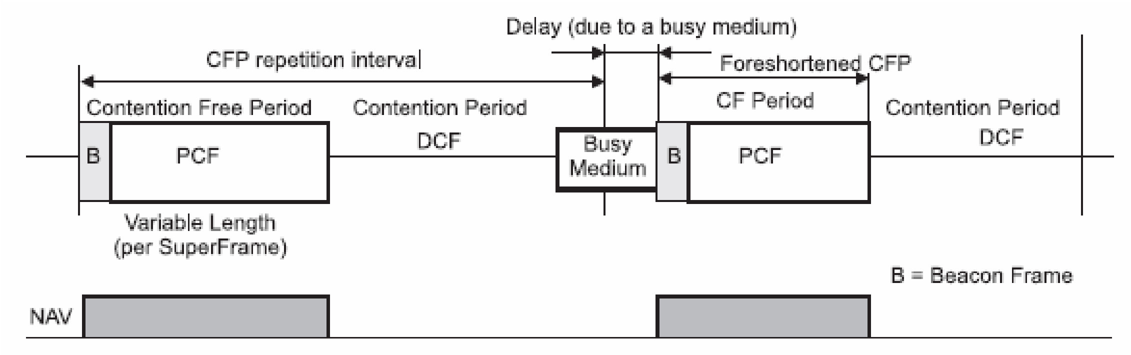


Figure 6-5: CFP/CP alternation

An improvement of this protocol is being studied in the 802.11 Task Group e (currently being finalized). This standard adds several mechanisms to the existing legacy 802.11 MAC in order to improve the QoS support. In addition to the CSMA/CA access technique of the Distributed Coordination Function (DCF), a Hybrid Coordination Function (HCF) is introduced. During the CFP, the Hybrid Coordinator (typically located in the Access Point) schedules the transmission opportunities of the terminals, based on individual requests. The Enhanced Distributed Channel Access (EDCA), which is the contention based channel access mechanism of HCF, supports soft QoS by introducing Access Categories (ACs): packets are delivered through multiple backoff instances per station, each backoff instance being parameterized with AC specific parameters. A mapping of the priorities on ACs is performed. Another feature of interest in 802.11e is the so-called Block Acknowledgement. After contending for a transmission opportunity, the terminal gaining access to the channel is able to transmit several packets with a single acknowledgement at the conclusion of the data burst.

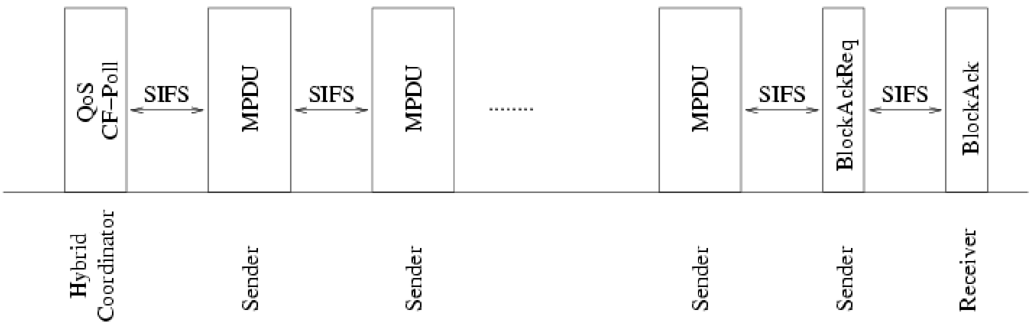


Figure 6-6: Block Acknowledgment mechanism in IEEE 802.11e

6.2.2.3 Simulation Overview

In this paragraph, we present some simulation results aiming at benchmarking the short range scenarios (e.g indoor and hot-spots) with the performance of existing 802.11a standard.

6.2.2.4 Simulation Parameters

The simulations have been run taking into account a set of parameters that are not supposed to change from one scenario to another (unless specified).

6.2.2.4.1 Link level parameters

Some inputs related to the PHY layer are required to conduct MAC level simulations. These have been previously identified as the transmitted power (considering that no transmit power control algorithm is implemented) and the PER vs. SNR curves that provide a direct performance of the PHY layer. These curves are used to implement an error model in the simulator (knowing the SNR through a computation of the link budget, we can derive the PER thanks to these curves).

The transmit power of the AP and station is 15dBm (40mW). This characteristic has been chosen as a feature for small devices with battery: the power consumption is a key aspect of such devices, and the transmit power plays a big role in the power consumption.

Another device receiver characteristic is the noise floor fixed at -90dBm.

The PER curves (used as inputs to the system simulator) are provided in the figures hereafter (Figure 6-7 and Figure 6-8). They are different depending on the constellation (BPSK to 64QAM) used, and the coding rate (1/2, 2/3 or 3/4) adopted. The two figures correspond to different packet sizes. It is reasonable to estimate a proportional relationship between the results and the packet size. The parameters of the various modes are summarized in Table 6-7.

Table 6-7 - PHY Modes of IEEE802.11a

Number	**Modulation**	**Code Rate R**	**PHY bit rate D**
0	*BPSK*	1/2	6 Mbit/s
1	*BPSK*	3/4	9 Mbit/s
2	*QPSK*	1/2	12 Mbit/s
3	*QPSK*	3/4	18 Mbit/s
4	*16-QAM*	1/2	24 Mbit/s
5	*16-QAM*	3/4	36 Mbit/s
6	*64-QAM*	2/3	48 Mbit/s
7	*64-QAM*	3/4	54 Mbit/s

The PER performance of the various modes is plotted on figure 1 for a typical 5 GHz indoor office Non Line Of Sight (NLOS) channel with 50ns r.m.s. delay spread, corresponding to the ETSI BRAN-A channel model.

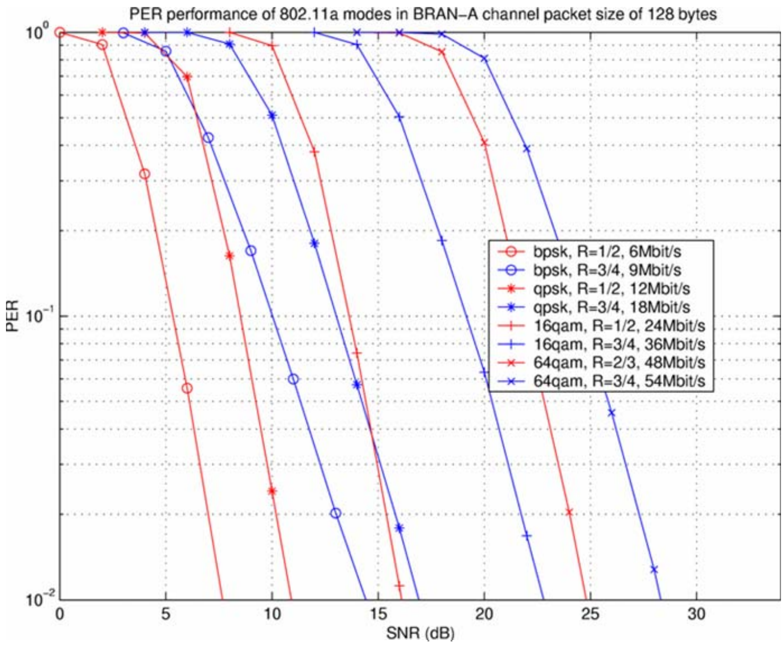


Figure 6-7 - PER performance of the 802.11a modes with 128 byte packets in a typical office environment

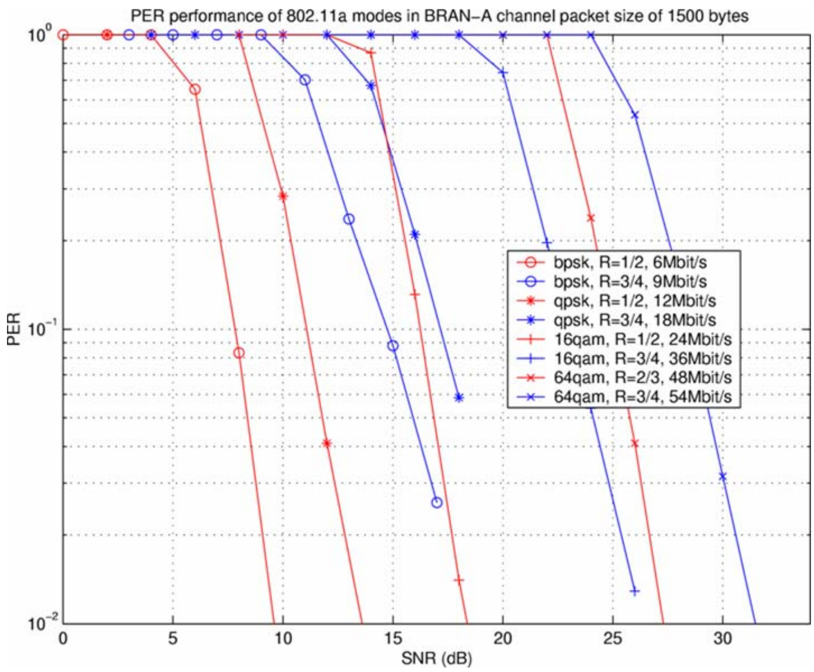


Figure 6-8 - PER performance of the 802.11a modes with 1500 byte packets in a typical office environment

6.2.2.4.2 Cell deployment and configuration

The terminals are dropped randomly within an isolated squared cell of 40 meters side (i.e. the largest distance between the AP in the centre and the STA is equal to 28 meters). This enables to offer a variety of path loss.

6.2.2.4.3 Channel Behaviour and control parameters

The environment is a typical one floor office with NLOS propagation. This is characterized by a path loss exponent of 3.6 (in the path loss model described in equation 7.1 of the "calibration" document [19]) and a shadowing standard deviation of 5.0 (the shadowing is assumed to have a log-normal distribution). Besides, as specified for the figures above, the channel model used is of type "BRAN-A" (typical indoor for short range). The channel bandwidth is fixed at 20MHz.

6.2.2.4.4 User Traffic Characterization

Two types of traffic are considered: VoIP and file transfer.

For the VoIP traffic, two vocoder rates are considered: R1=8 kbit/s and R2=64 kbit/s. Voice Packets inter-arrival time is Tvoice=20 ms, and the maximum delay tolerated for voice packet delivery is also Tvoice=20 ms. The sizes of the RTP, UDP and IP headers are respectively 12, 8 and 20 bytes, i.e. a total of 40 bytes. Since the 802.11 MAC header is 28 byte long, the MAC PDU length is 88 bytes for the 8 kbit/s vocoder and 228 bytes for the 64 kbit/s vocoder. There is no silence suppression. The delivery requirement is real-time (delay less than 20ms)

Depending on the simulations exposed below, two types of file transfer have been considered: FTP TCP/IP streams at 340 kbit/s, sent with 1500 bytes long frames with inter-request time distributed uniformly between 0 and 2 seconds. This is the case when FTP is used as a background traffic for voice. If FTP users are considered alone, they send data at full load during 60s.

In all cases, only UL connections are considered.

6.2.2.4.5 RRM mechanisms (On/Off)

Some mechanisms are used in the simulations. In what follows, a link adaptation mechanism is implemented to show optimal or bad performance of the system depending on this algorithm being switched on or off.

More information about the modes used to transmit the packets follow:

- ACKs are sent with the most robust PHY mode (BPSK 1/2);
- Broadcast packets (e.g. ARP) are sent in BPSK ½;
- In the current version of the implementation, we consider that the AP STAs have the same LA algorithms than the non AP STAs. They will react exactly in the same way.

6.2.2.4.6 Other characteristics

The RTS/CTS threshold is set such that packets are not fragmented (no RTS/CTS because the RTSThreshold is set to 3000 bytes).

EDCA is used as the channel access technology. It is compliant with the latest 802.11e draft. When a mix traffic is used, the voice and data traffic are assigned different EDCA parameters (AIFS=DIFS, CWmin,voice=3, CWmax,voice=7, CWmin,background_data,=15, CWmin,background_data=1024). This prioritizes the VoIP traffic and leaves the background DATA traffic (together with ARP packets, beacon and TCP acknowledgements) with the legacy 802.11a characteristics.

The ACKTimeOut is set to a PIFS and the Retry Limit for all packets is set to 7. In other words, when the transmitting STA does not receive any ACK a PIFS after the end of the DATA packets transmission, it considers that the transmission has failed and it retransmits the same DATA packet up to the given Retry Limit (i.e. 7).

Address Resolution Protocol (ARP) is used (this protocol must be successfully completed before any traffic can actually start).

6.2.2.5 Alignment with D7.2 configurations/parameters

With respect to what has been specified in the D7.2 [18] and the System Simulator Calibration [19] documents, there has been an attempt to have benchmarking results that can be compared to others and to what future simulations in WINNER will provide.

If we now refer to what has been stated in section 3 of D7.2 [18], we can compare the hypotheses made for results presented in 6.2.1.4 with those in D7.2.

First (table 3.1 of D7.2), the considered scenario is Scenario A.1, with a mobility of 2m/s (i.e. 7.2km/h, slightly above the indicated range of 0-5km/h – by doing that the environment is moving faster and the performances are more challenged). Besides, the traffic density is high: the scenarios with VoIP users are tested at and above the limit of supported traffic, the scenarios with FTP users are tested at full load.

Second (table 3.2 and section 5 of D7.2), let the environment characteristics be compared:

- The indoor model is used: the path loss model does not take into account floors and walls but only distance (see 3.3.1 of D7.2) with a path loss exponent fixed at 3.6 (typical indoor office). Referring to the path-loss model exposed in equation 7.2 in the calibration document [19], it becomes in this benchmarking results with a carrier frequency of 5.25GHz:

$$PL(d) = 36\log_{10}(d) + 42.03\left[dB\right]$$

- The mobility model is rather basic here compared to what is proposed in D7.2: users are static and the environment is moving around them at the specified speed (7.2km/h) – this enables fluctuations of the fast fading.
- The user related characteristics (table 3.2 and section 5 of D7.2 [18], table 1 of the calibration document [19]) can also be compared:
- Devices are of type 1 with one antenna, and maximum transmitting power of 200mW (they actually limit their transmit power at 40mW);
- Users are uniformly distributed in the cell – actually, in order to average the results over the cell, from 150 to 1500 positions in the cell are randomly selected to provide a uniform distribution and to average the results;
- All users are engaged in an uplink connection;
- There is no mobile relay station.

At last, as specified in table 6 of the calibration document, adaptive modulation and coding 9AMC or link adaptation LA) is used with the coding schemes that we define in Table 6-7 above.

The “figure of merits” used are as follows:

- **Coverage:** throughput vs. distance curves are provided to give an idea of what can be optimally reached (with an optimal LA algorithm and no interference) at a given distance;
- **Capacity:** the aggregated cell throughput is provided for users performing full load file transfers;
- **Delay and QoS:** these are critical FoM, especially for VoIP applications. Concerning those, a limit of T_{voice}=*20ms* is imposed for packets before their being dropped. The VoIP connection is considered “good” as long as the percentage of packets delivered within T_{voice} is higher than 99%. For FTP users, their satisfaction is measured in terms of percentage of dropped packets.

6.2.2.6 Results

Throughput versus distance performance is plotted on Figure 6-9 and Figure 6-10. The throughput on top of the physical layer is also plotted as a reference. It can be observed that although the PHY layer throughput is lower with long packet sizes, due to lower PER performance, the main effect on the MAC throughput is the overhead, which is much lower with long packets. The use of short packets (e.g. those used for VoIP) is very inefficient and leads to a very low throughput. The assumption used in the simulations is that the channel is slowly varying, so that the advised mode remains valid even until the next transmission.

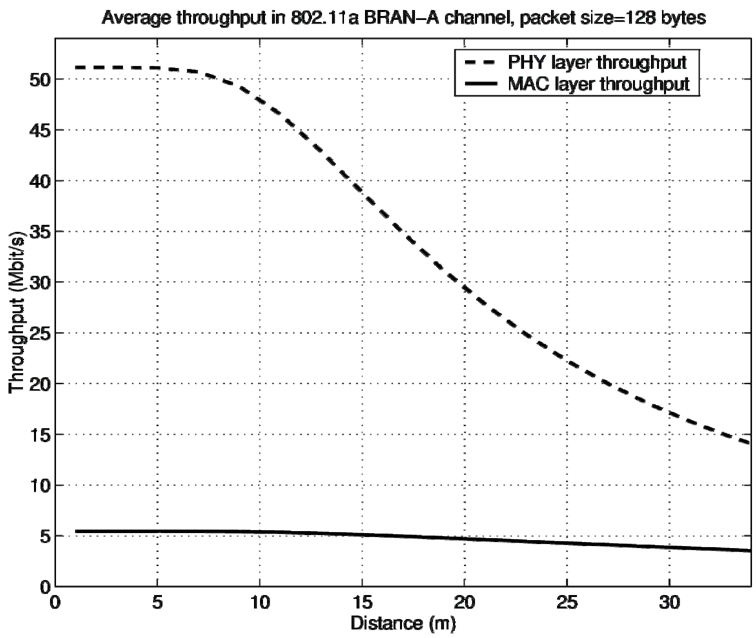


Figure 6-9 - Performance of 802.11a: average throughput versus distance, 128 byte packets (TxPower=200mW)

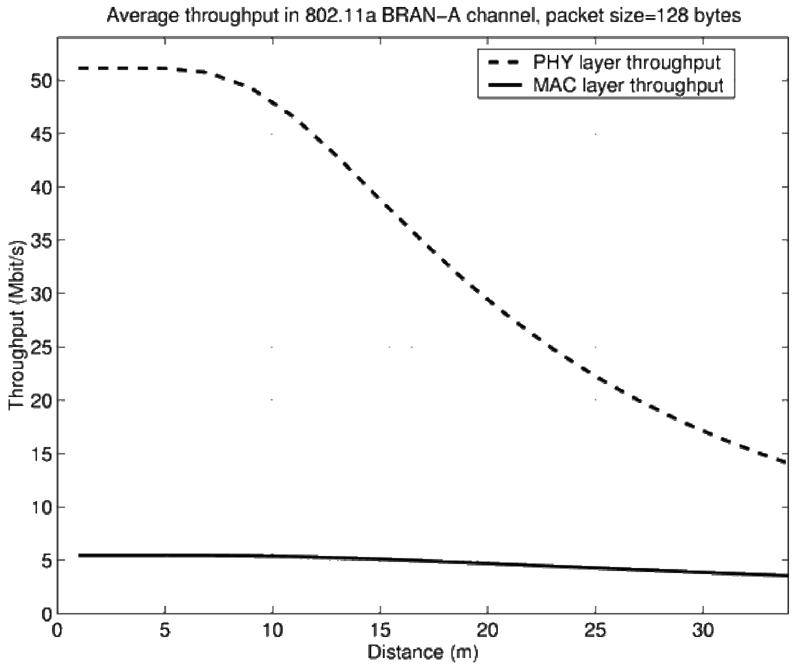


Figure 6-10 - Performance of 802.11a: average throughput versus distance, 1500 byte packets (TxPower=200mW)

In order to better understand curves provided on Figure 6-9 and Figure 6-10, we plotted the throughput versus distance performance of the various transmission modes of 802.11a. On Figure 6-11 and on Figure 6-12, the average MAC throughput of a single user is plotted versus the distance for each PHY bit rate, for propagation exponent n=3.6. It can be observed that for cells of 30m radius, link adaptation algorithms would rather directly work with the 12Mbit/s mode. However, the 6Mbit/s mode allows an improved range, especially in more severe propagation environments. The advantage of using 6Mbit/s mode disappears when the packet size increases.

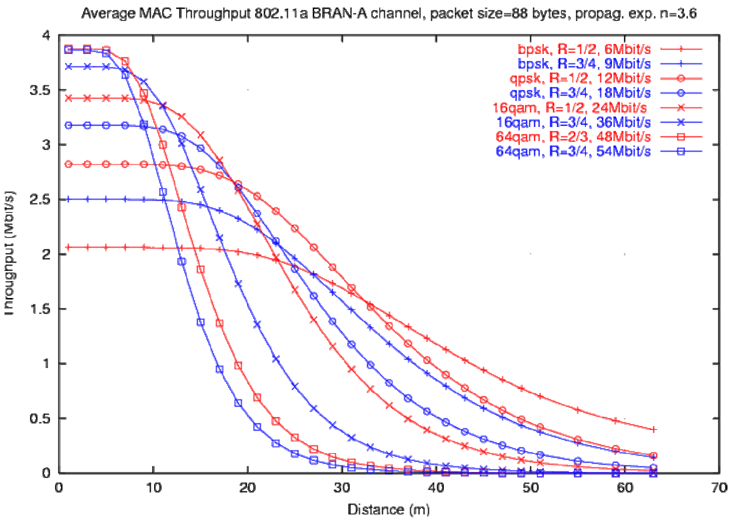


Figure 6-11 - Throughput vs. distance for VoIP (88 byte MAC PDU) traffic, with propagation exponent n=3.6 and shadowing σ_S=5 dB (TxPower=200mW)

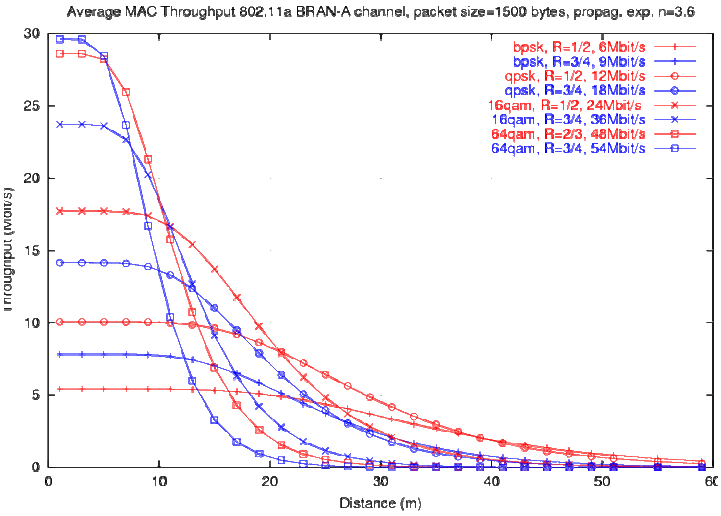


Figure 6-12 - Throughput vs. distance for data (1500 byte MAC PDU) traffic, with propagation exponent n=3.6 and shadowing σ_S =5 dB (TxPower=200mW)

Now, if the results (obtained with Matlab) presented above are compared with results obtained with a system simulator, we obtain very similar results (Figure 6-13). The black and blue curves represent the throughput above the MAC layer for a STA sending data to the AP at a fix rate of respectively 6 and 24 Mbps. The difference with results on Figure 6-12 is the transmitted power (40mW instead of 200mW): but these curves behave in the same way. Indeed, for a distance greater than 25m, the throughput provided at 24Mbps falls below the one at 6Mbps. When Link Adaptation is used (with or without a margin), the throughput is comparable to what is obtained on Figure 6-10 (with lower throughput because of the lower transmitted power).

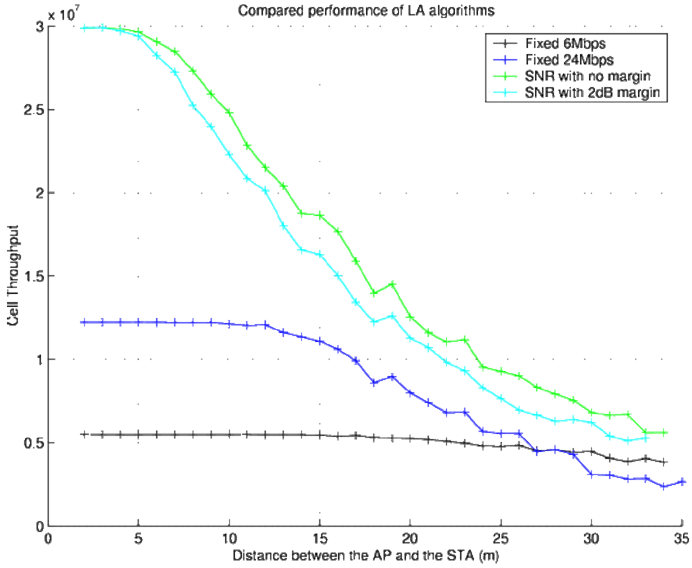


Figure 6-13 – Throughput vs. distance for data (1500 byte MAC PDU) traffic, with propagation exponent n=3.6 and shadowing σ_S =5 dB (TxPower=40mW)

Now, if we have no more a single connection, but multiple connections transmitting data at full load, the aggregated cell throughput does not exceed 12Mbps above the MAC layer. The cell is quite large and when the number of users increase, the collision rate increase and the throughput decreases proportionally.

Besides, if we look at VoIP connections, instead of FTP applications, we notice on Figure 6-14 that the maximum number of 64kbps VoIP users supported in the cell is around 40 if an optimum link adaptation algorithm is used. As explained in the previous subsection, a user is satisfied as long as 99% of its packets are delivered within 20ms.

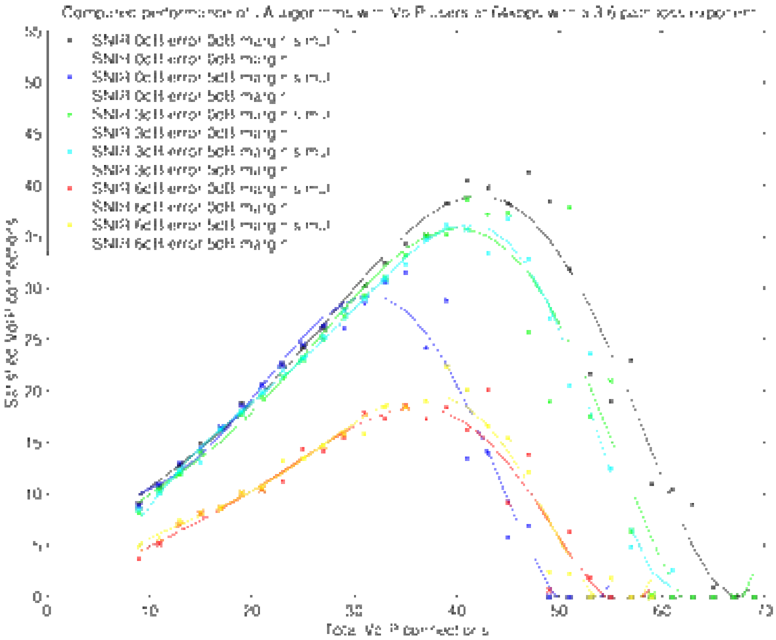


Figure 6-14 – Number of satisfied VoIP users in a cell with different link adaptation algorithms

6.2.3 WLAN HIPERLAN/2

The Medium Access Control (MAC) Layer of High Performance Radio Local Area Network Type 2 (HIPERLAN/2) is used as an example for a MAC that could be the base for the development of the next generation of mobile telecommunication systems that are expected to be wireless broadband systems. H/2 was standardized by the European Telecommunications Standards Institute (ETSI) within the project Broadband Radio Access Networks (BRAN), see [26]. HIPERLAN/2 (H/2) is member of a family of related systems, which are: HIPERLAN/2, HIPERACCESS [30] and HIPERLINK. The main objectives for the specification of H/2 were the support of QoS, Handover (HO) and data integrity, which is not present in standards of Institute of Electrical and Electronics Engineers (IEEE) 802.11. For a further description of the other systems refer to [31].

The H/2 system is a variant for short range (up to 200m) communication complementing UMTS systems and being a Wireless Local Area Network (WLAN) for private users. It offers high speed access (up to 54 Mbit/s gross data rate) to various networks including Universal Mobile Telecommunications System (UMTS) core networks, Asynchronous Transfer Mode (ATM) and Internet Protocol (IP) based networks. A typical HIPERLAN/ 2 network consists of one or several Access Points (APs) that are covering an area partly or totally. APs are sometimes also called Central Controller s (CCs). The Mobile Terminals (MTs) (or Wireless Terminal (WT)) are communicating with the AP. While moving in the geographical area covered, they are automatically attached to the best serving AP.

The APs are selecting a suitable frequency channel, which is of 20MHz bandwidth, from the list given in Table 6-8. In Europe there are 455 MHz assigned in the License-Exempt- Spectrum at 5 GHz, while in the United States of America (USA) there is 300MHz unlicensed spectrum assigned and only 100 MHz in Japan. In Europe the allowed Equivalent Isotropic Radiated Power (EIRP) regarding the maximum mean transmission power is 200mW in the lower and 1W in the higher frequency band. Furthermore it is stipulated to use Power Control (PC) and Dynamic Frequency Selection (DFS). This is not necessary for stations located in the USA, which are sharing the spectrum with IEEE 802.11a. In Japan all stations have to perform Carrier Sensing (CS) every 4ms.

Table 6-8 Frequency bands designated for HIPERLAN/2

Region	Frequency Channels	Restrictions		Total Number
		Power(EIRP)	Technology	of Channels
Europe	$5180\,\text{MHz} + N \cdot 20\,\text{MHz}, 0 \leq N \leq 7$	200 mW	DFS, PC	19
	$5500\,\text{MHz} + N \cdot 20\,\text{MHz}, 0 \leq N \leq 11$	1 W	DFS, PC	
USA	$5180\,\text{MHz} + N \cdot 20\,\text{MHz}, 0 \leq N \leq 7$	200 mW/1 W	-	12
	$5745\,\text{MHz} + N \cdot 20\,\text{MHz}, 0 \leq N \leq 3$	4 W (outdoor)	-	
Japan	$5180\,\text{MHz} + N \cdot 20\,\text{MHz}, 0 \leq N \leq 3$	-	CS (4 ms)	4

6.2.3.1 System Parameters

The specifications of the ETSI/BRAN systems are defining predominantly the lower three layers of the International Organization for Standardization (ISO) Open Systems Interconnection (OSI) reference model, standardized in [32]. A schematic sketch of the service model is depicted in Figure 6-15. On top of the physical layer (parameters in Table 6-9) is the Data Link Control (DLC) layer, with its sub-layers MAC, Radio Link Control (RLC) and Error Control (EC). The Convergence Layer (CL) adapts the HIPERLAN/2 protocol stack to different higher layer protocols. The protocol functions are is divided into two planes, namely the control and the user plane. All layers above the DLC are beyond the scope of WINNER and therefore not explained in detail. Among the functions usually residing in the RLC layer, only the Link Adaptation (LA) was regarded in the benchmarking, because it is an essential functionality required for the operation in a dynamic, event-driven simulation.

Table 6-9: HIPERLAN/2 OFDM Transmission Parameters

Parameter	Value
Sample frequency $F_s = 1/T$	20 MHz
Symbol duration T_u	$64 \cdot T = 3.2\,\mu$s
Guard time T_{GI} (demanded/optional)	$16 \cdot T = 0.8\,\mu$s / $8 \cdot T = 0.4\,\mu$s
Symbol interval T_S (demanded/optional)	$80 \cdot T = 4.0\,\mu$s / $72 \cdot T = 3.6\,\mu$s
Sub-carrier for data N_{SD}	48
Sub-carrier for pilot signals N_{SP}	4
Total number of sub carrier N_{ST}	52 ($N_{SD} + N_{SP}$)
Sub-carrier distance δ_f	0.3125 MHz ($1/T_U$)
Distance of outer sub-carrier	16.25 MHz ($N_{ST} \cdot \delta_f$)

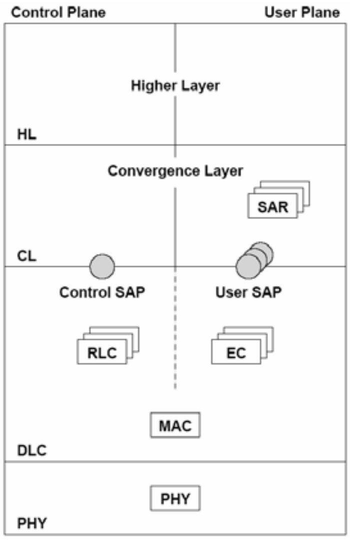


Figure 6-15: HIPERLAN/2 Service Model

6.2.3.2 Link Adaptation

In order to adapt the transmission to the channel quality, given by a Carrier to Interference Ratio (CIR) or Bit Error Rate (BER) the modulation scheme and the coding rate can be changed dynamically. A combination of modulation scheme and coding rate is called a PHY-mode, all possible PHY-modes are listed in Table 6-10 and their respective link level performance is shown in Figure 6-16.

Table 6-10: Modulation and Coding Schemes for HIPERLAN/2

Modulation	Coderate	Coded Bits per Symbol	Data Bytes per Symbol	Nominal Bit Rate MBit/s
BPSK	1/2	48	3	6
BPSK	3/4	48	4.5	9
QPSK	1/2	96	6	12
QPSK	3/4	96	9	18
16QAM	9/16	192	13.5	27
16QAM	3/4	192	18	36
		optional		
64QAM	3/4	288	27	54

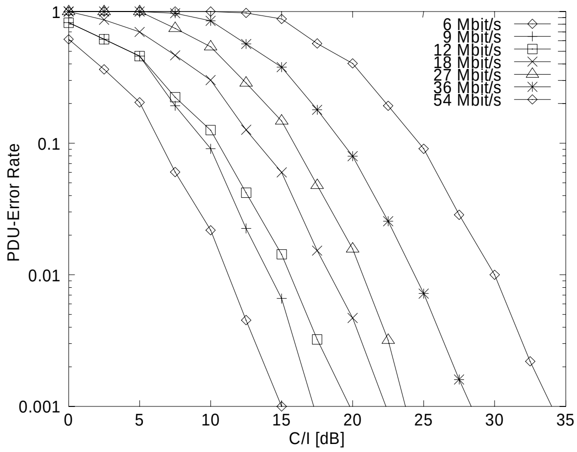


Figure 6-16: PDU-Error Probability for varying C/(I + N) and PHY-mode [22]

6.2.3.3 Simulation Overview

6.2.3.3.1 Scenario: Dense Urban Hot Area Coverage

In order to ensure comparability of results, the Manhattan grid scenario has been taken for the following investigations, see [23]. The most important parameters of the scenario are the block size of 200 m and the street width of 30 m. In the deployment scenarios in Figure 6-17:, each of the APs covers the range of two building blocks and one street crossing, resulting 430 m range.

This cell configuration requires a minimum of 4 carrier frequencies to ensure that in each direction, the co-channel cells are separated by at least one cell with another carrier frequency, see Figure 6-17: (left). Based on that structure, two possible variants can be considered. The APs can be placed at equal coordinates in adjacent streets, shadowed by the buildings (Figure 6-17:, middle). The second variant is that APs are placed on street crossings (Figure 6-17:, right), thereby covering horizontal and vertical streets. In this scenario, at least 8 frequencies are needed to ensure that co-channel cells are separated by cells using a different frequency.

6.2.3.3.2 Scenario: Wide Area Coverage

It is known that wireless broadband systems exhibit low coverage range at high bitrates [21]. In a conventional 1-hop hexagonal cellular approach, this leads to a large number of APs required for continuous coverage. We consider a coverage radius for a single AP of *R*=200 m and R=346 m with different cluster sizes (*N*=3,7,12).

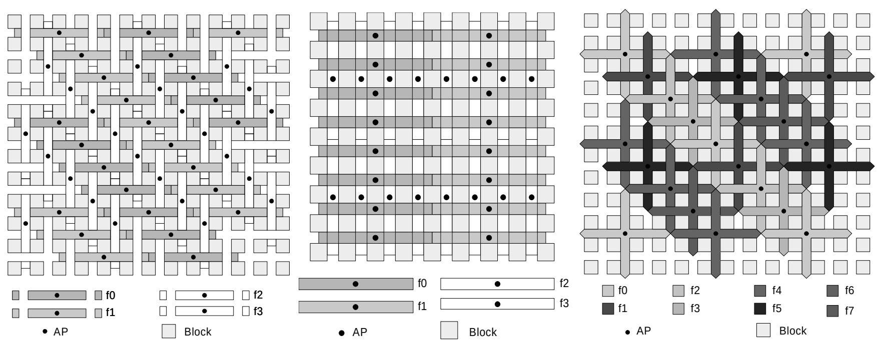


Figure 6-17: Possible H/2 AP deployments

Link-Level Performance: The basis for the calculation of the transmission errors is the ratio of Carrier to Interference and Noise power (C/(I+N)). The results of Link-level investigations [22] provide a Protocol Data Unit (PDU) error-probability related to the average C/(I+N) during reception of the PHY-PDU. This relation is shown in Figure 6-16. In our simulation model, collisions of interfering transmissions are detected and the resulting average C/(I+N) is calculated for each transmitted PHY-PDU to decide on success or retransmission.

Propagation Models: The COST259 Multi-Wall model has been used in the Manhattan scenario. This model [24] is an indoor propagation model at 5GHz, which takes into account the transmission through walls obstructing the LOS between transmitter and receiver. Unlike in the COST231 model [25], the attenuation non-linearly increases with the number of transmitted walls. Wall attenuations have been chosen according to the suggestions from the BRAIN project [25].

The Propagation Model used in the wide-area simulations is the Large-Open-Space model [23] and a path loss exponent of γ=2,5 has been used.

Other Parameters of the Simulation Model: Transmission Power has been set to a fixed value of 20dBm (100mW) for all simulations, no Power Control has been performed. The traffic load is assumed to be constant bitrate, which is a reasonable assumption when investigating the maximum achievable end-to-end throughput.

6.2.3.4 Results

This section presents the performance evaluation results obtained by stochastic-event driven simulation. Results for the Downlink (DL) direction are presented here only, since the main effects that can be observed are quite similar in Uplink (UL) and DL directions, a result which is partly due to the Time Division Duplex air interface studied.

6.2.3.4.1 Reference Scenarios

In Figure 6-21:-Figure 6-21 (left) the DL C/(I+N) and the related maximum End-to-End throughput are plotted over the distance of the MT from the AP when servicing the scenario by APs only, according to the Manhattan and the Wide-Area scenarios.

In the Manhattan Scenario, the C/(I+N) values are slightly higher for the deployment variant where the APs are placed on the street crossings (Figure 6-17:, right) when compared to the other options. Figure 6-21: (upper right) shows the resulting Throughput (TP), again versus the distance of the MT from the AP. At distances of 115 m and 345 m, some additional interference on the crossings is visible for a deployment according to Figure 6-17: (right)

In the wide-area cellular deployment, the C/(I+N) values degrade as expected with decreasing cluster size. For comparison, also the C/(I+N) for a single AP without Interference is shown. Figure 6-21: (lower right) shows that at the cell border (at a distance of 200 m), a maximum End-to-End throughput of ca. 8 Mbit/s can be provided in the very optimistic case of N=19.

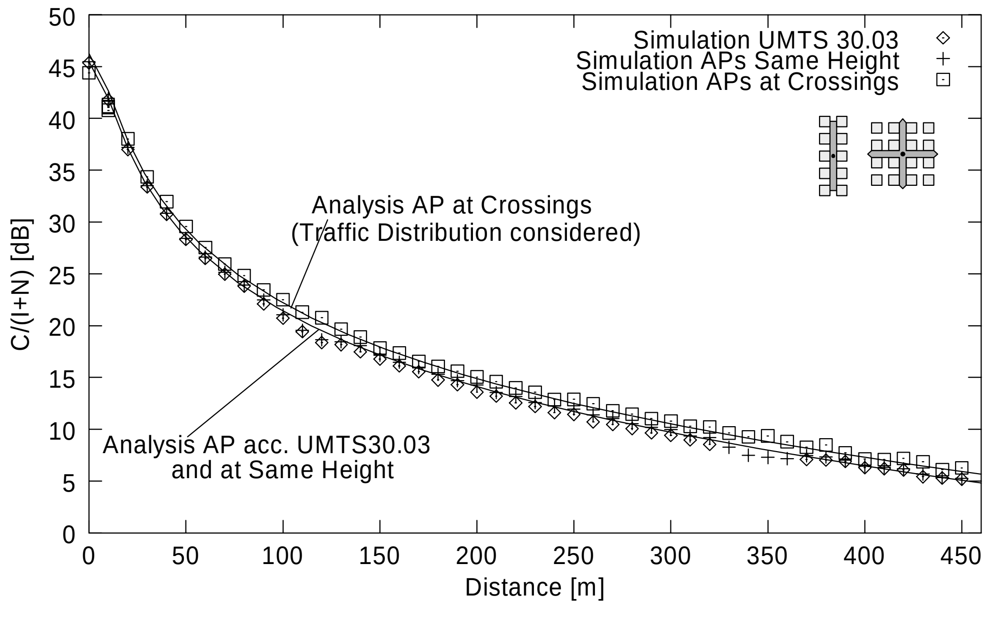


Figure 6-18: C/(I+N) in Manhattan Scenario (Lines: analysis, Markers: simulation)

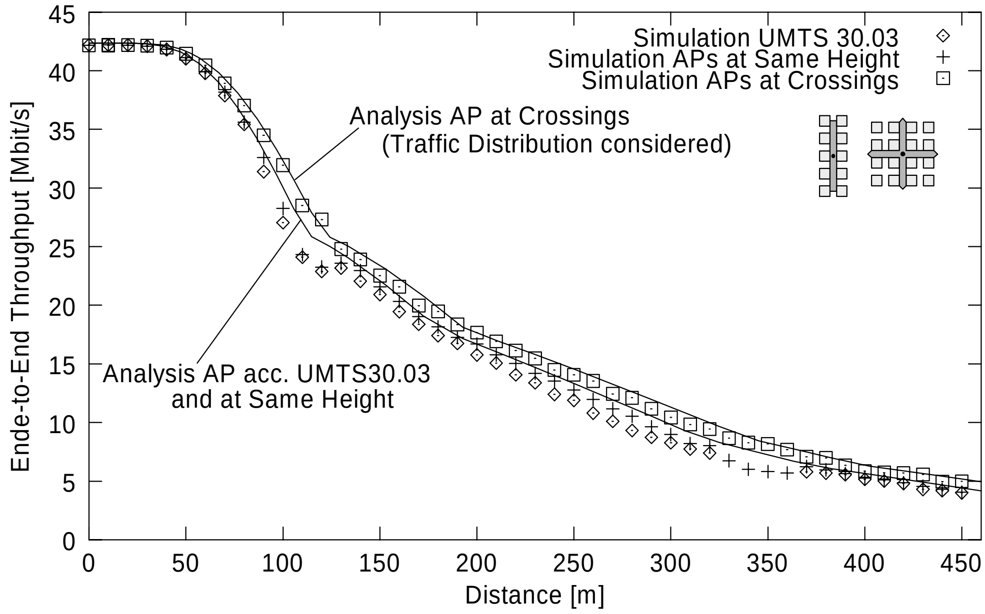


Figure 6-19: Top: C/(I+N) (left) and End-to-End-Throughput (right) in Manhattan Scenario (Lines: analysis, Markers: simulation)
Bottom: C/(I+N) (left) and End-to-End-Throughput (right) in wide-area scenario (Lines: analysis, Markers: simulation)

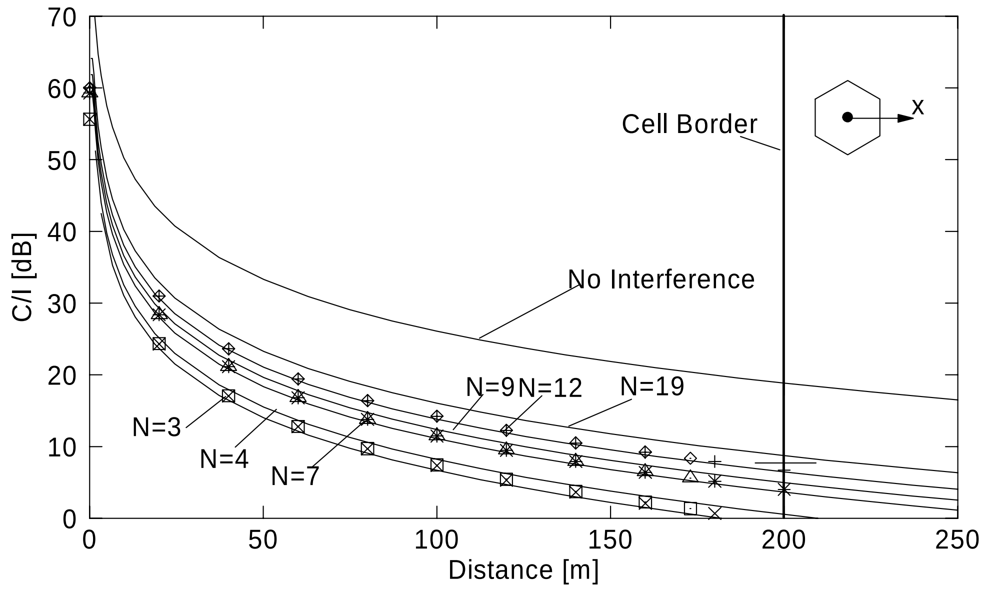


Figure 6-20: C/(I+N) in wide-area scenario (Lines: analysis, Markers: simulation)

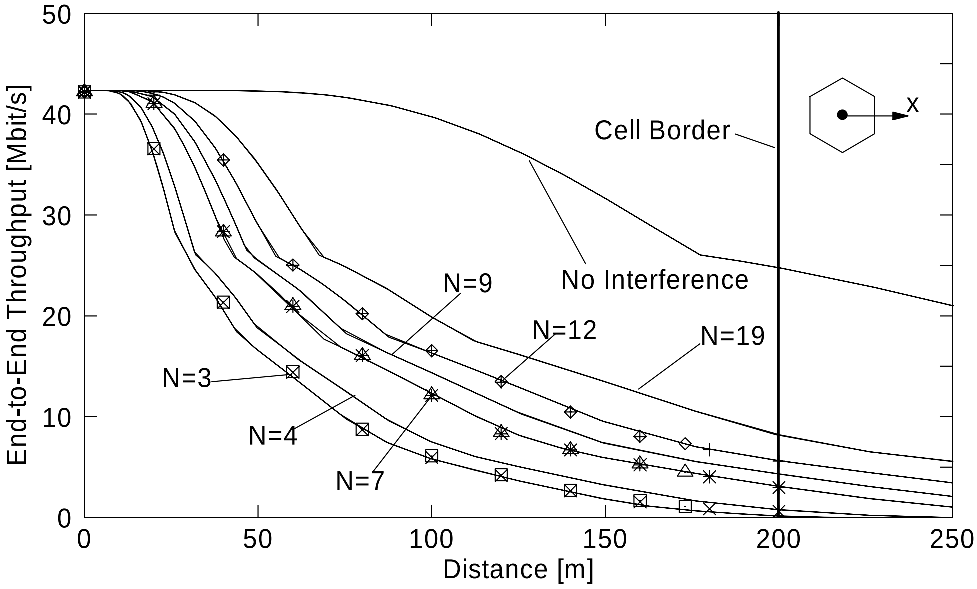


Figure 6-21: End-to-End-Throughput in wide-area scenario (Lines: analysis, Markers: simulation)

Table 6-11: Average Cell Capacity and spectral efficiency for a Cell with 10 MTs and Exhaustive Round Robin (ERR) Scheduling, comparing three Manhattan Single-Hop deployments

Scenario	Used # of Freq.	Cell Size [m2] / 103	Cell Capacity [Mbit/s]	Spect. Efficiency [bit·s^{-1}·Hz^{-1}·m^{-2}]
1-Hop (UMTS 30.03)	4	25,8	21,04	10,19
1-Hop (horiz./vert. depl.)	4	25,8	20,01	9,69
1-Hop (APs on cross.)	8	53,4	20,24	2,37

6.2.3.4.2 System Capacity and Spectral Efficiency

In addition to the End-to-End throughput studied in the previous sections, the system capacity, i.e. the aggregate traffic that can be carried in a well-defined service area and a certain amount of used spectrum is an important measure to assess a system's performance. To optimise a system, it is very important to have a clearly defined optimisation goal. Table 6-11: shows the average End-to-End cell throughput for the different 1-hop deployments in the Manhattan scenario. The AP deployment from Figure 6-17: (left) shows only small advantages over the horizontal/vertical placement (Figure 6-17:, middle). The placement on street crossings (Figure 6-17:, right) has the advantage that a larger area is covered per AP, reducing the number of needed backbone connections by a factor of 2. At the same time, a minimum of 8 carrier frequencies is needed to enable continuous coverage. This and the larger cell size lead to a substantial reduction in spectral efficiency, while the average cell throughput changes only slightly. Table 6-12: shows the average End-to-End cell throughput for the different 1-hop deployments in the wide-area scenario. Again, from the small cell size and the high cell throughput results a relatively high area spectral efficiency in the case of the 200 m-cells.

Table 6-12: Average Cell Capacity and spectral efficiency for a Cell with 10 MTs and Exhaustive Round Robin (ERR) Scheduling, comparing the wide-area cellular Single-Hop deployments

Scenario	Used # of Freq.	Cell Size [m2] / 103	Cell Capacity [Mbit/s]	Spect. Efficiency [bit·s^{-1}·Hz^{-1}·m^{-2}]
Standard 200m	3	104	6,84	1,10
Standard 200m	7	104	12,2	0,84
Standard 200m	12	104	16,42	0,66
Standard 346m	3	311	6,53	0,35
Standard 346m	7	311	11,42	0,26
Standard 346m	12	311	14,82	0,20

6.2.3.4.3 Delay Aspects

Figure 6-22 shows the end-to-end delay for high-rate connections with 1536 Mbit/s and for different channel error probabilities *p.* A constant use of the 54 Mbit/s PHY-mode is assumed. It clearly shows that the delay increases with increasing error probability, because of the increasing number of retransmissions necessary. It also shows that, if the channel quality is known, the system allows certain maximum delay gurantees. In Figure 6-23 the end-to-end delay for the 99% percentile of PDUs on 1Mbit/s connections is plotted versus the C/I for the different HIPERLAN/2 PHY modes. It shows that a 5ms maximum delay can be guaranteed for all users starting at a C/I of 10dB, if the lowest (6Mbit/s) PHY mode is used, and starting at 27dB for the highest (54Mbit/s) PHY mode.

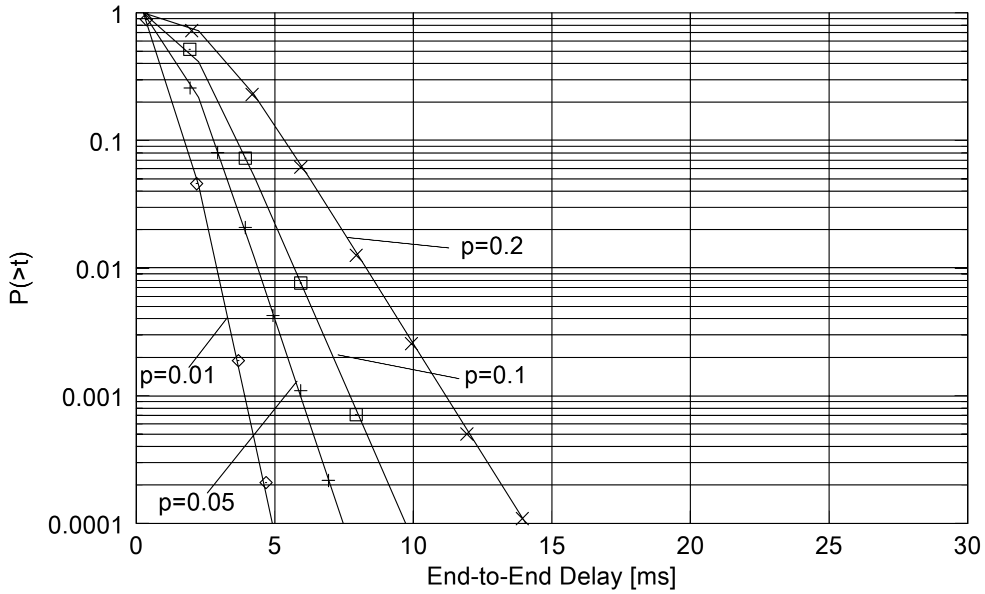


Figure 6-22: End-to-End-Delay for Connections with 1536 kbit/s for different channel error probabilities (Lines: analysis, Markers: simulation)

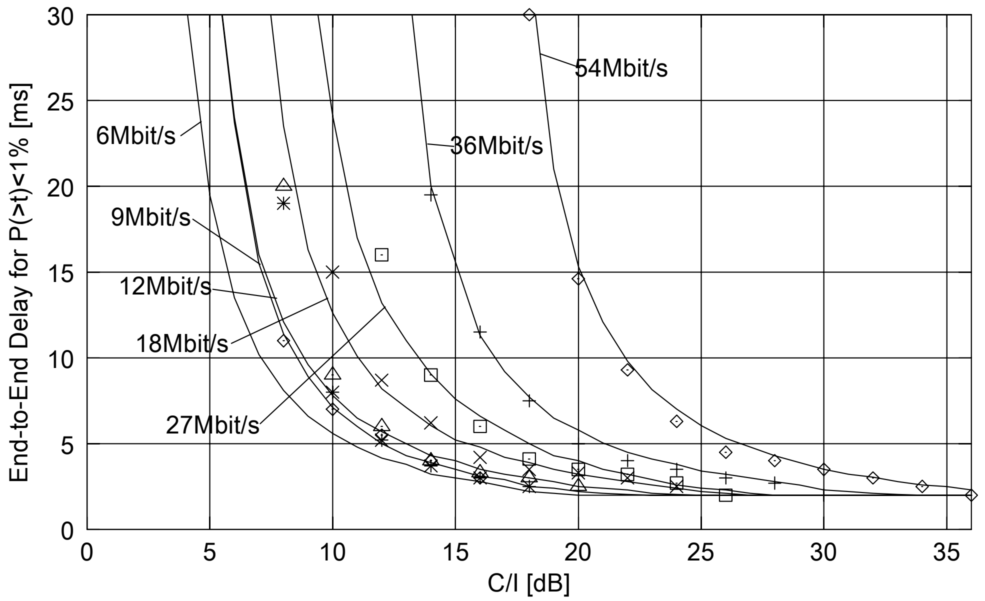


Figure 6-23: End-to-End-Delay for 99% of PDUs on 1Mbit/s-Connections plotted vs. C/I (Lines: analysis, Markers: simulation)

6.2.4 WMAN 802.16(a)

6.2.4.1 General

The IEEE802.16a specifies a BFWA system able to operate in the 2-11 GHz frequency range. The standard is born as an amendment to IEEE802.16-2001 which specified a BFWA system operating at frequencies between 10 and 66 GHz. By June the 24th 2004, a working group in charge of the revision of all the specification, produced the last version of the standard that consolidates the points left open by previous releases and unify in one document the two mentioned above (IEEE802.16-2004), [35]. Actually then, the IEEE802.16a wording can be misleading considering the evolution undertaken by specs but it will be still used in the following to indicate the low frequency option of the IEEE802.16 system.

Main goals addressed by IEEE802.16a standard, then, can be resumed as: non-professional installation of terminal, rapidly scalable infrastructure deployment, efficient spectrum usage, QoS support. In the lower range of working frequency, NLOS operation can be guaranteed. Users targeted by the system span from large business, through SME and SOHO down to residential.

The IEEE802.16 standard employs several enhanced features at the PHY and at the MAC layer gives the possibility to deploy both the conventional PMP architecture and the innovative Mesh (Multi-hop) architecture. A basic characteristic of the IEEE802.16a is that the support for mobility is not foreseen. Furthermore, also evolutionary steps of the standard (e.g. carried out by working group IEEE802.16e, [36]) which will include mobility, consider it like an additional opportunity for the standard and not a driver (i.e. no mobility at speeds of 250 km/h for real time applications will be provided). This characteristic allows maintaining greater network simplicity also in perspective in respect to 2G and 3G networks, which translates directly into lower costs for the network deployment.

The mixture of advanced features and flexibility provided by the IEEE802.16a makes this system really interesting considering the development of new communication systems. Currently there is a great attention about the IEEE802.16a definition and real performance statements. The WiMAX forum, [37], has been founded in order to grant demonstrable compatibility and interoperability among systems and components developed by OEMs. By definition of conformance and interoperability tests, WiMAX will be able to issue "WiMAX Certified™" label.

Even if great benefits are attended from IEEE802.16a there are still things to be clarified. In particular, some doubts are related with the position assumed in respect to other existing standards, and some open points are not clear regarding transfer speeds and actual coverage range. Furthermore no currently certified equipment is commercially available. Compared to other systems considered in another paragraph less data or performance assessments are available.

6.2.4.2 Basic remarks on network deployment

Some remarks are hereafter proposed, as an overview of 802.16 network deployment aspects. The purpose of these remarks is just to outline the basic feature of this kind of systems, to allow a quick comparison with the other systems considered in this chapter.

The term WiMax will be used here for the sake of simplicity to indicate 802.16-2004 systems. Some remarks on the evolution to 802.16e will be reported in the following sub-section 2.2.4.7.

- WiMax is basically a broadband wireless access system. First deployments are expected to take place for WDSL (i.e., "fixed" applications), which is related to the interest of FNO (Fixed Network Operators) and ISP (Internet Service Providers) / WISP (Wireless Internet Service Providers).
- WiMax will be a high-performance BWA systems - since the first versions - in terms of:
  - channel bandwidth (scalable solutions will be offered up to 20 MHz bandwidth)
  - overall throughput per cell (the order of magnitude of 63 Mbit/s per cell is theoretically achievable given a bandwidth of 20 MHz) - ref. to the section 2.4.4 for an analytical evaluation of the overall capacity considering also protocol-related constraints; in the same section, further details about the parameters affecting the overall capacity value will be provided.
  - overall throughput handled by a single WiMax Base Station (BS), which can manage more cells (a typical case is a multi-sector capable BS, which serves three or four sectors)
  - large numbers of served Secondary Stations (SS) per BS - even up to hundreds. It should be noted that in a first step the SS will be stand-alone (although small) equipment. A fixed interconnection is initially foreseen between the SS and the user

PC's. The connection of more PC's per terminal station is in principle possible (depending obviously on specific characteristics of the products available in the future). In the first deployment scenarios, the PC is the target terminal, because of the data-oriented nature of WiMax applications.

- The NLOS property of WiMax will ensure both indoor and outdoor coverage, and therefore a cellular deployment, up to a full coverage: rural, suburban, urban.
- A good indoor coverage is very important, given the kind of foreseen applications.
- The actual coverage characteristics (radio planning) for WiMax networks are currently under study. It is interesting to note that - from a qualitative point of view - the coverage characteristics are expected to be in principle not remarkably different with respect to the ones of 2G (and probably 3G). This aspect suggests interesting perspectives of co-location of WiMAX BS and 2G BTS / 3G Nodes B, and therefore also MNO (Mobile Network Operators) may be interested in this solution to offer complementary services with respect to their own.
- The access network architecture of WiMax systems is flat and simple. No Base Station Controller is foreseen; in the more probable network model, the BS is directly connected to Access Routers.
- The core network is in principle built simply as an IP core network, supported by an IP backbone. Proper functional servers (for authentication, IP address assignation, etc.) are necessary, connected to the IP backbone.
- It is therefore possible to integrate WIMAX access, in a more or less strict way, in existing core networks of ISP/FNO.
- This is also possible in principle with respect to MNO core networks; the degree of integration - from the feasibility and convenience viewpoints - is under investigation.
- As already mentioned, in the first deployment, WiMax will be a data-oriented access network. A comparison is therefore meaningful with HSDPA; in qualitative terms, WiMax will offer a higher overall throughput than HSDPA, more flexibility in bandwidth allocation in UL and DL differently from HSDPA, but obviously will not support full mobility.

It is finally noted that the technology evolution will lead to significant enhancements and is open from the Network and Application viewpoints (3.5).

6.2.4.3 Medium Access Control layer

A basic characteristic of the IEEE802.16 is represented by a MAC layer which is mostly common for all frequency range foreseen for operation. The capability to work in different frequency ranges (i.e. between 10 and 60 GHz or between 2 and 11 GHz) or to operate in different environments (either LOS or NLOS) are substantially related with the different radio interfaces defined inside the standard. The capabilities of the MAC layer allow IEEE802.16 to support a wide array of data networking services, addressing users that span from corporate to residential.

The MAC layer actually manages the fixed wireless access deployment, providing ways for BS to be connected to public networks, for SSs to get service, and for users to be served with multiple services with different QoS requirements and with solid protection for security and privacy. The MAC layer is connection oriented, meaning that all data communications are managed in context of connections and all procedures are carried out referring to connection addresses.

The MAC layer is composed of three sub layers:

- The MAC Service Specific Convergence sub layer. It interfaces higher layers (ATM Convergence Sub layer, Packet Convergence Sub layer are defined). Functionalities provided by the sublayers are related with accepting higher layer protocol data units (PDUs), their classification, processing and delivering to appropriate MAC SAP and respective connection.
- The MAC Common Part Sublayer. It carries the key functionalities for sharing the radio resources. The MAC CPS manages network entry for Secondary Stations and management of connections over which data are transmitted to and from the Base Station (BS). It will be described in detail below.
- The MAC Privacy Sublayer which provides authentication, secure key exchange, and encryption.

In the figure below the MAC layer structure is evidenced.

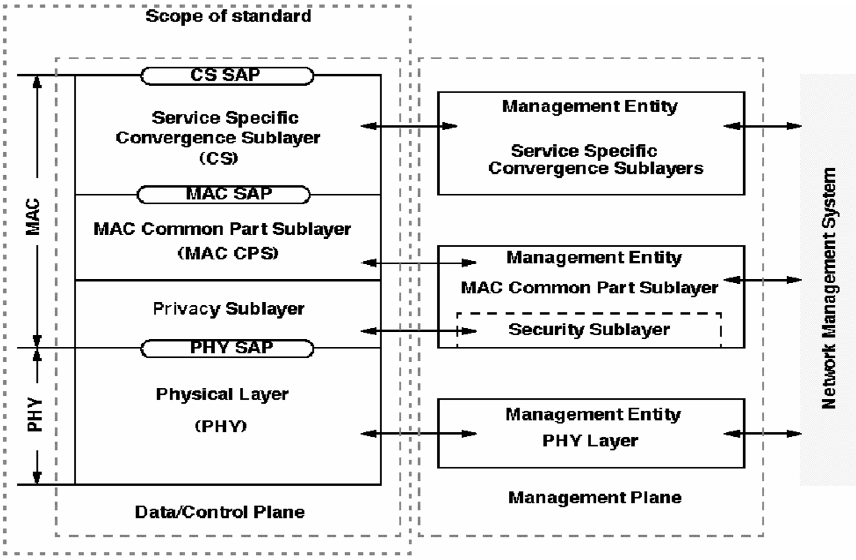


Figure 6-24: IEEE802.16 protocol layering

The standard allows management of both PMP and Mesh modes of operation. In the PMP mode, traffic from/to any SS passes through the base station (BS). In the PMP architecture the BS is the only transmitter operating in the downlink direction. The IEEE802.16a does not foresee the coordination between BSs. No handover procedure intended for supporting mobility across the network is provided by the standard.

Differently, within the Mesh mode, traffic can be routed through other SSs, without passing through a BS. The Mesh BS differs form other SS only because it is the entity that interfaces the wireless network to backhaul links. This document considers only the PMP deployment.

In order to support mobility an amendment (called as IEEE802.16e) is under drafting. The scope of the amendment is to enhance IEEE802.16a to support mobility at vehicular speeds and specifies a system for combined fixed and mobile broadband wireless access. Operation will be limited to licensed bands suitable for mobility below 6 GHz. Fixed IEEE Std 802.16-2004 subscriber capabilities will not be compromised by mobility support introduction.

6.2.4.4 MAC Common Part Sublayer (MAC CPS) in PMP mode

In determining performances of the IEEE802.16a system, the Common Part Sub layer, (CPS), plays a major role. It defines procedures to enter and acquire portion of communication medium and settles rules for sharing the available radio resources. Practically, it schedules the usage of air link between connected stations in accordance with QoS agreed by users, while maximizing the data throughput of the system. The MAC CPS provides link adaptation and Automatic Repeat Request, ARQ, (optionally and only in the 2-11 GHz frequency range) functionalities.

The first procedure an SS undergoes in order to start communication, is the Network Entry procedure. This process is realized through a sequence of operation that starts with DL channel synchronization, proceeds through DL and UL parameter acquisition, initial ranging process (for obtaining UL synchronization), capability negotiation and authentication, and finishes with registration and establishing of IP connectivity.

All data communications, either for management or data traffic, are managed in the context of connections. Connections allow to provide adequate QoS levels for each data service foreseen. Upon entering the network (and already from the initial ranging phase), three pairs of connections (uplink and downlink) are established between the SS and the BS which are dedicated to management purpose. These connections are identified as the Basic, Primary and Secondary connections with the respective Connection Identifiers (CIDs). The basic connection is used by the BS MAC and SS MAC to exchange short, time-urgent MAC management messages. The primary management connection is used by the BS MAC and SS MAC to exchange longer, more delay tolerant MAC management messages. Finally, the Secondary Management Connection is used by the BS and SS to transfer delay tolerant, standards-based [Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP), SNMP, etc.] messages.

For providing bearer services, BS may set up several connections each one identified by a proper CID. MAC CPS builds MAC PDU by adding the proper MAC header (with the appropriate CID) to the classified SDUs coming from Convergence sublayer. Optionally CRC is calculated and added for purpose of error detection.

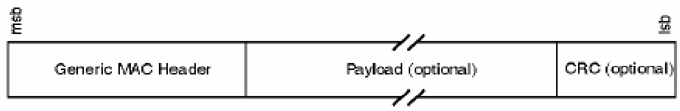


Figure 6-25: MAC PDU structure

In the PMP architecture, the BS is the only responsible for allocation of resources both in uplink and downlink directions. The scheduling provided by the MAC layer can be subdivided into two main tasks: the first step aiming to share the usage of the airlink between users and the second step for scheduling of individual packets at BS and SSs.

Regarding the airlink scheduling, the main subdivision regards uplink and downlink. The downlink is generally broadcast. All information regarding the structure of physical frame being transmitted for downlink direction is specified by the DL-MAP management messages transmitted by the BS. If not explicitly indicated by the DL-MAP itself, all SSs capable of listening to a portion of the downlink sub frame shall listen. The SSs check for connection identifiers, i.e. CIDs, in the received PDUs and retain only those PDUs addressed to them.

In the uplink direction, SSs share the uplink to the BS adhering to transmission protocol that control contention between users. Allocations are granted by BS on a per SS-base and addressed through the Basic CID assigned to each connected SS. Through UL-MAP control message the BS assigns proper allocations to all SSs.

MAC manages also link adaptation in order to optimize resource sharing according to traffic requirements and radio link state.

In case of optional features used at the physical layer, such as AAS, STC or MIMO, part of DL and UL sub frames are reserved only for capable SSs. Details about frame structures are reported below.

The second task, i.e. packet scheduling, is performed in both BS and SS MAC. Through connections, it is possible for MAC to map and furnish adequate QoS levels for each data service foreseen. This is accomplished through four different classes of traffic: Unsolicited Grant Service (UGS), real time Polling Service (rtPS), non real time Polling (nrtPS) and Best Effort (BE). Characteristics of service classes are:

- Unsolicited Grant Service: guarantees a constant bandwidth allocation not subject to bandwidth request. The allocated flow is permanently reserved and wasted if no traffic is to be transmitted. Low latency and low jitter guaranteed.
- Real time Polled: include variable bit rate services likewise real time applications. It guarantees bandwidth and provides the possibility to exceed it when more bandwidth is available. SS are polled via unicast polling, bandwidth is granted in response to BW request. Low latency and low jitter guaranteed.
- Non-Real time Polled: same as rtPS but no latency or jitter bound is guaranteed. Bandwidth is granted in response to BW request
- Best-Effort: no guarantee is provided. Bandwidth is granted upon request and availability.

In order to provide appropriate and agreed QoS level, each connection is mapped to a dedicated Service Flow. Service flows are biunivocally mapped to CIDs and carry all information regarding QoS requirements (like Maximum sustained traffic rate, Minimum reserved traffic rate, Maximum latency, Tolerated jitter…). The concept of a service flow is central to the operation of the MAC protocol. The service flows defines the QoS parameters for the traffic that is exchanged on a particular connection. In particular, they are integral to the bandwidth allocation process.

The packet scheduling task, as stated above, is carried out separately by BS and SS MAC layer. In particular in the DL direction BS allocates bandwidth for each active connection (CID) while in the UL direction each SS receives grants for transmission addressed to its specific Basic CID while the SS requires bandwidth specifying its needs in terms of byte for each one of its active connection. Upon receiving grant for transmission, the SS is then responsible for allocating data packets from its connections. The MAC layer allows several methods for SSs in order to make bandwidth requests.

On one side, the BS reserve uplink resources for allowing SS to require bandwidth allocation. It is the so called Polling procedure. Polling can be broadcast, multicast or unicast depending of the CID towards which it is reserved. In case of broadcast or multicast polling resources are reserved for contention between users following an appropriate contention resolution algorithm called full contention algorithm (based mandatory on a truncated binary exponential backoff algorithm). In case of OFDM physical layer an optional contention based BW request procedure called focused contention bandwidth request can be applied. In case of unicast polling resources are reserved to an individual SS (identified by its Basic CID) and there is no contention.

On the other side, the SS can make bandwidth requests either using stand-alone bandwidth request headers or Piggyback request (that can be optionally adopted). Furthermore, SS can make bandwidth requests either following polling indications coming from the BS or autonomously.

6.2.4.5 Simulation Overview

The scope of simulation activity performed during this phase of work, consisted in assessing basilar performances of the IEEE802.16a MAC protocol, considering only some relevant features of those mandatory specified by the standard. . The PHY layer considered is the so-called WMAN OFDM PHY, which implement and OFDM modulation based on a 256 - point transform. More details on the PHY layer considered are provided in Annex II.

6.2.4.6 Simulation Parameters

6.2.4.6.1 Simulation environment

Simulation activities have been conducted deploying a MAC protocol simulator using Network Simulator 2 (NS2) platform. NS2 is a discrete event ("event driven") simulator targeted at networking research. Ns2 provides substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless (local and satellite) networks.

6.2.4.6.2 Simulation assumptions

The software developed simulates an IEEE802.16a based system characterized by the following configuration:

- Radio Interface: OFDM (256 point Fast Fourier Transform)
- Duplexing: TDD
- Channel bandwidth: 20 MHz
- Frame duration: 10 ms.
- Sampling factor (*n*): determines the subcarrier spacing and the useful symbol time. For licensed channel bandwidths which are multiples of 1.75 MHz and 2.75 MHz, and license-exempt, the sampling factor is 8/7; for any other bandwidth: 7/6 (in our case 7/6).
- *G:* it is the ratio of CP (Cyclic prefix) time to "useful" time. Available possibilities were four: 1/4, 1/8, 1/16, 1/32; we have chosen ¼ in order to deal with delay spread values for NLOS operation in suburban areas

From the above choices follows:

- Sampling frequency F_s: $F_s = n * BW = 23.33 Mhz$
- Subcarriers spacing (Δf): $\Delta f = F_s / N_{FFT} = 91.13 Khz$
- Useful symbol time (T_b): $T_b = 1/\Delta f = 10.97 \mu s$
- Cyclic prefix (T_g): $T_g = G * T_b = \frac{1}{4} * T_b = 2.74 \mu s$
- Symbol duration time (T_{SYMB}): $T_{sym} = T_b + T_g = 13.71 \mu s$
- Maximum theoretical number of OFDM symbols per frame: 729

6.2.4.6.3 Implementation details

<u>Frame structure</u>

In order to simulate the MAC layer behaviour, a TDD frame structure compliant with specification has been defined. As explained in previous section, the MAC frame in the TDD mode is formed by two sub frames, the uplink and downlink. The first sub frame consists of a PHY header, the Long Preamble (2 OFDM symbols), a certain amount of OFDM symbols for transmitting the FCH, DL-MAP and immediately after, the UL-MAP management messages. The exact number of OFDM symbols needed for

DL-MAP and UL-MAP transmission depends on the actual number of information elements that are to be transmitted. In turn this number depends on the number of active connections that are to be served. The remaining part of the DL sub frame is left for transmission of a number of bursts, equal to the number of IE inside the DL-MAP.

The UL sub frame allocates a ranging window composed of 4 ranging slots each one composed of a long preamble, one OFDM symbol for transmitting the RNG-REQ message and a time gap equal to the maximum round trip delay corresponding to the intended coverage range of the cell. The ranging window is left empty in order to simulate the presence of the ranging procedure which actually has not been implemented.

The ranging window is followed by a Bandwidth request region comprised of 10 contention slots available for Bandwidth request. Each Transmission Opportunity is 2 OFDM symbols long. After these two contention regions there are allocation for uplink PHY PDUs to be transmitted by SSs.

In order to consider time needed for switching from transmission to reception state and vice versa, the TTG (Tx to Rx Time Gap) and RTG (Rx to tx Time Gap) have been inserted.

The frame structure implemented is reported in the following figure.

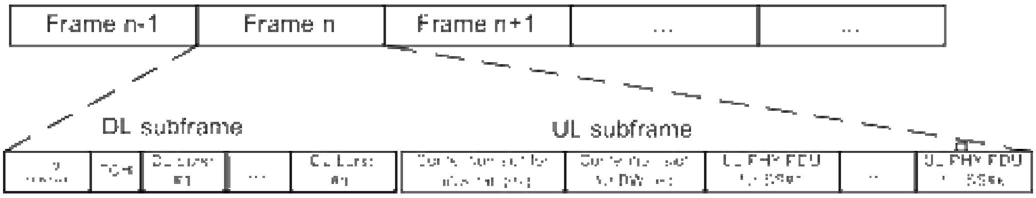


Figure 6-26: Implemented frame structure.

Each UL PHY PDU is composed by a short preamble followed by one UL burst which in turn contains a number of MAC PDU each one preceded by its own MAC header specifying the CID to which it is addressed to.

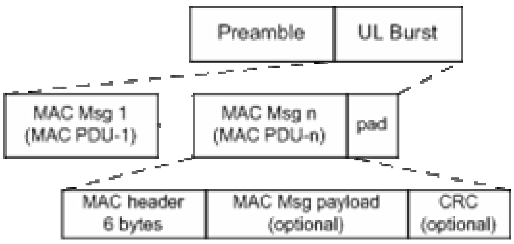


Figure 6-27: Detailed UL burst.

<u>MAC Management messages</u>

Only DL-MAP and UL-MAP messages have been implemented. In particular, the DL-MAP completely defines the sequence of bursts that complete the downlink sub frame and in particular the Connection ID to which they are addressed to. The UL-MAP has the role to inform all Subscriber Stations under the coverage of Base Station when and in which way they have the access to the radio channel. The UL-MAP defines in particular the beginning of the DL sub frame through the Allocation Start Time field. The Allocation Start Time is determined considering the following information, all available to the BS:

- RTD_{MAX} : indicates the maximum Round Trip Delay related with the farther SS
- TTG indicates the time gap left for switching the BS from Tx to Rx
- “DlData”: is an integer number of OFDM symbols used for transmitting Downlink Data. This value is deduced by the DL-MAP message.
- The Allocation Start Time results then:
- $RTD_{MAX} + TTG + DlData$

<u>UL - DL sub frame switching</u>

The standard does not give indications about determination of the DL and UL sub frame duration but explicitly requires that the switching point between the two regions has to be variable. Because of the lack of indication, we decided to determine the maximum DL sub frame duration in terms of OFDM symbols.

The reason for this resides in that a certain amount of bandwidth (even if fixed) for the DL is always guaranteed. In this way, if BS hasn't enough data to fill all the allocation, it will occupy only one part of it; on the other side, if the amount of data coming from all CIDs in the DL direction is larger than the maximum allocation, not all data will be transmitted, but only those ones that fit the given allocation. This maximum DL sub frame length is assured even if UL allocation would require more bandwidth allocation. The position of the switching is spread to all connected SS through the Allocation Start Time field inside the UL-MAP message. The maximum number of OFDM symbols established to be available for downlink data transmission is set to 388 OFDM symbols, with assumptions reported above; when this limit is reached the uplink sub frame is approximately 70-75% of the downlink sub frame.

Connections

As stated above, the IEEE802.16 MAC layer is connection oriented, i.e. each data communication between BS and SS is managed through connections. In the simulator only two kinds of connections has been considered: broadcast connections allowing the transmission of MAC management messages (DL-MAP and UL-MAP) to all SSs, and unicast connections between the BS and each one of the SSs. In order to maintain simulation complexity low, only one connection per SS can be allowed. Furthermore, no ranging procedure has been implemented so data connection are instantiated through an extern file at the beginning of the simulation.

Burst profiles

All available burst profiles have been considered and are reported in the following table together with uncoded and coded Block sizes:

Table 6-13: Coded and uncoded block sizes.

PHY mode m	Modulation and coding rate	Uncoded block size [byte] (BpS_m)	Coded block size [byte]
1	QPSK ½	24	48
2	QPSK ¾	36	48
3	16-QAM ½	48	96
4	16-QAM ¾	72	96
5	64-QAM 2/3	96	144
6	64-QAM ¾	108	144

Adaptivity of the physical layer has not been implemented. All simulation considered all SSs with same of modulation-coding couple.

Bandwidth request

In order to simulate bandwidth request procedure in UL, a bandwidth request window has been left in the UL sub frame, as explained above. In order to acquire resources to transmit in UL, SSs enters a full contention procedure based on the mandatory contention resolution method which is a truncate binary exponential backoff algorithm. In the following figure the detailed structure of the BW request window are reported.

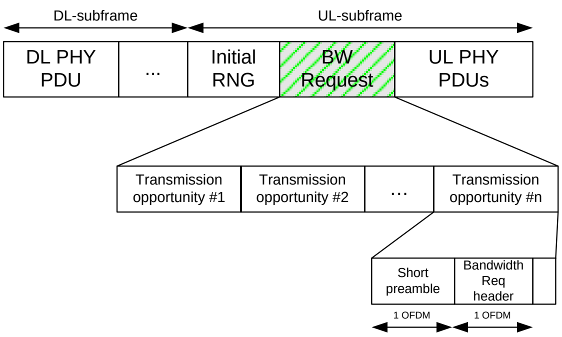


Figure 6-28: Bandwidth Request phase.

Once acquired initial allocations, SSs update BS knowledge about their needs using both stand-alone bandwidth request header and piggyback requests (whose implementation is optional). As stated in the standard, piggyback requests are only incremental while stand-alone request header can be both incremental or aggregate. Both kinds of requests are sent using part of the uplink allocation reserved by BS.

MAC PDU length

Fragmentation and packing features has not been implemented. The absence of such features required that MAC PDU dimensions were kept low in order to avoid waste of bandwidth otherwise filled with padding. The dimension fixed for all MAC PDU has been fixed then at 72 Byte. Considering a fixed MAC header of 6 Byte the MAC SDU resulted equal to 66 Byte. The CRC has not been considered in the simulations.

72 Byte MAC PDUs can be allocated always in a limited number of OFDM symbols with all available modulation and coding couples available:

- QPSK ½: 3 symbols per MAC PDU
- QPSK 3/4: 2 symbols per MAC PDU
- 16 QAM 1/2: 3 symbols per 2 MAC PDU
- 16 QAM 3/4: 1 symbol per MAC PDU
- 64 QAM 2/3: 3 symbols per 4 MAC PDU
- 64 QAM 3/4: 2 symbols per 3 MAC PDU

The drawback of this choice is, from the other side, the MAC overhead that is introduced which lower considerably the performances. Fragmentation appears to be a desirable feature for the system.

6.2.4.6.4 Simulation scenario

The simulation scenarios considered deal with a unique BS serving a variable number of Subscriber Stations. Number of SSs varies from a minimum of 10 SS to a maximum of 50 SS (PMP deployment).

The area considered for simulations is a square region with side 350 m long. This choice has been made due to some limits imposed by NS-2, as increasing distances the computational complexity grows to a point to slow simulations. For avoiding this problem a smaller area has been chosen as platform for simulations. This assumption influences slightly performances in terms of throughput as coverage distance impacts on initial ranging window dimension reserved in the UL sub frame. Ideal propagation conditions (i.e. no error due to channel) have been considered. This ideal assumption allowed estimating the ultimate performances of the protocol.

6.2.4.7 Results

6.2.4.7.1 Throughput analysis

The throughput analysis conducted allowed to evaluate the peak user data rate and the net throughput of the system versus offered traffic and variable number of users.

Evaluation has been done through analytical calculation and confirmed also with simulation (simulations have been run for reciprocal validation vs calculations). The analysis allowed to evaluate peak throughput provided by the system in ideal conditions. In particular, it has been possible to evaluate the influence on protocol efficiency of some parameter settings like, cyclic prefix length and intended maximum coverage range.

Analytical evaluation of throughput

The analytical evaluation of throughput can be done determining the net bandwidth (i.e. the actual number of symbols that are transmitted per frame) available for data transmission. In a frame of T_{FRAME} duration, the theoretical maximum number of symbols that can be transmitted is given by:

$$N_{sym/FRAME}^{MAX} = \frac{T_{FRAME}}{T_{sym}}$$

The resulting theoretical gross bitrate corresponding to this max number of symbol per frame can be immediately obtained by:

$$Grossbitrate = \frac{BpS_m \cdot 8}{T_{sym}} \left[\frac{bit}{\sec} \right],$$

where BpS_m indicates the uncoded block size in byte that can be transmitted over an OFDM symbol when PHY mode*m* is employed (see Table 6-13: Coded and uncoded block sizes.).

Actually, due to the presence of PHY overhead (Long and Short preambles), MAC management messages, Ranging and Bandwidth request phases, the actual bitrate achievable is lower. In order to obtain a realistic evaluation of throughput, all these contributions have to be estimated. The PHY/MAC overhead extension can be estimated deterministically considering their specific dimensions as defined in the standard:

- Long preamble: N_{LP} =2 OFDM symbols
- FCH message: N_{FCH} = 1 OFDM symbol
- Short preamble (N_{SP} = 1 OFDM): each UL PHY PDU, one for each SS connected (let n_{SS} be the number of SS actually served) is preceded by 1 Short Preamble. The comprehensive number of symbols occupied by short preambles can then be calculated as: $n_{SS} \cdot N_{SP}$.
- DL-MAP, UL-MAP, DCD and UCD messages length: the comprehensive number of symbols occupied by these messages is variable and depends mainly on number of active connection but also on the actual characteristics of the channel. Let be , N_{MACmsg} this variable number of OFDM symbols.
- Bandwidth Request phase: it is formed by a certain number of request opportunities. Let's indicate with N_{BWReq} = 2 OFDM the duration in symbols of each BW request opportunity, and with n_{BWReq} the number of BW request opportunities. The dimension of Bandwidth request phase is then: $n_{BWReq} \cdot N_{BWReq}$.
- Initial Ranging phase duration: the duration of such phase, as reported above, depends from the maximum coverage range of the cell because each initial ranging opportunity lasts for enough time for allowing all SS in the coverage range to acquire the UL synchronization and entering the network. Each ranging transmission opportunity comprehends a time gap equal to the RTD_{MAX} (Round Trip Delay max for the cell) plus time for RNG request message transmission (3 OFDM symbols). Let's indicate with T_{RNG} the duration of each ranging opportunity, and with n_{RNG} the number of ranging opportunities. Differently from quantities above, the Ranging phase cannot be quantified in an exact number of OFDM symbols.

The actual number of symbols that can be derived as follows:

Let's $T_{FRAME}^{NET} = T_{FRAME} - n_{RNG} \cdot T_{RNG}$ be the actual portion of frame available for transmission of symbols.

$N_{sym/FRAME}^{NET} = \left\lfloor \frac{T_{FRAME}^{NET}}{T_{sym}} \right\rfloor - (N_{LP} + N_{FCH} + n_c \cdot N_{SP} + N_{MACmsg} + n_{BW\,Req} \cdot N_{BW\,Req})$ indicates the net number of OFDM symbols available for transmission. Once determined this number it is possible to derive the net amount of bytes which can be actually transmitted. This number depends, once again, from the modulation/coding couple selected for transmission, i.e. form the uncoded block size that can be transmitted per each OFDM symbol, BpS_m (see Table 6-13: Coded and uncoded block sizes.).

Such amount of bytes can be spent for MAC PDU transmission. Subsequently the number of MAC PDU that can be transferred is given by:

$$n_{MACPDU} = \frac{N_{sym/FRAME}^{NET} \cdot BpS_m}{MACPDU_{length}}$$

MAC PDU_{length} expresses the length in Byte of each MAC PDU.

The net bitrate that can be provided by the system then is given by:

$$NetThroughput = \frac{n_{MACPDU} \cdot (MACPDU_{length} - MACheader[Byte] - CRC[Byte]) \cdot 8}{T_{FRAME}} \left[\frac{bit}{\sec}\right]$$

the standard specifies a 6 bytes length MAC header for each MAC PDU. The 4 bytes CRC field, instead can be optionally calculated.

It has to be noted that the n_{MACPDU} has been calculated assuming that MAC PDUs can fill all the available OFDM symbols continuously. This assumption requires actually that fragmentation is practiced. Without fragmentation, only MAC PDUs multiple in length of the uncoded block size (BpS_m) of the PHY mode used can fill exactly an integer number of OFDM symbols. MAC PDUs lengths that do not satisfy these constraints would cause waste of bandwidth in padding.

6.2.4.7.2 Peak user data rate

The analytical evaluation and simulations allowed to estimate the peak throughput provided by the system in ideal conditions. In particular, it has been possible to evaluate the influence on protocol efficiency of some parameter settings like, cyclic prefix length and intended maximum coverage range. Our evaluation of peak user data rate considered a coverage radius spanning from 0.5 km up to 30 km, 4 ranging slots and 10 BW request opportunities, MACPDU length of 72 Byte with no CRC field.

In the table below are reported values for peak data rate for all burst profiles per Hz.

These values are obtained from above equations simply considering:

$$netbitrate = \frac{NetThroughput}{ChannelBandwidth}\left[\frac{bit}{\sec \cdot Hz}\right]$$

Table 6-14: Net bit rate per Hz available.

	net bit rate bps/Hz		
Tg	$1/4T_b$		
Cell radius	500m	10 km	30 km
64QAM3/4	2,69	2,62	2,47
64QAM2/3	2,39	2,33	2,19
16QAM3/4	1,80	1,75	1,64
16QAM1/2	1,20	1,16	1,10
QPSK3/4	0,90	0,87	0,82
QPSK1/2	0,60	0,58	0,55

	net bit rate bps/Hz		
Tg	$1/32T_b$		
Cell radius	500m	10 km	30 km
64QAM3/4	3,29	3,21	3,03
64QAM2/3	2,93	2,85	2,69
16QAM3/4	2,20	2,14	2,02
16QAM1/2	1,46	1,43	1,34
QPSK3/4	1,10	1,07	1,01
QPSK1/2	0,73	0,71	0,67

Results evidenced the relevant impact of coverage radius and G (i.e. Cyclic prefix length) setting on peak throughput performances. In fact, the standard specifies different values for cyclic prefix in respect to useful symbol time T_b, spanning from $1/4T_b$ to $1/32T_b$. The radio environment influences the selection of the appropriate guard time (cyclic prefix) for each OFDM symbol. The differences in throughput corresponding to these two extreme cases are relevant. The coverage radius, on the other side, influences the maximum round trip delay (RTD) that the system should support. The RTD influences the ranging phase duration (each ranging slot comprises 3 OFDM symbols plus RTD max interval in order to avoid possible interference between tx and rx data). Also the number of ranging slots reserved per frame influences the net throughput.

Considering the theoretical gross bitrate of a system with analogous settings, the resulting overhead, due to MAC and PHY spans, in average, between 13% (with 500 m cell radius) and 20% (with 30 km cell radius). As stated above, the effect of coverage range on throughput is related to the dimensioning of the initial ranging window. Due to the no trivial effect of coverage range, a proper dimensioning of the Initial ranging phase should be done.

Tables above evidence the impact on net throughput imputable to cyclic prefix dimensioning. From the above analysis, passing from $T_g = 1/4T_b$ to $T_g = 1/32T_b$, about the 20% of net throughput becomes available.

As result form the analytical evaluation of net throughput, further factors influencing the peak user data rate are: number of Bandwidth (BW) request opportunity (2 OFDM symbols each), MAC management messages length in symbols, MAC SDU dimension in byte (which influence the overhead introduced by MAC headers (6 byte header per MAC PDU) and the optional presence of CRC (4 byte per MAC PDU).

6.2.4.7.3 Throughput vs Offered traffic

The simulator allowed to obtain throughput curves versus increasing offered traffic when 10 SSs are to be served. Results are reported in the following figure. Particular assumptions for the simulation run are:

- Cyclic prefix length: ¼ T_b
- Channel Bandwidth: 20 MHz
- MAC PDU dimension: 72 Byte
- No CRC

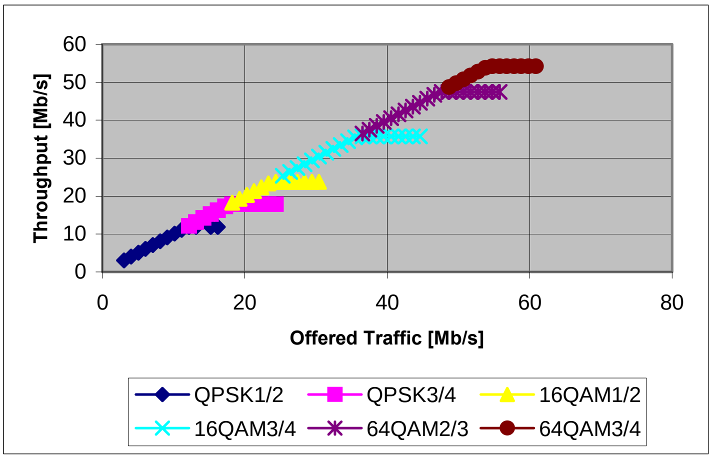


Figure 6-29: Overall simulated throughput.

Simulated and calculated values for net throughput agree tightly. From the table below, the overhead introduced by the MAC protocol appears to be in the order of 15%.

Table 6-15: Throughput values versus number of users.

Modulation/Coding	Maximum Theoretical Throughput	Max Simulated Throughput for 10 SS	Max Simulated Throughput for 30 SS	Max Simulated Throughput for 50 SS
QPSK 1/2	14 Mb/s	11,86 Mb/s	11,29 Mb/s	10,77 Mb/s
QPSK 3/4	21 Mb/s	17,82 Mb/s	17,00 Mb/s	16,15 Mb/s
16QAM 1/2	28 Mb/s	23,73 Mb/s	22,65 Mb/s	21,56 Mb/s
16QAM 3/4	42 Mb/s	35,70 Mb/s	34,01 Mb/s	32,31 Mb/s
64QAM 2/3	56 Mb/s	47,46 Mb/s	45,39 Mb/s	43,28 Mb/s
64QAM 3/4	63 Mb/s	53,58 Mb/s	50,98 Mb/s	48,53 Mb/s

The increase in the number of served SSs corresponds to a parallel decrease of throughput mainly related with the increasing dimension of the DL-MAP and UL-MAP messages.

6.2.4.7.4 Delay analysis

A basilar evaluation of the protocol behaviour in respect to delay has been conducted. The analysis ealt with a comparison of the end-to-end delay introduced by protocol when serving several SS under different traffic conditions with two proportional scheduling strategies.

IEEE802.16a defines a complete architecture able to provide a tight respect of QoS constraints that are negotiated by users. As explained in the general description of the standard, there are four different service classes defined (UGS, rtPS, nrtPS and BE). Each connection is then mapped to a dedicated service flow which carries information regarding the service class and the specific traffic parameters contracted by users. Anyway, the specific packet scheduling policies definition is left to manufacturer's implementation.

The end-to-end delay evaluation, conducted using the protocol simulator developed, aimed to test the basic behaviour of the MAC protocol under different traffic sources and different load conditions.

The scheduling policy considered is a general proportional round robin algorithm employed in both DL and UL direction.

In DL, BS is completely informed about status of all queues and from this knowledge is able to subdivide the available bandwidth between all connected users and to compile the DL-MAP message. The DL-MAP message will inform all SS of the DL sub frame structure that will be transmitted in the current frame.

Packet scheduling is based on a proportional round robin algorithm which allocates portions of frame to all active users following two basic criteria:

- First (packet policy): the BS allocates to a particular connection (CID) a portion of DL sub frame proportional to number of bytes currently present in the queue of that CID.
- Second (flow policy): the BS evaluates the incoming packet flow in each queue (CID). Allocations will be proportional to the evaluated incoming flows.

In UL, the local scheduler of each SS is responsible of keeping the BS up-to-date about its bandwidth requirements which, in the current implementation, are related with only one connection. The BS provides SSs with appropriate grants on the base of information received through bandwidth requests messages. As stated above, bandwidth requests messages implemented can be MAC request headers (which can be aggregate or incremental) and Piggyback requests (which can be only incremental). SSs, which manage only one connection, create bandwidth requests following analogous criteria to those used in DL, i.e. based on packet or flow policies. Grants generated by BS will be proportional to the requests received from all connected SSs. The SS scheduler, in our implementation, is only responsible for creation of proper bandwidth requests, while actual packet scheduling is performed again by BS.

In order to be able to smooth out the variability of traffic sources, which was found to be severe in the UL direction when request have to be generated (i.e. meaning higher overhead transmitted), a simple SMA (Smoothing Moving Average) algorithm has been implemented. The SMA can be applied with both policies described above and can be tuned independently between BS and SSs.

The updated value is obtained by: $A_d = A_{d-1} + F * (M_d - A_{d-1})$

A_d represents the new trend number while A_{d-1} the old one. M_d is the updating value (fluctuating). The smoothing factor, $F \in [0,1]$, allows to regulate the capability of the algorithm of following the traffic fluctuation.

Results reported consider the BS serving 20 SSs with increasing offered load. The traffic sources considered are, respectively, CBR ON/OFF and Poisson traffic sources.

<u>CBR ON/OFF Traffic</u>

This type of traffic is characterized by setting a Burst time of 1004 ms and an Idle time of 1587 ms. The resulting average transmitting state is of about 38% of the simulation time. The burst profile used is the most robust (QPSK 1/2) one and the number of stations is set to 20.

With CBR ON OFF traffic types, evaluations considered a smoothing factor F equal to 1.

6.2.4.7.5 Downlink Sub frame

In the following figure it is presented the CCDF of the downlink delay with a Flow policy.

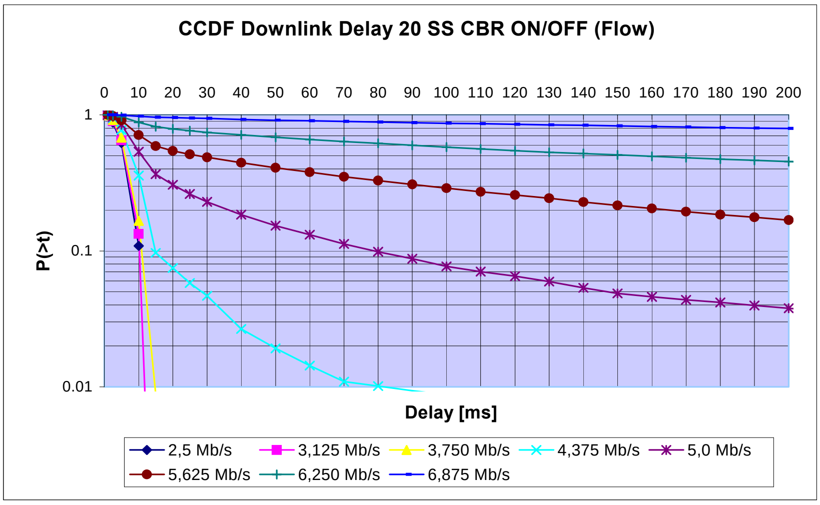


Figure 6-30: CCDF DL, end to end delay, CBR ON/OFF, flow policy

The different curves represent different Average overall Offered Traffic in DL direction. The above curves, gives the probability that packets should wait in their transmitting queue due to saturation of the available DL bandwidth.

Considering the maximum DL sub frame length in symbols has been settled equal to 388 OFDM symbols, the maximum available DL capacity is about 7.25 Mbps (analytically evaluated). Considering variability of traffic sources associated with each SS, effects of DL saturation appear already with an average overall traffic of 4.375 Mbps.

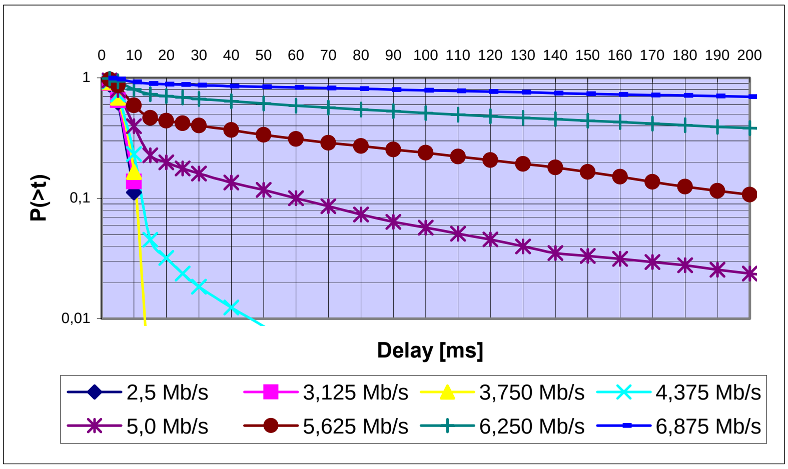


Figure 6-31: CCDF DL, end to end delay, CBR ON/OFF, packet policy

The same simulation has been conducted using the packet policy; with offered traffic approaching to the maximum capacity, this discipline leads to low delays. As it is clear from the above figure, for example, comparing it to the fourth “Flow” curve, the per-cent-rate of packets with a delay exceeding 15 ms is lower (about 3%).

The following graph underlines differences between scheduling policies regarding average end-to-end delays.

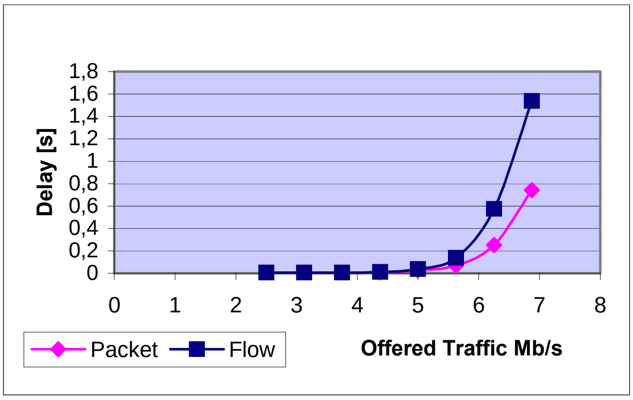


Figure 6-32: CBR ON/OFF Packet vs Flow policies comparison

Approaching to a congestion state, the "packet" politic reduces considerably the average end-to-end delay: With an Average Offered Traffic of 6-7 Mb/s the gain is of about 200 ms.

6.2.4.7.6 Uplink Sub frame

The analysis made on the uplink sub frame follows is similar to that of downlink sub frame. The traffic considered is characterized by the same burst and idle times.

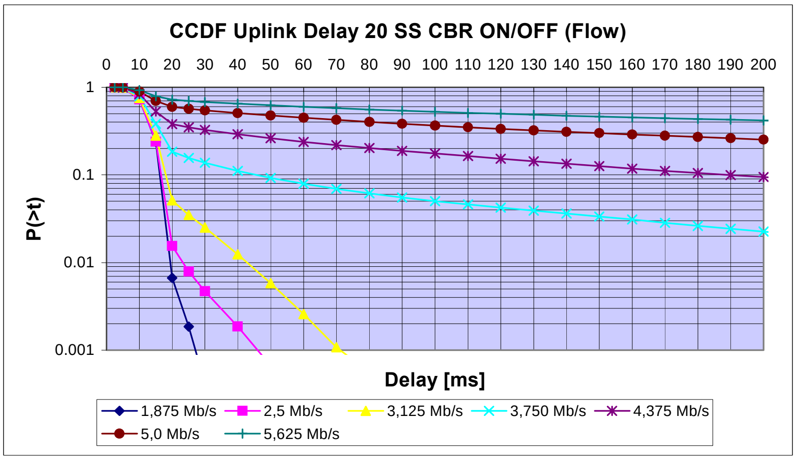


Figure 6-33: CCDF UL, end to end delay, CBR ON/OFF, flow policy

In this case, the average end-to-end delay is considerably increased due to the delay introduced by bandwidth request algorithm. All the curves are shifted of at least one frame (10 ms) in comparison to the CCDF of the downlink delay. In the UL, Subscriber Stations try to enter the contention resolution process more times in a simulation; this concept has a confirmation given by a flex in the curves around 20 ms of delay more evident when dealing with low traffic rates. For low traffic rates, the low per-cent-rate of packets exceeding this threshold is due to the latency introduced by the backoff algorithm; for higher traffic rates this latency is mixed with the delay introduced by the overcoming of the uplink capacity.

The same simulation has been conducted with the "Packet" policy.

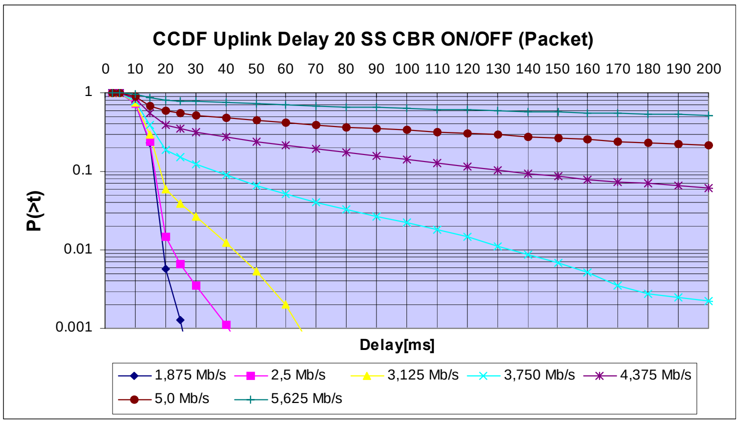


Figure 6-34: CCDF UL, end to end delay, CBR ON/OFF, packet policy

Also in this case the packet policy gives advantages even if not in a substantial way as in the downlink case, always during a congestion phase.

The strategy adopted in the simulator for determination of switching point between uplink and downlink with this kind of traffic a lot variable, brought more advantages for the uplink traffic; in fact during periods with low concentration of transmitting BS sources, the frame space reserved to the uplink transmission increases up to the point that the uplink traffic is carried with low delays than the downlink traffic. On the contrary, during periods when a limited number of Subscriber Stations are active, the downlink traffic is not able to exploit the frame portion not used by the uplink due to the limit of 388 OFDM symbols imposed. This is confirmed by the following graph where, in comparison to the downlink one, the average end-to-end delays during congested phases are considerably lower.

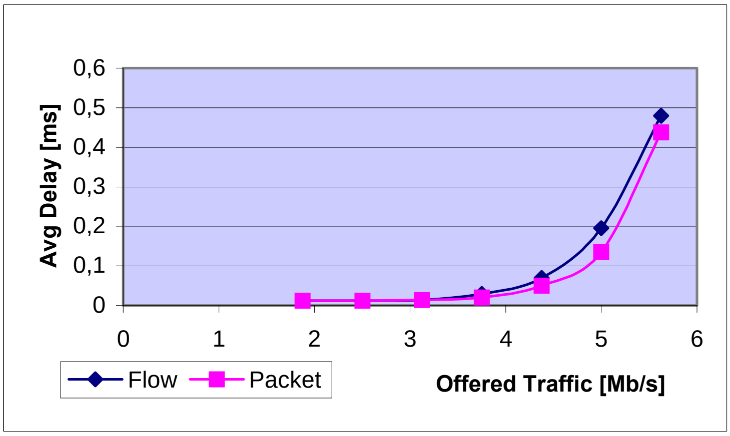


Figure 6-35: CBR ON/OFF Packet vs Flow policies comparison (UL).

<u>Evaluations using Poisson Traffic</u>

With this kind of traffic the Smoothing Moving Average appeared to be necessary due to variable highly packet inter-arrival times. The SMA smoothing factor (F) choice has been kept different between uplink and downlink. This difference is base on the fact that BS is perfectly aware about its queues state, while each SS has to generate proper bandwidth request messages (i.e. overhead) in order to inform the BS about its needs.

Analysis conducted with Poisson traffic are similar to those carried out for CBR ON OFF.

6.2.4.7.7 Downlink sub frame

In DL a SMA factor set equal to 0.5 provided better performances respect to the setting (F = 1) that allows to strictly follow the fluctuation of the Poisson traffic.

To notice that while the packet policy get worst by decrementing the SMA factor the flow policy improves leading to lower end-to-end delays.

After these, we have though to decrease the SMA factor to 0.5 in order to notice if any changes but also in this case the packet policy has resulted more performing.

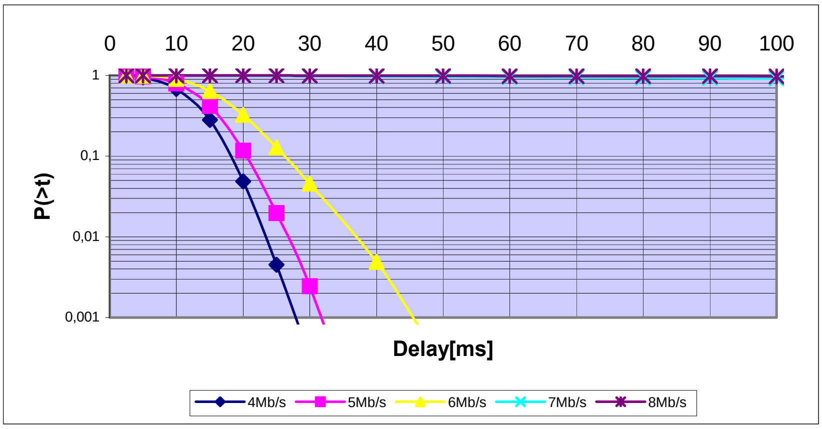


Figure 6-36: CCDF DL, end to end delay, Poisson, flow policy, F=0.5.

With this parameters tested it is possible to say that the packet policy in general performs better but it is not sure that the results remain the same varying ulteriorly the SMA factor.

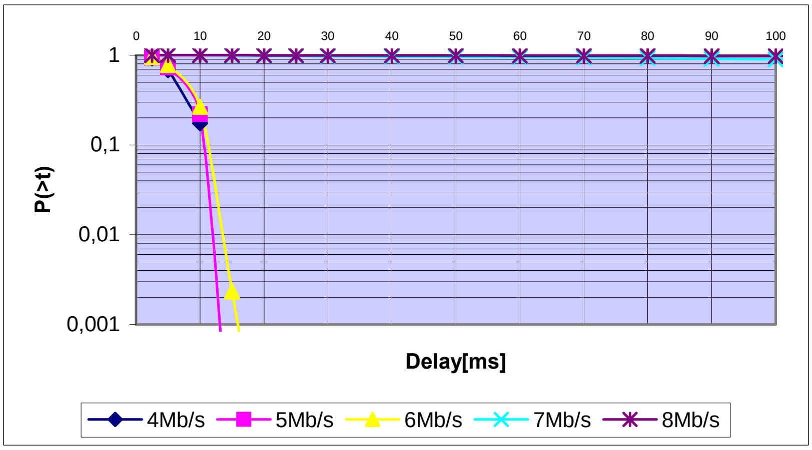


Figure 6-37: CCDF DL, end to end delay, Poisson, packet flow policy, F=0.5.

6.2.4.7.8 Uplink sub frame

The strong fluctuations in perception of bandwidth by the Subscriber Stations would lead on one side to an introduction of overhead onto the channel due to the attempt to satisfy requests and on the other to the generation of fluctuations at the Base Station scheduler. In fact, as the uplink scheduler follows a proportional policy, it would give more bandwidth to SS that have asked more, taking bandwidth out of the other SS. These SS will ask for more bandwidth in next frames causing fluctuations; this process difficultly would come to an end without a low SMA factor. For all the following simulations an SMA factor of 0.19 is used as it has been considered a good compromise between following fluctuation and introducing overhead but the specific optimal case has not been studied.

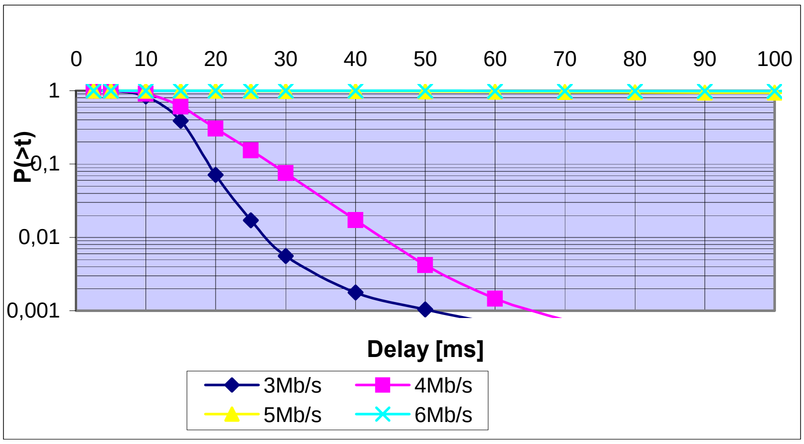


Figure 6-38: CCDF UL, end to end delay, Poisson, flow policy, F=0.19.

These graphs deal with the CCDF of the uplink delays comprising delays due to the bandwidth request phase. Also in this case an evaluation about the Packet and Flow policies has been provided.

Results of previous paragraphs are not confirmed; the flow policy performs better than that packet one. Maybe this fact is due to the decreased SMA factor that keeps off flexibility to the policy.

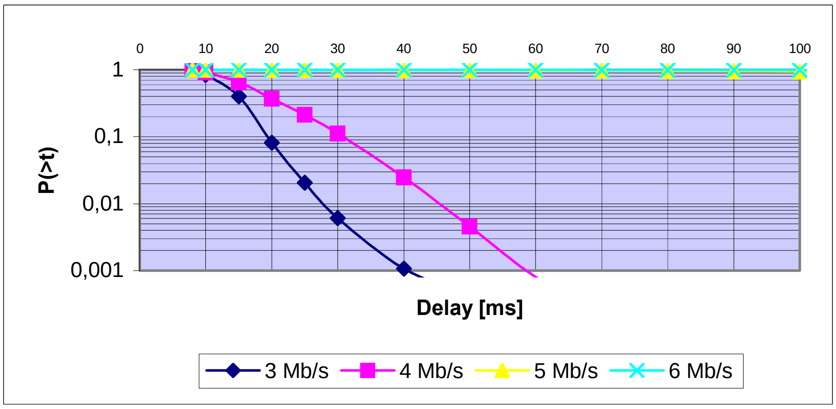


Figure 6-39: CCDF UL, end to end delay, Poisson, packet policy, F=0.19.

6.2.4.8 Conclusions

Performance evaluations have dealt with Access Delay, End-to-End delay, Throughput with variable number of Subscriber Stations and traffic sources.

It is figured out that the overall MAC overhead of the IEEE 802.16a system can vary approximately from a minimum of 15%, when dealing with 10 SS, to a maximum evaluated of 23% with 50 SS. The Throughput performances decrease as the number of SS grows with a rate of 4,8% every 20 SS. In comparison to the Gross Bit Rate of 63 Mb/s, with the most performing modulation and coding scheme, with 50 SS, the maximum Throughput evaluated has been of 48,5 Mb/s that has grown to 53,6 Mb/s with 10 SS.

It is clear that the optional features of packing and fragmentation are powerful to optimize the system throughput mostly when the number of Subscriber Stations sharing the MAC frame, grows. With packets of 72 bytes length (Data + MAC overhead), the lack of these features is partially made up but at a loss of an increased overhead; with this packet configuration throughput is reduced of the 8%.

The choice of the switching point between the uplink and downlink sub frame has advantaged, with CBR and Poisson traffic, the downlink sub frame in a substantial manner; with CBR ON/OFF traffic, the uplink has more flexibility and can exploit frame portion not used by the downlink traffic; uplink end-to-end delays remain superior to the downlink case but grow faster when approaching of saturation.

We conducted an analysis on which was the best policy for Bandwidth Request to adopt between the “Packet” and the “Flow” one. In general the Packet policy appears to be better performing but only in condition near to saturation with a maximum gain evaluated of 100 ms in terms of average delay.

6.3 Advanced link metrics

6.3.1 Introduction

Several algorithms, such as routing or Adaptive Modulation and Coding (AMC, a.k.a. Link Adaptation), rely on the quality of the link to make their decision. Basically, the AMC will select the best combination of constellation and coding rate so that the signal is transmitted at the highest possible speed with the guarantee of a good service (low PER) to minimize the delay. Similarly, routing will make sure that the selected route allows the minimum end-to-end transmission delay (low PER) with an optimized channel utilization (high data rates).

To do that, there is a need to evaluate the link quality to select the "best" route or the best constellation/rate. The link quality can be, for example, measured in terms of PER thresholds. And, to get an estimate of the PER, two solutions are available:

direct estimation: by a computation of the received ACK (if such a protocol is implemented)

indirect estimation: by measuring a channel metric and by assuming that the PER is a function of this metric

Once a solution is selected (direct or indirect estimation, or even a hybrid method combining the two), the adaptive algorithm must be defined. In the case of AMC, several algorithms are proposed that are based either on direct or indirect estimation. In the case of routing, a similar process can be followed.

In this initial contribution, the focus will not be on the routing algorithm itself, but on the adequate selection of the channel metric allowing an indirect accurate estimation of the PER. By using channel metrics, this report will not be dependent on the implementation of an ARQ algorithm. The only limit presented by indirect estimations is the availability of some channel information. The discussion will therefore deal with the necessary inputs that a chip should provide to the MAC layer to make the indirect estimation possible.

The following section will consequently be dedicated to the description of various candidate/recommended link quality metrics, followed by initial simulation results displaying the accuracy of such metrics. An on-going activity within WINNER consists in adapting this metrics to the systems that will eventually be adopted (in terms of complexity and feasibility evaluation), and in using adequately these metrics for some algorithms, including, in the context of WP3.2, the routing algorithms. At last, discussions open the way to possible real life implementations of algorithms based on the proposed metrics.

6.3.2 Description of several link quality metrics

As one of the FITNESS deliverables puts forward [59], the goal is to be able to forecast the PER at any instant for a particular user, provided the user's channel matrix H, interference channel matrices $H_1 \dots H_k$ and thermal noise energy N_0. This can be envisaged in terms of curves of the form PER(C) where C=C(H, $H_1, \dots, H_k$, N_0) is the scalar performance metric. The identification of a metric C that can adequately map all channel, interference, and noise parameters on to a single scalar is indeed a challenging task. Besides, the definition of C also encompasses MIMO cases (H_1 being a NxM channel matrix, where M is the number of transmit antennas, and N the number of receive antennas), and multi-carriers systems (such as OFDM, with Hi being a vector of channel frequency coefficients for every useful sub-carriers of dimension $1xN_u$, where $N_u=48$ in the IEEE 802.11a/g or ETSI HiperLAN/2 case).

In what follows, the authors will not deal with the MIMO case. An extension to this case is a possibility depending on the choices made in the WINNER consortium, in order to extend the following approach in the case of multi-antenna systems. Therefore, some restricting assumptions are made for the considered approach on multi-carrier systems, illustrated in the IEEE 802.11a/g case.

In the following, we assume that the complex channel frequency coefficients Hi are known to the receiver for every useful sub-carrier of index i in $0, N_u-1$ with $N_u=48$ in 802.11a/g. We denote by G_i the squared modulus of the coefficients ($G_i=|H_i|^2$). In addition, the noise plus interference variance σ_n^2 is also known and remains constant throughout the packet. The transmitted signal power is σ_s^2. The interference effect is comprised in the noise (considered as Gaussian).

- **Reception SNR**

At the output of the channel the SNR equals:

$$SNR = \frac{1}{N_u} \sum_{i=0}^{N_u-1} \frac{G_i \sigma_s^2}{\sigma_n^2} \tag{46}$$

- **Shannon Capacity**

The Shannon Capacity equals:

$$C = \frac{1}{N_u} \sum_{i=0}^{N_u-1} \log_2 \left(1 + \frac{G_i \sigma_s^2}{\sigma_n^2} \right) \tag{47}$$

Channel capacity only makes use of the squared modulus of the channel coefficients.

Formula above can also be computed with a margin on noise power, i.e. assuming the noise variance is greater than σ_n^2.

- **Channel Gain Variance (M1) and efective SNR based metrics (M2) and (M3)**

The first one was suggested in [60]:

$$M_1 = \frac{1}{N_u} \sum_{i=0}^{N_u-1} \left(|H_i| - |H| \right)^2 \tag{48}$$

Where $|H|$ is the average channel gain

$$|H| = \frac{1}{N_u} \sum_{i=0}^{N_u-1} |H_i| \tag{49}$$

Contrary to Shannon Capacity, M_1, has to be combined with the SNR in order to determine the PER. Just like the Shannon capacity, this metric has low implementation complexity. However, its relationship with the PER performance is questionable.

Two more metrics, M_2 and M_3, have also been investigated. They are both based on the evaluation of the effective SNR that was introduced in [61] and later applied to COFDM in [62]. While their computation and assumption made are extensively described in the Annex, here only the final expression is provided:

$$M_2 = \min_{n<RB_m; c\in C(n,d_f)} \left[\frac{1}{d_f} \sum_{i=1}^{d_f} \frac{G_{\Pi_3^M(p_c(i))} \sigma_s^2}{\sigma_n^2} \right] \tag{50}$$

$$M_3 = H^{-1} \left[\frac{1}{RB_m} \sum_{n<RB_m; c\in C(n,d_f)} \frac{1}{N(n,d_f)} H\left(\frac{1}{d_f} \sum_{i=1}^{d_f} \frac{G_{\Pi_3^M(p_c(i))} \sigma_s^2}{\sigma_n^2} \right) \right] \tag{51}$$

The implementation complexity of these metrics is reasonable, although greater than that of Shannon Capacity or M_1 metric. The codewords at distance d_f can be found by exhaustive search. However, another difference with [61] and [62] comes from the puncturing. Without puncturing, in the equation (49) and (50), the codewords c considered for the n^{th} node are the same as for node 0 shifted by *n/r* bits. In other words, the most likely error events are the same all along the trellis. Therefore, $N(n, d_f)$ is independent of *n* and only a few codewords have to be stored (e.g. in 802.11a, $\forall$ n $d_f = 10$ and $N(n, d_f) = 11$). With puncturing, the free distance is reduced and may be reached only for a subset of nodes, depending on the value of *n mod P*. For instance, with the coding rate R=¾ of 802.11a, error events starting at nodes such that *n=0 mod 3* have minimum Hamming weight of 6. The free distance equals 5 and is reached only for 6 error events starting at nodes such that *n=1 mod 3* and 1 error event when *n=2 mod 3*. Thus, since we restrict the computation to codewords at the free distance, only two thirds of the nodes have to be considered. Obviously, metrics based on code knowledge are perfectible, and are worth improving as shown in the performance analysis later in this document.

6.3.3 Performance of the various link quality metrics

In this section, we try to analyze the performance of each link quality metric presented in the previous section. In a first step, we focus on 64QAM constellation, which is the worst scenario in 802.11a for metrics M_2 and M_3. However we do not puncture the rate ½ code in this first step. This will be done afterwards. We assume 128 bytes packet size. The channel is assumed perfectly known both for decoding and for link quality metric computation.

6.3.3.1 SNR metric and validity of effective SNR

Contrary to the classical approach that consists in using the SNR as a link metric, Figure 6-40 suggests that a better approach might be to estimate the offset of the PER vs SNR curve for a given channel with respect to the AWGN reference. On Figure 6-40, we shifted curves so that they coincide with the AWGN reference at PER=5%. The fact that they are almost parallel with the reference is confirmed, since the difference never exceeds 1dB. This is equivalent to using the effective SNR, as defined in the Annex.

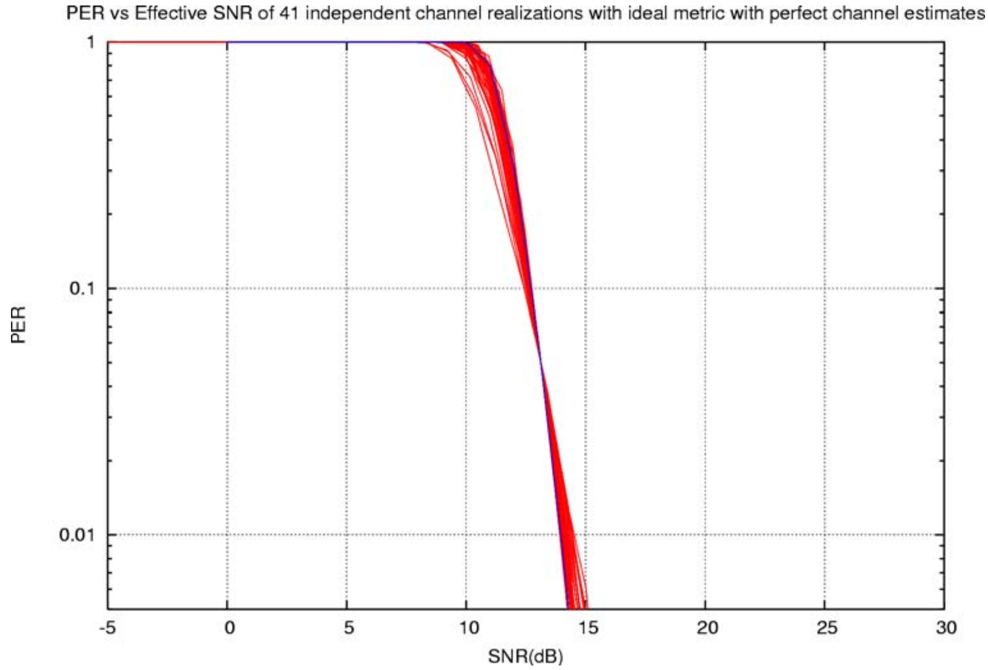


Figure 6-40: Ideally shifted PER vs SNR curves for 40 independent channels, BRAN-A model

6.3.3.2 Performance of M_1 metric

On Figure 6-41, we plot the SNR offset w.r.t. the AWGN reference at a PER of 5% for 40 independent trials. We then derive a Look-Up Table (LUT) by least-squares fitting. As stated in [60], the precision of M_1 is much better than by using the SNR. This is confirmed on Figure 6-42, where the PER vs Effective SNR is plotted for the same channels.

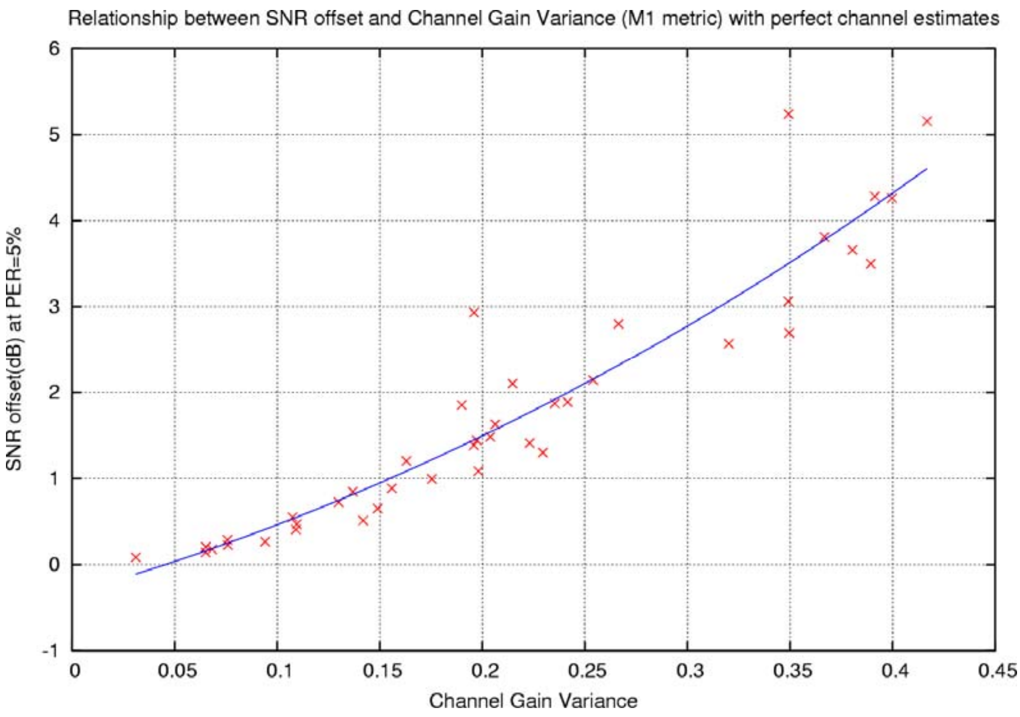


Figure 6-41: SNR offset at PER 5% with metric M_1 on 40 independent channels, BRAN-A model

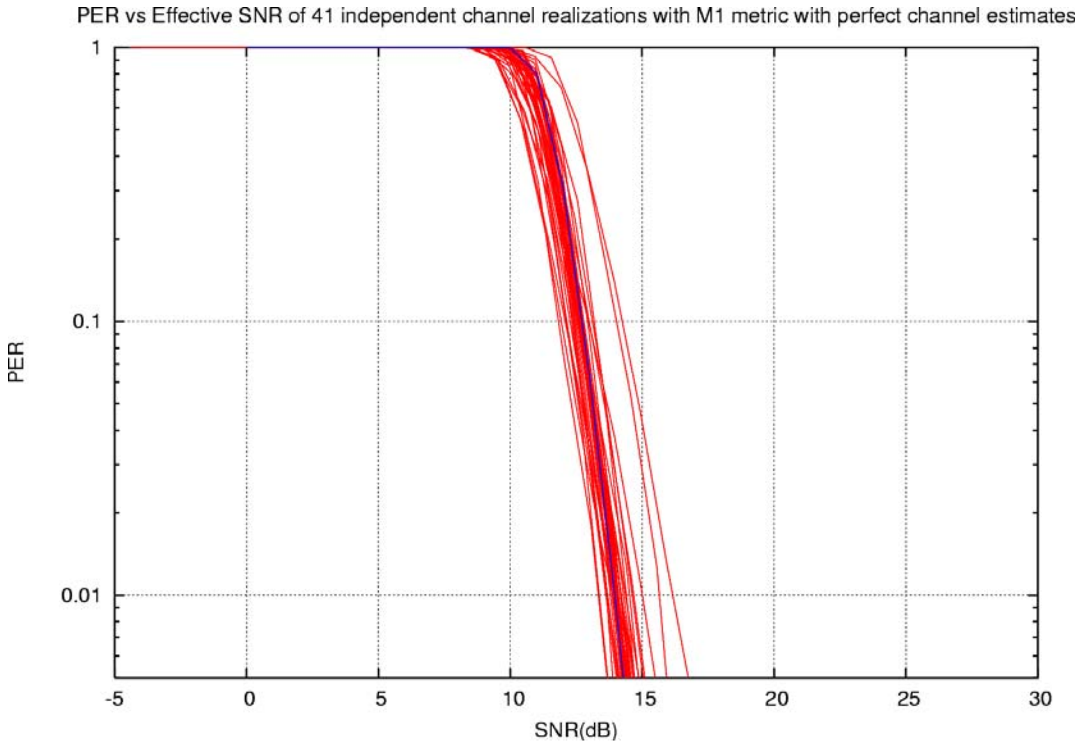


Figure 6-42: PER vs. effective SNR with metric M_1 on 40 independent channels, BRAN-A model

6.3.3.3 Performance of Shannon Capacity metric

On Figure 6-43, we plot the PER vs Shannon capacity for the 40 independent BRAN-A channel trials. The blue curve represents a LUT obtained by LS fitting. This LUT will provide a reasonably good estimate of the PER, as shown on Figure 6-44: by taking a 1.5 dB margin, PER estimation errors will never exceed 10%.

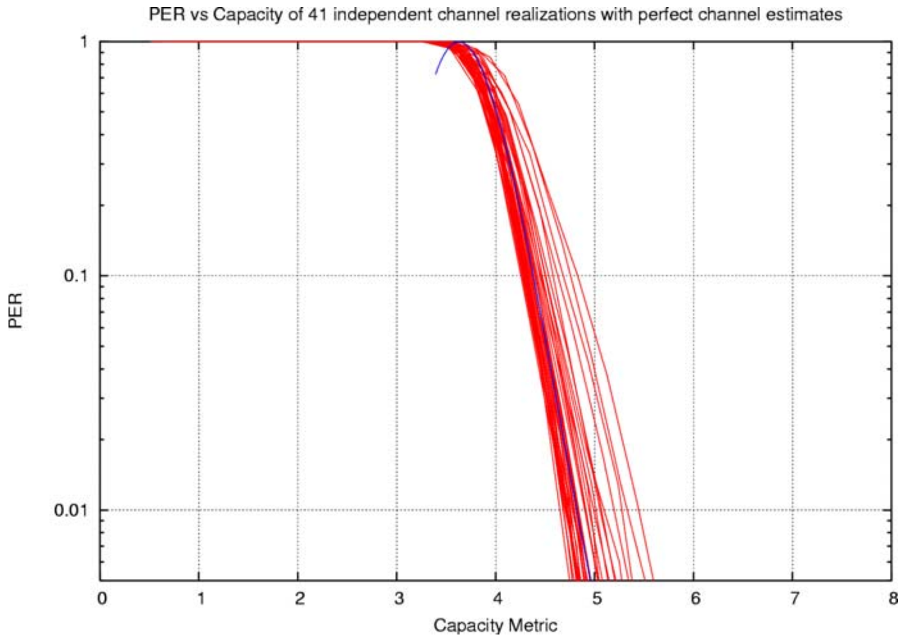


Figure 6-43: PER vs Shannon Capacity for 40 independent channels trials, BRAN-A model

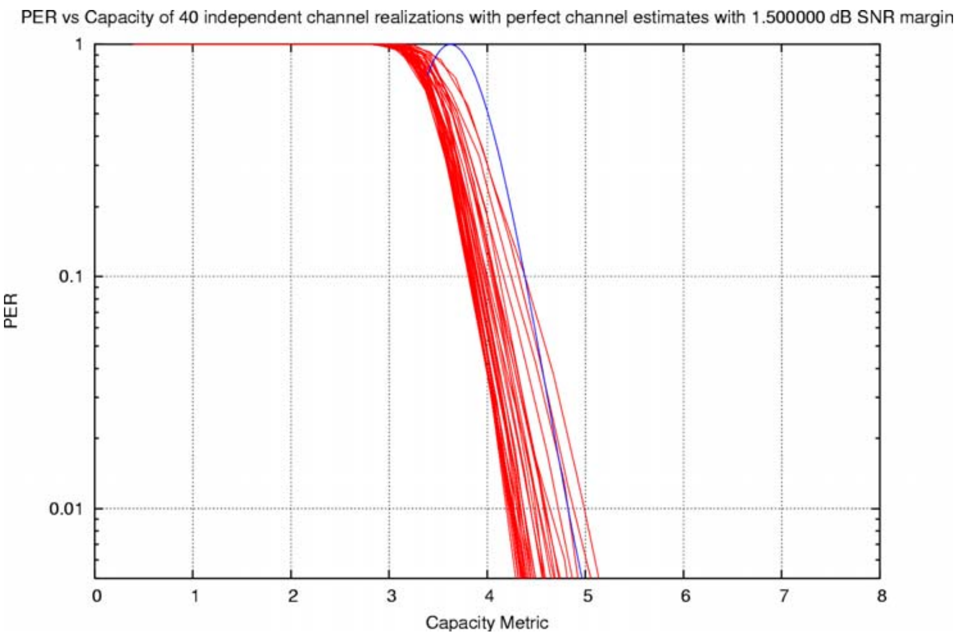


Figure 6-44: PER vs. Shannon Capacity for 40 independent channels trials, BRAN-A model, 1.5 dB SNR margin

6.3.3.4 Performance of M_2 and M_3 metrics

On Figure 6-45 and Figure 6-46, we represented the PER vs. effective SNR performance for metrics M_2 and M_3. The AWGN performance is plotted in blue for reference. It can be observed that M_2 is overly conservative compared to M_3, which is logical since M_2 assumes that the frame error rate is imposed by the most likely error event. The important parameter is the range of Effective SNR at a given PER. At a PER=1%, this range is of about 3dB for M_2 and 2dB for M_3. These figures have to be compared with the range of 7dB if the SNR was used.

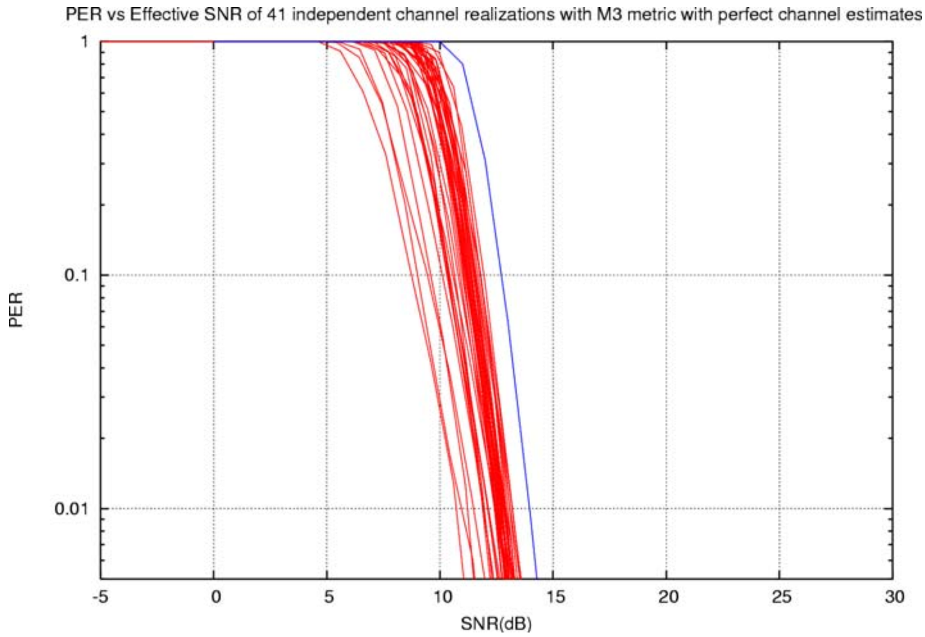


Figure 6-45: PER vs. effective SNR for M_2 link quality metric, 40 independent channels, BRAN-A model

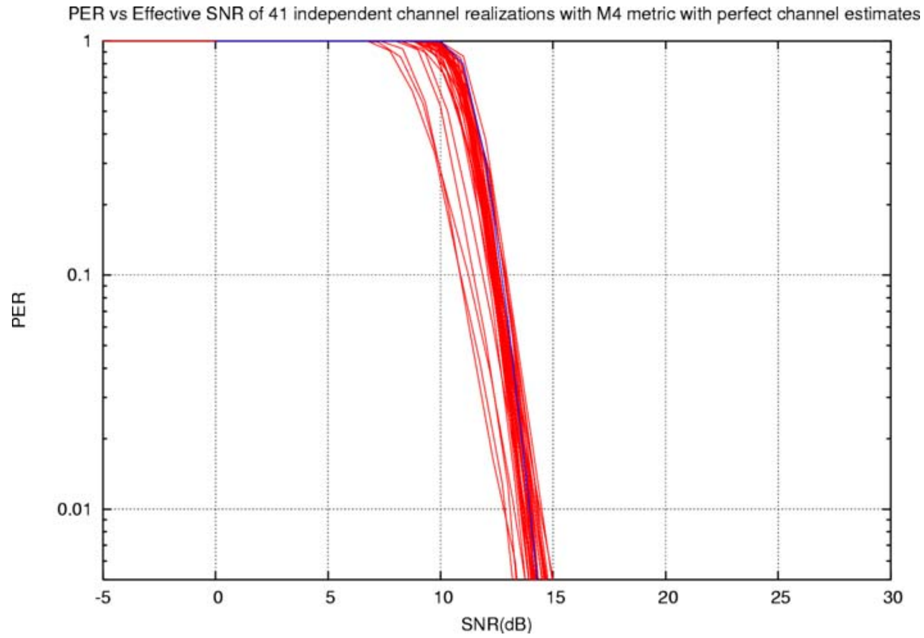


Figure 6-46: PER vs. effective SNR for M_3 link quality metric, 40 independent channels, BRAN-A model

6.3.3.5 Comments on interference and RSS measurement precision

In this report, we did not take into account interferers. The PER actually depends on the number, duration and power of simultaneous interferers, and also on the channel on which each interfering signal is received. Including such a complete interferer model in the link quality metrics presented before would be very complex. We propose in a first step to adopt a rather conservative approach which consists in adding the maximum interfering power in the packet σ_i^2 to the noise power σ_n^2 in the equations presented before. The problem remains that the interfering power depends on the sub-carrier index. This problem was briefly addressed in FITNESS [64] in the context of HSDPA and can be studied in future work.

Another issue which will occur when designing routing algorithms based on advanced link metrics is the impact of imprecision in the RSS measurements due for instance to AGC gain inaccuracies and RSS signalling quantization. These errors cannot be suppressed by a better PHY abstraction.

As a result, it might turn out that the practical solution to routing algorithms in real conditions be a hybrid data fusion of several estimators, including channel metrics. The problem of advanced routing taking the link metrics into consideration will be tackled next.

6.3.4 Conclusions

In this contribution, we completed a first study on appropriate link quality metrics for OFDM PHY layer performance estimation. We showed that link metrics involving not only the SNR but also the channel frequency coefficients can bring down the uncertainty on the SNR required to achieve a low-enough 1% PER from 7dB to about 2dB.

Three metrics within those studied show good performance, achieving a 2dB uncertainty range. The first one is the Shannon capacity. The advantages of this metric are that it is not very complex to compute, it can be easily extended to multiple antennas, and it is likely to be even more accurate in future systems when coding schemes approaching the capacity will be implemented (e.g. Turbo-Codes, LDPC). Such systems are realistic and under consideration within WINNER. The channel gain variance metric (M1) has also good performance, but its extension to multiple antennas is not straightforward. Another metric which shows promising results was derived from union bound theory. Its advantage is that it takes into account the exact coding and interleaving scheme. Therefore, it could potentially yield a more precise PER estimate. However, it is more computationally complex than capacity, and cannot be directly extended to Turbo-Codes for instance.

6.4 Conclusions

Some of the well known systems: 3G (HSDPA), WLAN IEEE 802.11, WLAN HiperLan/2 and WMAN 802.16a, where described and their deployment concepts were illustrated with respect to the identified scenarios. Furthermore, system level simulation results were shown to evaluate the performance of identified enhanced conventional deployment concepts in terms of traffic performance and feasibility. This can be used as a reference for the benchmarking against the future system, since the simulation results are aligned with the identified scenarios and the defined set of input and output parameters (as figures of merit - FoM) aligned with the recommendations of the D7.2, ensuring the “comparability” of the results.

7. Annex II: Additional Information on Multi-hop Concepts

7.1 IEEE 802.16

7.1.1 Physical Layer and Frame Structure

The target frequency range identified by WINNER is below 11 GHz, and hence, only the standard IEEE 802.16a is taken into account in the following. In the IEEE Standard 802.16a [82] there exist several PHY specifications for the 2 - 11 GHz spectrum. The most common and meanwhile agreed physical layer within WiMAX is Orthogonal Frequency Division Multiplexing (OFDM) transmission scheme with 256 point Fast Fourier Transformation (FFT). Both, Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes are specified, which differ only slightly. However, only the TDD mode is specified both for PMP and Mesh mode [82]. Since a comparison will be carried out in later sections we concentrate on the TDD mode here. Moreover, TDD is most likely to be better suited for multi-hop support.

On the physical layer data bits enter the channel coding block, which is composed by three steps: randomizer, Forward Error Correction (FEC) and interleaving. These procedures are applied in reverse order at the receiver. The FEC is realized by a Reed Solomon (RS) outer code, which is concatenated with a rate Compatible Convolutional (CC) inner code. After interleaving the data bits are entered serially to the constellation map of the modulator. In Table 7-1 the available combinations between channel coding and modulation is reported.

Table 7-1. Channel coding per modulation

Type	Modulation	Uncoded block size [byte]	Coded block size [byte]	Overall coding rate
BpS_1	QPSK	24	48	1/2
BpS_2	QPSK	36	48	3/4
BpS_3	16-QAM	48	96	1/2
BpS_4	16-QAM	72	96	3/4
BpS_5	64-QAM	96	144	2/3
BpS_6	64-QAM	108	144	3/4

Besides Quadrature Phase Shift Keying (QPSK) also M-ary Quadrature Amplitude Modulation (QAM) with 16 and 64 constellations is supported. In combination with different code rates link-adaptation becomes possible with 6 different modulation and coding schemes, $BpS_1 – BpS_6$.

Since the modulation and coding and structure of the basic bursts are almost the same for PMP and Mesh mode, no further distinction is made (check the bursts / preambles!?). However, the frame structure is quite different. Since the frame structure is closely related to the mechanism of medium access and resource request and allocation mechanism, the frame structure will be introduced in the respective sub-sections of the MAC description.

7.1.2 Medium Access Control

The MAC layer is divided in three sub-layers based on the functions that every layer supplies. The Service-Specific Convergence Sub-layer (SSCS) acts like interface to higher layers. It is designed for ATM and packet services in order to guarantee the QoS to the specific flow. It provides the mapping functions between the upper layers and the MAC Common Part Sub-layer (MAC CPS) connections. Below the SSCS, the MAC functions are provided by the MAC CPS. Further below is the Privacy Sub-layer (PS). The primary task of the PS is to set up secure connections among SSs [91].

The MAC CPS is necessary to take advance of the technology of the PHY layer. The protocol must provide functions like access control, bandwidth management, scheduling, power control, etc.

The MAC is connection-oriented. All data communication is in the context of a connection. The connection has to be established, maintained, and finally terminated. Shortly after registration, connections are associated with a service flow, to provide a reference against which to request bandwidth.

Service flows provide a mechanism for QoS management on the uplink and downlink. The service flow identifies the QoS parameters of the PDUs transmitted on the respective connection.

The SS has a unique 48-bit MAC identifier, which is used during initial ranging and during association with the BS. Connections in the PMP mode are identified by a 16-bit Connection Identifier (CID). The type of service and other current parameters of the service are implicitly in the CID and can be accessed via a lookup-table indexed by the CID. In the Mesh mode, the nodes are identified by a 16-bit Node Identifier (Node ID), which is transferred in the Mesh sub-header, which follows the generic MAC header. For addressing nodes in the neighbourhood, 8-bit Link Identifiers (Link ID) are used. The Link ID is transmitted as part of the CID in the generic MAC header. Service parameters are added to the CID in the Mesh mode.

7.1.2.1 Point-to-Multi-Point Mode

The frame structure for the PMP mode using OFDM as transmission scheme is illustrated in Figure 7-1[82], [91].

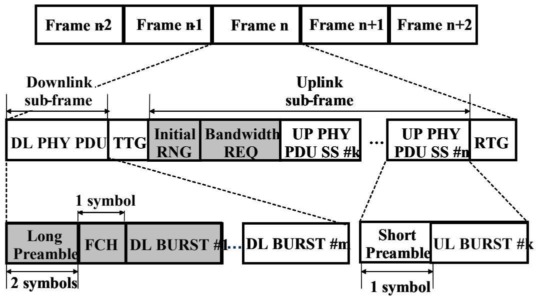


Figure 7-1. Frame structure for PMP mode

A frame consists of a downlink sub-frame (DL sub-frame) and an uplink sub-frame (UL sub-frame). The Tx/Rx transition gap (TTG) and Rx/Tx transition gap (RTG) shall be inserted between sub-frames to allow terminals to turn around from reception to transmission and vice versa. The SSs share the uplink to the BS on a demand basis. The DL sub-frame provides to SSs information regarding available channels in the wireless network and resources allocated for data transmission both in uplink (UL) and downlink (DL) direction. Furthermore, the DL sub-frame serves for data transfer in the DL. The UL sub-frame is used by terminals that want to gain access to the network, by users that require a new resource allocation and to deliver data from the SSs to the BS on the UL. The data transmission is in the context of connection between the BS and the SS, which is identified by a Connection ID of 2 byte in the MAC header.

The DL sub-frame consists of only one DL PHY PDU that starts with long preamble of length 2 OFDM symbols (L_{LP}) used for PHY synchronization purpose. Immediately following, the Frame Control Header (FCH) field specifies the burst profile and length of the DL BURST #1. The FCH is one OFDM symbol length and is encoded with the most robust PHY mode QPSK 1/2. The DL BURST #1 is followed by multiple DL BURSTs, each DL BURST consists of an integer number of OFDM symbols.

The characteristics (modulation and coding scheme, etc.) of downlink and uplink channels are broadcast by means of the Downlink Channel Descriptor (DCD) and Uplink Channel Descriptor (UCD) message, respectively. Bandwidth allocations for data transmission in downlink and uplink direction are provided by the BS respectively in the Downlink MAP (DL-MAP) and Uplink MAP (UL-MAP) message. These MAC management messages, i.e. DCD; UCD, DL-MAP and UL-MAP, could be contained partially in the FCH field and the rest in the DL BURST #1.

An UL sub-frame consists of contention based intervals scheduled for initial ranging and bandwidth requests purposes in which respectively the Ranging Request (RNG-REQ) and Bandwidth Request (BW-REQ) message can be transmitted. Immediately following, UL PHY BURSTs are assigned to different SSs for uplink data transmission, which have been signaled in the UL-MAP. Every UL PHY BURST is made up of a short preamble of length 1 OFDM symbol (L_{SP}) and an integer number of OFDM symbols.

7.1.2.2 Mesh Mode

The MAC CPS controls the efficient exploitation of the medium that is shared by multiple users. The frame structure for the Mesh mode is illustrated in Figure 7-2.

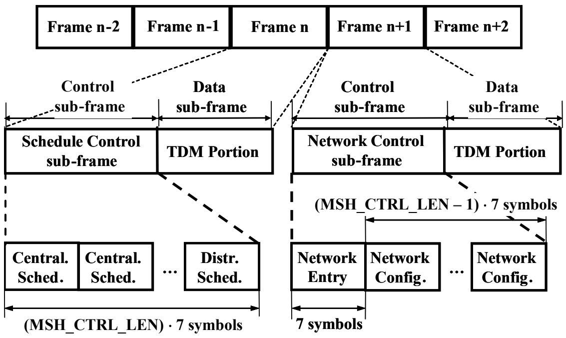


Figure 7-2 Frame structure for Mesh mode

A Mesh frame consists of a Control sub-frame and a Data sub-frame, which are fixed in length. The length of the Control sub-frame, L_{CS}, expressed as number of OFDM symbols, is fixed to:

$$L_{CS} = 7 \cdot MSH\_CTRL\_LEN \tag{52}$$

The MSH_CTRL_LEN can have a value between 0 and 15 and is distributed by the Mesh BS [82]. Two types of Control sub-frames exist, the Network Control sub-frame and the Schedule Control sub-frame. During frames in which the Schedule Control sub-frame is not scheduled the Network Control sub-frame is transmitted.

The Network Control sub-frame serves primarily for new terminals that want to gain access to the network. It is used to broadcast network information to all SSs and it provides means for a new node to gain synchronization and initial network entry into a Mesh network. This type of Control sub-frame occurs periodically, whereas the period is a network parameter that can be varied. The respective management messages are not described in detail in this section since the overhead introduced by the Network Control sub-layer will be neglected due to the typically large period for this type of control frame.

The efficiency on MAC layer primarily depends on the structure of Schedule Control and Data sub-frames. These sub-frames and their purposes are described in detail in the following sub-sections.

7.1.2.2.1 Data Sub-frame

The data transfer within the Data sub-frame is connection-oriented. One link shall be used for all the bidirectional data transmissions between these two SSs. Downlink and uplink sub-frames are not distinguished. The Data sub-frame serves for the transmission of user data in variable length PHY BURSTs (Figure 7-3).

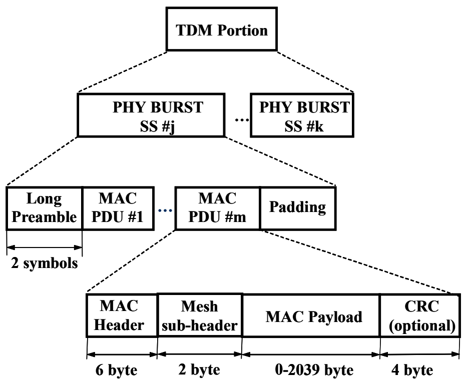


Figure 7-3. Burst structure of MAC PDU

The PHY BURST starts with a Long Preamble L_{LP} (2 OFDM symbols), but neighbouring SSs can negotiate to use the Short Preamble (1 OFDM symbol) [82]. MAC PDUs are inserted immediately following the preamble in order to fulfill the allocated resources.

The MAC PDU consists of fixed length MAC Header, a fixed length Mesh sub-header, a variable length MAC Payload and an optional Cyclic Redundancy Check (CRC). Since the size of the payload is variable, the length of the MAC PDU can vary between 6 and 2051 byte.

7.1.2.2.2 Schedule Sub-frame

The Schedule Control sub-frame is used to determine the amount of allocated transmit resources for a link, which is served within the Data sub-frame. The Schedule Control sub-frame carriers the MAC management messages, see Figure 7-4.

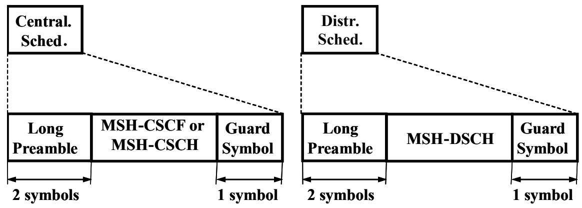


Figure 7-4. MAC management messages in the Schedule Control sub-frame

The Mesh BS decides on the number of distributed scheduling messages in the Schedule Control sub-frame and indicates this information by the MSH_DSCH_NUM field, which is transmitted in the Network Control sub-frame. From the total of MSH_CTRL_LEN control messages the first (MSH_CTRL_LEN - MSH_DSCH_NUM) are allocated to transmission bursts containing Mesh Centralized Scheduling (MSH-CSCH) and Mesh Centralized Configuration (MSH-CSCF) messages. The remainder MSH_DSCH_NUM are allocated to transmission bursts containing Mesh Distributed Scheduling (MSH-DSCH) messages (see Figure 7-4). MSH-CSCH and MSH-DSCH messages shall handle portions of the network where respectively the distributed and centralized scheduling is applied.

The data transmissions are in the context of a link, which is established between two SSs. A link is set up by means of a resource request, which is initiated by a SS, and a resource grant. A resource grant corresponds to a fraction of time in the Data sub-frame. If the Mesh CS mode is selected, the Mesh BS assigns the granted resources for each link in response to resource requests. MSH-CSCH messages are used for this purpose. In case of Mesh DS mode, the neighbouring SS responds to the request with a corresponding grant for the link between the involved two SSs. MSH-DSCH messages are used for this purpose. The Mesh CS and Mesh DS modes can be deployed simultaneously.

7.1.3 Resource Allocation and Random Access

One key feature of an air-interface is the acquisition of transmit resources, i.e. resource requests and resource grants in IEEE 802.16a. In the following it is assumed that the station already has been registered and associated with the corresponding BS, respectively Mesh BS. However, the SS was inactive for a longer time and, hence, no resources are explicitly reserved for that individual station, neither in the DL nor in the UL.

7.1.3.1 PMP Mode

SSs initially need to request transmission resources from the BS, which in turn constructs the frame and signals the UL and DL frame structure within the FCH and DL Burst#1 (see also description in previous sub-section 7.1.2.1).

Consequently, the transmission in the DL can be started instantaneously, since the BS generates the frame. Therefore, the challenge is the allocation of transmit opportunities in the UL, i.e., from the SS to the BS.

Requests of SSs for transmission are based on the 16-bit CID in PMP mode. Different methods are foreseen in the standard to acquire resources from the BS.

7.1.3.2 Mesh Mode

In the Mesh mode the resource allocation is closely related to the scheduling mechanisms, i.e., whether centralized or distributed scheduling or both is used. In case of centralized scheduling, the Mesh BS determines the schedule of all SSs and therefore, the allocation of transmit resources in the DL can be easily managed. This is different in case distributed scheduling is used. In the latter case the transmit permission in DL and UL is managed between the SSs. For distributed scheduling there is no difference between DL and UL with respect to the resource allocation scheme, and it will be described in the respective sub-section.

7.1.4 Distributed and Centralized Scheduling

In the PMP mode only centralized scheduling is supported. The BS assigns the resources for the connections between the BS and associated SSs. Since the resource management is only defined for one-hop connections, i.e., between BS and SSs, the scheduling is closely related to the mechanisms for the resource request and resource assignment, based on the frame structure defined for PMP mode. Hence, a further description is not provided. The specific scheduling algorithm, e.g. round-robin, earliest-due-date, weighted-fair queuing, etc., is not specified in the standard, but is left to the individual manufacturer.

In Mesh mode, two modes for scheduling are supported: distributed and centralized scheduling. Both modes serve to manage the resource allocation for multi-hop connections. I.e., mechanisms are specified, how resource allocations can be organized and realized in excess of the direct communication range of an individual station. Specifically, for the centralized scheduling scheme, where the Mesh BS organizes the order of transmission, SSs receive their schedule from the Mesh BS indirectly via relaying capabilities of other SS, in case they are out of range of the Mesh BS. In the following sub-sections a detailed description of the two scheduling schemes for the Mesh mode is given.

7.1.5 Protocol Efficiency for PMP and Mesh Mode – Analytical Derivation

In the following two sections the protocol efficiency, respectively the overhead, for the Mesh and PMP mode are derived.

7.1.5.1 Protocol Efficiency for PMP Mode

The number of MAC PDUs in a frame as function of OFDM symbols for data transmission is derived as

$$\boldsymbol{N^{k,m}_{MAC\ PDU} = \left\lfloor \frac{L_{DP} \cdot BpS_m}{k} \right\rfloor} \tag{53}$$

where k denotes the MAC PDU length in bytes. The MAC PDU length when the PMP mode is applied contains a header of 6 bytes, a CRC of 4 byte and a payload that can vary in the interval between 1 and 2041 byte.

$$\boldsymbol{L^k_{MAC\ PDU} = L^k_{Payload} + 10 = k,\ \ k \in [11\ ;\ 2051\ byte]} \tag{54}$$

If the PMP air-interface is deployed in a multi-hop scenario the Mesh sub-header of 2 bytes is embedded, hence the MAC PDU length is:

$$\boldsymbol{L^k_{MAC\ PDU} = L^k_{Payload} + 12 = k,\ \ k \in [13\ ;\ 2051\ byte]} \tag{55}$$

The number of OFDM symbols available for data transmission within the PMP frame is equal to:

$$\boldsymbol{L_{DP} = \sum_{i=1}^{n_c} \left(DL\ BURST_i + UL\ BURST_i \right)} \tag{56}$$

where n_c is the number of bidirectional connections. The number L_{DP} depends on the control fields introduced by the MAC protocol:

$$\boldsymbol{\begin{aligned} L_{DP} = {} & N_{symbols} - L_{LP} - L_{FCH} - L_{RNG} - L_{BW-REQ} + \\ & - L_{DL\ BURST_1} - L_{SP} \cdot n_c \end{aligned}} \tag{57}$$

where $N_{symbols}$ is the total number of OFDM symbols in each frame, and L_{LP} and L_{FCH} are respectively the length of the long preamble and the frame control header. L_{RNG} and $L_{BW\text{-}REQ}$ correspond to the length of the initial ranging and the bandwidth request field, respectively. $L_{DL\ BURST\ \#1}$ is the overhead introduced by UL-MAP, DL-MAP, UCD and DCD messages, and L_{SP} is the length of the short preamble.

$$\begin{aligned} L_{DL\ BURST_1} &= \left\lceil \frac{OH_{UL-MAP} + OH_{DL-MAP} + OH_{UCD} + OH_{DCD}}{BpS_1} \right\rceil \\ &= \left\lceil \frac{214 + 6 \cdot n_c}{BpS_1} \right\rceil \end{aligned} \tag{58}$$

The overhead introduced by the initial ranging and the bandwidth requests can be evaluated respectively as:

$$L_{RNG} = \left\lceil \frac{N_{RNG} \cdot \left(OH_{RNG-REQ} + OH_{MAC\ PDU} \right)}{BpS_1} \right\rceil \tag{59}$$

$$\boldsymbol{L_{BW-REQ} = \left\lceil \frac{N_{BW-REQ} \cdot \left(OH_{BW-REQ} + OH_{MAC\ PDU} \right)}{BpS_1} \right\rceil} \tag{60}$$

The overhead $OH_{MAC\ PDU}$ = 10 byte takes into account the 6 byte MAC header and 4 byte CRC. N_{RNG} and $N_{BW\text{-}REQ}$ are the number of contention based opportunities in the initial ranging and bandwidth request fields. $OH_{RNG\text{-}REQ}$ is 12 byte and $OH_{BW\text{-}REQ}$ is 0 byte since the BW-REQ message consists of only a MAC header and CRC.

7.1.5.2 Protocol Efficiency for Mesh Mode

Similar to the efficiency derivation for the PMP mode, the calculation starts with the number of MAC PDUs in one frame.

$$N^{k,m}_{MAC\ PDU} = \left\lfloor \left(\sum_{i=1}^{n_c} PHY\ BURST_i - n_c \cdot L_{SP} \right) \cdot \frac{L_{Minislot}}{k / BpS_m} \right\rfloor \tag{61}$$

where n_c is the number of active connections in the wireless network, k denotes the MAC PDU length, BpS_m (m=1, 2, 3, 4, 5, 6) represents the chosen PHY mode according to the Table 7-1, and L_{SP} corresponds to the length of the short preamble. Besides the long preamble the standard foresees that neighbouring stations can negotiate also the usage of the short preamble, which has been assumed in this case. The MAC PDU length, comprising a header of 6 byte, a Mesh sub-header of 2 byte and CRC of 4 byte, and is in the range of 13 and 2051 byte, assuming a minimum payload of 1 byte.

$$L^k_{MAC\ PDU} = L^k_{Payload} + 12 = k, \quad k \in [13\ ;\ 2051\ \text{byte}] \tag{62}$$

In Mesh mode every resource allocation is specified as multiple of one minislot size $L_{Minislot}$, expressed as number of OFDM symbols [82], [85].

$$L_{Minislot} = \left\lceil \frac{L_{DS}}{256} \right\rceil \tag{63}$$

The number of available OFDM symbols L_{DS} for data transmissions in the Data sub-frame is equal to:

$$L_{DS} = N_{symbols} - L_{CS} \tag{64}$$

where N_{symbol} is the number of OFDM symbols in a frame.

Since in Mesh mode the frame is composed of the Control and the Data sub-frame the number of MAC PDUs, $N^{k,m}_{MAC\ PDU}$, that are transmitted within the Data sub-frame is affected by the duration of the Control sub-frame.

$$\boldsymbol{L_{CS} = 7 \cdot MSH\_CTRL\_LEN} \tag{65}$$

The number of OFDM symbols reserved for the Control sub-frame is defined by the network parameter MSH_CTRL_LEN. The calculation of this parameter is different whether the centralized or the distributed scheduling is adopted.

If the scheduling is performed in centralized manner:

$$MSH\_CTRL\_LEN = \frac{L_{MSH-CSCF} + L_{MSH-CSCH} \cdot \#SS^*}{7} \tag{66}$$

otherwise if it is performed in distributed way:

$$\boldsymbol{MSH\_CTRL\_LEN = \frac{L_{MSH-DSCH} \cdot \#SS^*}{7}} \tag{67}$$

where $L_{MSH\text{-}CSCF/\text{-}CSCH}$ is the overhead introduced by the MSH-CSCF/-CSCH message, $L_{MSH\text{-}DSCH}$ the overhead due to the MSH-DSCH message and SS* the number of stations which transmit control messages. It is worth to be noted here that it is not necessary to transmit the MSH-CSCF message in every frame, reducing the overhead for CS.

Efficiency on MAC layer for Mesh CS

The overhead introduced by MSH-CSCF and MSH-CSCH messages, expressed as number of OFDM symbols, can be evaluated:

$$L_{MSH-CSCF/-CSCH} = 7 \cdot \left\lceil \frac{OH_{MSH-CSCF/-CSCH}}{4 \cdot BpS_1 - OH_{MAC\ PDU}} \right\rceil \quad (68)$$

with the overhead for the MSH-CSCF in byte:

$$OH_{MSH-CSCF} = 3 + \left\lceil \frac{N_{CH}}{2} \right\rceil + 2 \cdot \sum_{i=1}^{N_{NODE}} N^i_{CHILD} + 3 \cdot N_{NODE} \quad (69)$$

and for the MSH-CSCH:

$$OH_{MSH-CSCH} = 4 + N_{FLOW} \quad (70)$$

The overhead $OH_{MAC\ PDU} = 12\ byte$ in the denominator takes into account the MAC header, the Mesh sub-header and CRC, and the MSH-CSCF/-CSCH message comprises 4 OFDM symbols, which are encoded with the most robust PHY mode BpS_1. N_{CH} is the number of available channels in the band assigned to the Mesh BS. N_{FLOW} is the maximum number of links that can be established. This value is strictly related to the network topology. Let N^i_{CHILD} be the number of neighbours with a distance from the Mesh BS (expressed as number of hops) one higher than the distance from the SS_i and the Mesh BS, N_{FLOW} is equal to:

$$N_{FLOW} = \sum_{i=1}^{N_{NODE}} N^i_{CHILD} \quad (71)$$

N_{NODE} is the number of SSs in the wireless network, under control of the Mesh BS.

Efficiency on MAC layer for Mesh DS

The overhead introduced by MSH-DSCH messages, expressed as number of OFDM symbols, can be evaluated:

$$L_{MSH-DSCH} = 7 \cdot \left\lceil \frac{OH_{MSH-DSCH}}{4 \cdot BpS_1 - OH_{MAC\ PDU}} \right\rceil \quad (72)$$

With

$$\begin{aligned} OH_{MSH-DSCH} &= 6 + 3 \cdot N_{SCHED} + 2 \cdot N_{REQUEST} \\ &+ 4 \cdot N_{AVAILABILITY} + 5 \cdot N_{GRANT} \end{aligned} \quad (73)$$

where N_{SCHED} is the number neighbours of which the distributed scheduling information is carried in the message, and $N_{REQUEST}$, $N_{AVAILABILITY}$ and N_{GRANT} are respectively the number of allocation requests, availabilities and grants reported in the message.

7.2 Routing

7.2.1 Relay Locations

7.2.1.1 Analytical Development of Optimal Relay Locations

We observe that the spectral efficiency for a given link has an approximate linear dependency on the SNR measured in dB, as shown in Figure 7-5 (plotted for bit-interleaved coded modulation [132]).

We can approximate the spectral efficiencies a (for the direct link) and x_i (for hop i in the multihop link) as

$$a \cong K_1 \log_{10} \gamma \quad (74)$$

$$x_i \cong K_1 \log_{10} \gamma_i \quad (75)$$

where γ and γ_i are the SNR for the link S-R and link i, respectively, and K_1 is a constant of proportionality. The expressions (74) and (75) remain valid for other digital modulation schemes as long as the linear dependency as in Figure 7-5 is preserved (straight line passing through origin).

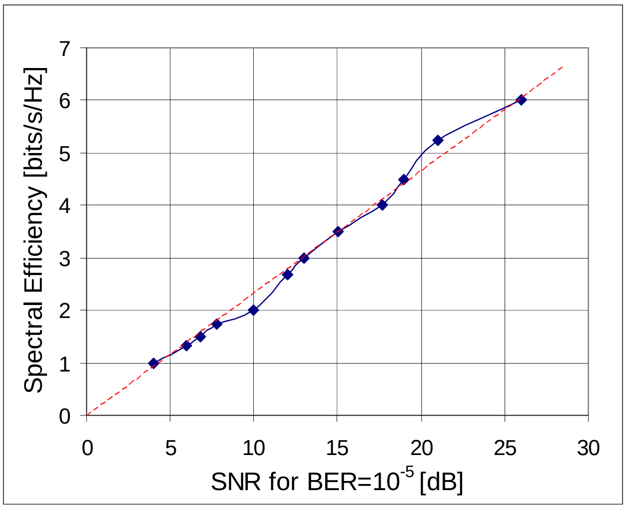


Figure 7-5: Spectral Efficiency vs. SNR.

Assuming that all transceivers at S, R, and the n-1 fixed relays are identical (in terms of transmit power, transmit and receive antenna gains, and receiver noise figure), we can express the mean SNR (excluding shadowing) for the single hop and link i in the n-hop case as

$$\gamma = K_2 \left(\frac{D}{d_0}\right)^{-p} \tag{76}$$

.

$$\gamma_i = K_2 \left(\frac{d_i}{d_0}\right)^{-p} \tag{77}$$

respectively, where:

D = distance S-R

d_i = the length of the hop i

d_0 = the reference close-in distance in radio propagation

K_2 = a constant of proportionality

p = the path loss exponent

In general the constant K_2 captures the radio link system gains, or the link costs in terms of radio resources:

$$K_2 = \frac{P_T G_T G_R \lambda^2}{(4\pi d_0)^2 P_N} \tag{78}$$

The total time required to pass the message over the multi-hop link can then be expressed as

$$\begin{aligned} T &= \frac{M}{BK_1} \sum_{i=1}^{n} \frac{1}{\log_{10} \gamma_i} \\ &= \frac{M}{BK_1} \sum_{i=1}^{n} \frac{1}{\log_{10}\left(K_2 \left(\frac{d_i}{d_0}\right)^{-p}\right)}. \end{aligned} \tag{79}$$

Improving the aggregate end-to-end spectral efficiency means reducing the message transfer time.

To minimize the sum in (79) we consider the function

$$f(d_i) = \frac{1}{\log_{10}\left(K_2\left(\frac{d_i}{d_0}\right)^{-p}\right)} \tag{80}$$

The function *f(di)* is strictly convex for the interval

$$\frac{K_2^{\frac{1}{p}}}{e^2} < \frac{d_i}{d_0} < K_2^{\frac{1}{p}} \tag{81}$$

and strictly concave for

$$\frac{d_i}{d_0} < \frac{K_2^{\frac{1}{p}}}{e^2} \tag{82}$$

for i = 1, 2, …, n.

The convexity and concavity intervals as in (81) & (82) can also be expressed in SNR terms as in (83) and (84), respectively:

$$1 < \gamma_i < e^{2p} \tag{83}$$

$$\gamma_i > e^{2p} \tag{84}$$

For the interval where *f(di)* is strictly convex, applying Jensen's inequality [133], we obtain

$$\sum_{i=1}^{n} \frac{1}{\log_{10}\left(K_2\left(\frac{d_i}{d_0}\right)^{-p}\right)} \geq \frac{n}{\log_{10}\left(K_2\left(\frac{\sum_{i=1}^{n} d_i}{n d_0}\right)^{-p}\right)} \tag{85}$$

with equality if

$$d_1 = d_2 = \ldots = d_n \tag{86}$$

On the other side, if $f(d_i)$ is strictly concave,

$$\sum_{i=1}^{n} \frac{1}{\log_{10}\left(K_2\left(\frac{d_i}{d_0}\right)^{-p}\right)} \leq \frac{n}{\log_{10}\left(K_2\left(\frac{\sum_{i=1}^{n} d_i}{n d_0}\right)^{-p}\right)} \tag{87}$$

also with equality if (86) is true.

7.2.1.2 Analysis of Relay Location in Two-Hop Links

To determine in which conditions a relay located at equal distances from source and destination has a better performance, we consider the difference between message transfer time for a two-hop link with hop lengths (d_1, d_2) and the message transfer time for an imaginary comparison two-hop link with hop lengths ($(d_1 + d_2) / 2$, $(d_1 + d_2) / 2$). The sign of the difference – positive or negative - indicates in which cases the "relay in the middle" scenario is optimal.

The relay's geometric possible locations for $d_1, d_2 \leq D$ shown in Figure 7-6, are within the intersection of two circles of radius D and centers S and R, respectively.

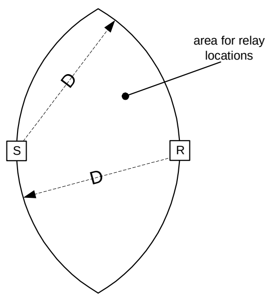
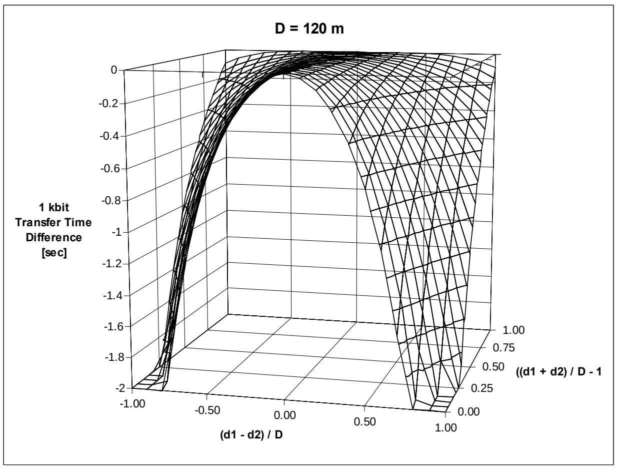


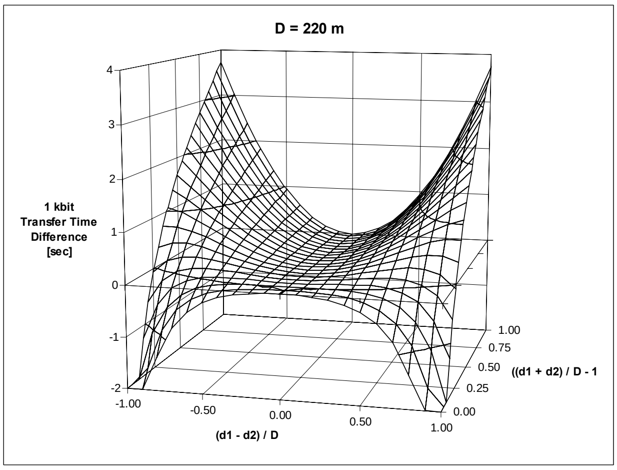
Figure 7-6: Geometric possible locations for relays in a two-hop link

Assuming the system parameters P_T = *30 dBm*, G_T = *10 dB*, G_R = *10 dB*, P_n = *-95 dBm*, d_0 = *2 m*, λ = *0.06 m*, an intermediate hop is "short" if

$$d_i < d_0 \frac{K_2^{\frac{1}{\alpha p}}}{e^2} \approx 119.6m \tag{88}$$

Since $d_1 + d_2 \geq D$, in Figure 7-7, for graphical reasons we have plotted the time difference against the changed and normalized variables $(d_1 + d_2) / D - 1$, $(d_1 - d_2) / D$. Various surfaces are plotted for different values of the distance SR *D*, which determines the 2-hop link SR to have "short" intermediate hops or not. For distances SR smaller than the value in (88) of about *120 m*, both intermediate hops are always "short" and the message transfer time is largest when the relay is placed in the middle of the segment SR. We can see from the plots that the best relay location in this case is as close as possible to either S or R, which seems to suggest that for "short" intermediate hops the single-hop link outperforms the 2-hop link (within the framework of the initial assumptions, the 2-hop link with the relay placed at one end is equivalent with a single-hop link). For larger distances SR, for example *D = 220 m*, the situation changes as follows: if the relay is located close to the straight line SR (where $d_1 + d_2 = D$), the two hop lengths are "short", and the best relay locations are close to S or R; however for relays located farther of the SR line, the best relay location is at equal distances from S and R. For even larger distances SR, the middle of the segment SR is optimal location for the relay (with the exceptions of locations very close to either S or R), and the use of the relay improves the overall performance compared with the single-hop.





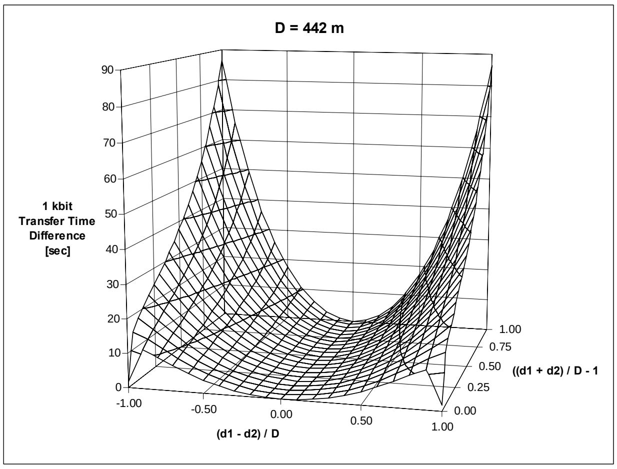


Figure 7-7: Difference between message transfer times of a two-hop link and the link with the two hops of length $\overline{d}^{\,i}$.

To find out the threshold value when the 2-hop link becomes more efficient than the single-hop SR, we set the condition that the message transfer time is the same for the single-hop and for the 2-hop link with the relay placed in the middle of the segment SR:

$$\frac{1}{\log_{10}\left(K_2\left(\frac{D}{d_0}\right)^{-p}\right)} = \frac{2}{\log_{10}\left(K_2\left(\frac{D}{2d_0}\right)^{-p}\right)} \tag{89}$$

Simplifying (89) we obtain

$$\frac{D}{d_0} = \frac{1}{2}K_2^{\frac{1}{p}} \tag{90}$$

or, $D = 442\ m$ for the same systems parameters as before. As seen in Figure 7-7, for SR distances $D = 442\ m$ and larger, the optimal relay location is the middle of the segment SR. In SNR terms (90) can be expressed as:

$$\gamma = 2^p \tag{91}$$

We can conclude that, for the link SR, if the signal to noise ratio of the single hop is larger than 2^p, any 2-hop link would have a lower performance than the single hop. If the signal to noise ratio is below 2^p, the best aggregate spectral efficiency is achieved by the 2-hop link with the relay placed in the middle of SR.

7.2.1.3 Information-Theoretic Analysis of Spectral Efficiency

The Shannon capacity, C_S, over an additive white Gaussian noise (AWGN) channel is

$$C_S = W \log_2\left(1 + \frac{P_{rx}}{N}\right) \quad (92)$$

with the received signal power, P_{rx} , and average noise power, N at the receiver. On each link of an MH connection with n hops the data rate has to be n times higher than the data rate of the corresponding one-hop connection to end-up in the same end-to-end throughput. The required SNR for the MH connection becomes

$$\left(\frac{P_{rx}}{N}\right)_{\text{MH}} = \left(1 + \left(\frac{P_{rx}}{N}\right)_{\text{one}}\right)^n - 1 \quad (93)$$

On the one hand, the required transmit power has to be increased to achieve the same data rate, but on the other hand the distance on the individual links decreases with an increasing number of hops. To incorporate this dependency in the capacity calculation, the receive power is substituted by the transmit power, P_{tx}, and the distance between source and destination, d, based on a simple one-slope pathloss model

$$P_{rx} = P_{tx} + c_0 - 10\gamma \log(d) \quad (94)$$

with a distance-power gradient (attenuation exponent), γ, and a constant c_0 taking into account the attenuation at 1 m. With this substitution in formula (93) the Transmit-power-to-Noise Ratio (TNR) for the MH connection as function of the TNR of the corresponding one-hop connection becomes

$$\left(\frac{P_{tx}}{N} c_0\right)_{\text{MH}} = \left[\left(1 + \left(\frac{P_{tx}}{N} c_0\right)_{\text{one}} \left(\frac{1}{d}\right)^\gamma\right)^n - 1\right] \cdot \left(\frac{d}{n}\right)^\gamma \quad (95)$$

Shannon Capacity with Frequency Re-use

In the previous derivation it has been assumed that transmissions cannot take place at the same time to preserve orthogonality of the received signals in the time domain. However, concurrent transmissions increase the interference but can be acceptable if the required signal-to-interference-and-noise ratio (SINR) is sufficiently high. Furthermore, different interference suppression or cancellation techniques can be adopted to decrease the interference, respectively increase the SINR. In addition intelligent combining of all transmitted signals at the relays and final destination can increase the received signal power, exploiting the diversity of the radio channel. To consider the frequency reuse to come up with a higher spectral efficiency, the Shannon formula is updated to take into account the interference from simultaneous transmissions. The new parameter that is introduced is the reuse distance, d_{reuse}, measured in meters. If we assume equidistant spacing of relays the reuse distance can take values equal to $i \cdot d/n$, with $i \in [1; n - 1]$. Considering the same end-to-end throughput as for a one-hop connection, the average SINR for multi-hop communication for a four hop connection with frequency reuse becomes

$$\left(\frac{P_{tx}}{N} c_0\right)_{\text{MH}}^{\text{reuse}} = \frac{\left[\left(1 + \left(\frac{P_{tx}}{N} c_0\right)_{\text{one}} \left(\frac{1}{d}\right)^\gamma\right)^{\frac{n}{2}} - 1\right] \cdot \left(\frac{d}{n}\right)^\gamma}{1 - \left[\left(1 + \left(\frac{P_{tx}}{N} c_0\right)_{\text{one}} \left(\frac{1}{d}\right)^\gamma\right)^{\frac{n}{2}} - 1\right] \cdot \left(\frac{1}{k}\right)^\gamma} \quad (96)$$

It is assumed that two stations simultaneously transmit at a reuse distance of two hops, i.e. $d_{reuse} = d/2$, resulting in half of the number of transmission cycles compared to multi-hop communication without frequency reuse. In the equation the separation of simultaneous transmission is incorporated in the variable k. If no interference cancellation is introduced $k = 2$. The more interference is suppressed the larger k becomes, unless the whole denominator in equation (96) becomes 1. In this case the power needed for the four hop connection becomes equal to the power needed for two hops only. Hence, for an n-hop connection with even values n the capacity is doubled since two transmissions can take part at the same time, which is reflected in the exponent n in equation (96) that is divided by 2 compared to equation (95).

7.2.2 Models and Algorithms for Routing/Forwarding in multi-hop wireless networks

7.2.2.1 Routing protocols for mobile and fixed relay nodes

7.2.2.1.1 Dynamic Source Routing (DSR)

The Dynamic Source Routing (DSR) protocol is an on-demand (reactive) routing protocol that is based on the concept of source routing [135], [136]. Mobile nodes are required to maintain route caches that contain the source routes of which the mobile is aware. Entries in the route cache are continually updated as new routes are learned. In DSR different packet types are used for route discovery, forwarding, and route maintenance.

Route discovery:

DSR uses two types of packets for route discovery, Route Request (RREQ) and Route Replay (RREP) packet. When a mobile node has a packet to send to some destination, it first consults its route cache to determine whether it already has a route to the destination. If the node does not have an unexpired route to the destination, it initiates route discovery by broadcasting RREQ packet. This route request contains the address of the destination, along with the source node's address and a unique identification number. Each node receiving the packet checks whether it knows of a route to the destination. If it does not, it adds its own address to the route record of the packet and then forwards the packet along its outgoing links. To limit the number of route requests propagated on the outgoing links of a node, a mobile only forwards the route request if the request has not yet been seen by the mobile and if the mobile's address does not already appear in the route record.

A RREP is generated when either the route request reaches the destination itself, or when it reaches an intermediate node, which contains in its route cache an unexpired route to the destination. By the time the packet reaches either the destination or such an intermediate node, it contains a route record yielding the sequence of hops taken and the respective addresses of the intermediate nodes.

If the node generating the route reply is the destination, it places the route record contained in the route request into the route reply. If the responding node is an intermediate node, it will append its cached route to the route record and then it will generate the route reply. To return the route reply, the responding node must have a route to the initiator. If it has a route to the initiator in its route cache, it may use that route. Otherwise, if symmetric links are supported, the node may reverse the route in the route record. If symmetric links are not supported, the node may initiate its own route discovery and piggyback the route reply on the new route request.

Care must be taken to avoid collisions between route requests propagated by neighbouring nodes (insertion of random delays before forwarding RREQ). To avoid multiple host replying simultaneously from their cache, each host delays its reply slightly. Time of delay depends on distance in hops from RREQ source. Also reply storm may be eased by preventing a node from sending RREP if it hears another RREP with a shorter route.

Route maintenance:

Route maintenance is accomplished through the use of Route Error (RERR) packets. RERR packets are generated at a node, when the data link layer encounters a fatal transmission problem. When a route error packet is received, the hop in error is removed from the node's route cache and all routes containing the hop are truncated at that point.

Improvements of the protocol are possible, when a relay node that forwards a data packet observes the entire route in the packet and updates its route cache. Closely related to this approach, the routing table can also be updated by nodes that work in promiscuous mode and overhear unicast packets being sent to other nodes, comprising route information.

The main advantage of DSR is that routes are maintained only between nodes which need to communicate. This reduces overhead of route maintenance. Furthermore, route caching can further reduce route discovery overhead. A single route discovery may yield many routes to the destination, due to intermediate nodes replying from local caches.

The main disadvantages are that the packet header size grows with the route length due to source routing. So the protocol has a major scalability problem due to the nature of source routing.

7.2.2.1.2 Ad-Hoc On-Demand Distance Vector (AODV)

One of the most promising approaches is the Ad-Hoc On-Demand Distance Vector (AODV) routing algorithm, which already became an IETF draft [134]. The AODV routing protocol is a reactive routing protocol intended for use by mobile stations and determines routes to destinations in ad hoc networks. It offers quick adaptation to dynamic link conditions, low processing and memory overhead, and low network resource utilization. AODV is based on the Destination Sequence Distance Vector (DSDV) routing algorithm and uses destination sequence numbers to ensure loop freedom. The destination sequence number is created by the destination to be included along with any route information it sends to the requesting station that allows the source to select the most recent and fresh route.

In AODV different packet types are used for route discovery, forwarding, and route maintenance.

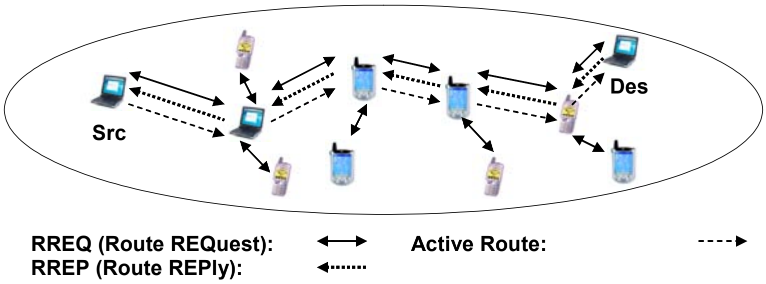


Figure 7-8. Operation of AODV

Route discovery:

AODV uses two basic packet types for route setup, the Route Request (RREQ) and Route Reply (RREP) packet. During route-discovery when a route to a new destination is needed the station broadcasts an RREQ packet, see Figure 7-8. Each station receiving the request caches a route back to the originator of the request, so that the RREP packet can be transmitted via unicast from the destination along a path to that originator. The packet is re-broadcast by every station as far as this packet has not been received before, which can be checked by means of its unique packet ID.

To decrease the overhead during flooding, a hop-count is introduced that limits the forwarding. With expanded ring search the hop-count is increased when no responds to a request is received after a pre-defined time-out, resulting finally in flooding the whole network with RREQ packets.

A route is determined when the RREQ reaches a station that knows a valid route to the destination (e.g., the destination itself). This route is made available by transmitting a RREP back to the originator of the RREQ, see Figure 7-8.

Forwarding:

Due to reverse-path setup based on caching the RREP packet, every station on the route knows the next-hop neighbour they should address to forward a packet to the destination. AODV provides a routing table with information for all known routes. An entry in the routing table contains the destination, the next neighbour to forward a packet, and a metric, e.g. the hop-count, for that route. As shown in Figure 7-8, a packet that contains the source and destination address can be forwarded along the path that the RREP packet traversed before with the help of the routing table without further information.

Route Maintenance:

When a link break in an active route is detected, the broken link is invalidated and a Route Error (RERR) message is typically transmitted back to the source to notify other nodes that the loss of that link has occurred. To detect a link break each station periodically transmits HELLO packets with their ID. Updating the neighbour table in each station upon receiving a HELLO packet a link is defined as not available after a respective timer has exceeded for that neighbour.

The station will setup a new route by means of the mechanisms described above. This results in large delays and overhead, especially when a long route breaks near to the destination. Hence, a local route-repair is defined in AODV. After a station has recognized a link break it tries to recover the route by means of transmitting an RREQ packet like it is defined for initial route discovery. If this attempt fails it will send a RERR back to the source, otherwise it receives a RREP with a fresh route to the destination.

It is not guaranteed that the station that locally recovers a broken link succeeds, resulting in undesirable delay and wasting bandwidth, since the source and intermediate stations continue transmitting packets to that station. Therefore, local route-repair should be favored the closer the link break happens to the destination, whereby closer is defined by the routing metric, e.g., the number of hops or the geographical distance. Typically, local route-repair is initiated if the link break happens closer to the destination than to the source.

The main advantages of AODV are that routes are maintained only between nodes that have to communicate, similar to DSR, and that the route information is locally stored in each intermediate relay based on reverse path setup procedure, and the whole route information has not to be sent in the packet header, making AODV more scalable than DSR.

However, AODV has some deficiencies, especially in the presence of very dynamic network topologies. If routes break during an active communication, a route-repair process should be initiated, that imposes additional overhead and needs considerable time. For that purpose respective methods for local repair are proposed for AODV. But the proposed solutions seem to be far from optimum, since they need considerable time and transmit resources.

7.2.2.1.3 Greedy Perimeter Stateless Routing (GPSR)

Location-based routing protocols use information about geographical position of nodes in the network. Distinct from topology-based approaches location-based routing comprises only the forwarding step. There is no root establishment nor a root maintenance procedure. Every node chooses the next-hop neighbour or a set of next-hop neighbours according to their mutual position with respect to destination. The only additional information that is required is the position of the destination. Acquiring this information is comparable to the route establishment in topology-based routing schemes. This process is termed location service.

Location Service:

The responsibility of a location service is to make a position of every node available to every other node in the network. The simplest approach to acquire the location of the final destination is to flood the network with a respective request. Thus, every node receives position of every other node in the network. However, pure flooding results in large overhead.

Different to the simple flooding, grid-based location services use grids to make location information available to every node in the network. In distinct from the flooding approach every node stores positions only some set of nodes in the network. An example is the Grid Location Service (GLS) proposed in [137]. It divides the area that contains the ad-hoc network into a hierarchy of squares. In this hierarchy, n-order squares contain exactly four (n-1)-order squares, forming a so called quadtree. Each node maintains a table of all other nodes within the local first-order square. The table is constructed with the help of periodic position broadcasts which are scoped to the area of the first-order square.

Another approach is based on the concept of quorum systems, which is well known from information replication in databases and distributed systems. Information updates (write operations) are sent to a subset (quorum) of available nodes, and information requests (read operations) are referred to a potentially different subset. When these subsets are designed such that their intersection is non-empty, it is ensured that an up-to-date version of the sought-after information can always be found.

Forwarding:

If the position of the final destination is known the packet can be forwarded. The simplest approach is again flooding. However, this is quite inefficient and the efficiency can be improved by means of “restricted flooding”. In this case a source node is using its own position and position of the destination and computes region to be flooded. This approach is, e.g., used by the Distance Routing Effect Algorithm for mobility (DREAM) [138]. The sender of a packet will forward the packet to all one-hop neighbours that lie in the direction of the destination. In order to determine this direction a node calculates the region that is likely to contain the destination, called the expected region. which can be a cirvel around the position of the destination as it is known by the sender. Since this position information may be outdated, the radius of the expected region is set to a fraction of the time the has elapsed since the last position update of the destination and the max. expected velocity a node might move. Given the expected region, the direction is defined by the angle between the lines that cover the target region.

While the restricted directional flooding approach used by DREAM leads to a certain reliability in terms of being able to reach the destination, the price to pay for this reliability is high: flooding wastes bandwidth, which is a very scare resource in mobile ad-hoc networks. While restricting flooding to a certain direction does reduce the overhead, it does not eliminate it.

A more efficient forwarding strategy is “greedy” forwarding. If a node knows the positions of its neighbours, the locally optimal choice with respect to the potentially largest progress of the next hop is the neighbour geographically closest to the destination. It is however worth mentioning that the node closest to the destination might only be reachable with high packet error rate. Hence, a node with less progress being closer to the sender might result in better performance with respect to forwarding efficiency, i.e. forwarding progress per transmission attempt. Forwarding in this regime follows successively closer geographic hops, until the destination is reached.

A simple beaconing algorithm provides all nodes with their neighbours' positions: Periodically, each node transmits a beacon to the broadcast address, containing only its own identifier (e.g., IP address) and position. The great advantage of greedy forwarding is its reliance only on knowledge of the forwarding node's immediate neighbours. The state required is negligible, and dependent on the density of nodes in the wireless network, not the total number of destinations in the network.

Recovering from local maximum in greedy mode:

Unfortunately, greedy routing may fail to find a path between sender and destination, even though one does exist. This is the case when the node to forward the packet is closer to the destination than any of the node in its transmission range. Greedy routing therefore has reached a local maximum from which it cannot recover.

Greedy Perimeter Stateless Routing (GPSR) protocol uses perimeter forwarding as a recovering procedure. In GPSR, a packet enters perimeter routing mode when it arrives at a local maximum. It returns to greedy mode when it reaches a node closer to the destination than the node where the packet entered perimeter routing mode.

Perimeter routing is based on planar graphs, i.e., graphs with no intersecting edges. A set of nodes in an ad hoc network can be considered a graph where the nodes are vertices and an edge exists between two vertices if they are close enough to communicate directly with each other. The graph formed by a network is generally not planar, so special algorithm have to be applied for constructing planar sub-graph. Based on the planar sub-graph, GPSR's perimeter routing performs a simple planar graph traversal to find a path towards the destination. The general concept of perimeter routing is to forward the packet on faces of the planar sub-graph which are progressively closer to the destination. On each face the packet is forwarded along the interior of the face by using the right-hand rule. With the right-hand rule a packet on the next edge if forwarded counter clockwise from the edge on which it arrived. Whenever the line between source and destination intersects the edge along which a packet is about to be forwarded it is checked if this intersection is closer to the destination than any other intersection previously encountered. If this is true, the node switches to the new face bordering on the edge which the packet was about to traverse. The packet is then forwarded on the next edge counter clockwise to the edge it was about to be forwarded before switching faces. This algorithm guarantees that a path is found from the source to the destination if there exists at least one such path in the original non-planar graph.

In GPSR, the packet header contains additional information such as the position of the node where it entered perimeter routing, the position of the last intersection that caused a face change, and the first edge traversed on the current face. Therefore, each node can make all routing decisions based only on the information about its local neighbours. This includes the detection of an unreachable destination, when a packet traverses the first edge on the current face for the second time.

7.2.3 Wireless Multi-hop Routing Problem

The model of the wireless multi-hop routing problem here presented is an extension of the well known multi-commodity flow problem where the network is represented by a graph with a capacity associated with each link and connections are represented as flows that compete for the limited link capacities. The objective is to maximize the fraction of offered traffic admitted in the network. A survey on multi-commodity flow problems can be found in [145]. Differently from wired networks, with wireless multi-hop networks we cannot associate a capacity to each link in the graph, since parallel transmissions on different links may be prevented due to interference and the inability of stations to transmit and receive at the same time. We assume that when a transmission occurs on a link between a couple of nodes, the transmissions of nodes directly connected to the transmitting or the receiving nodes are prevented. Such an assumption assure that the hidden terminal problem does not occur both for the considered transmission and that of the acknowledgement packet, like for IEEE 802.11 systems.

Let us consider a graph $G = (V, E)$ whose N vertexes, $V = \{1, 2, ..., N\}$, represent the wireless stations and the M edges (i, j) connect stations within transmission range.

Consider now a set of K pairs of vertexes (s^k, t^k) for k=1,2 …K representing the source and destination node associated with K commodities for which a path should be found on the graph G. For each commodities k can be introduced the following variables.

$f^k_{i,j}$ = Units of flow of commodity k routed on link from i to j;

F^k = Total units of flow to be sent from s^k to t^k;

$f_{i,j} = \sum_{k=1}^{K} f^k_{i,j}$ = Total units of flow on arc from i to j.

Then for each node n, represented in G by a vertex n, the following sets are defined.

$$A(n) = \{ j \in V | (n,j) \in E \}$$
$$B(n) = \{ j \in V | (j,n) \in E \}$$

Where $A(n)$ represents the set of nodes that can be reached from node n while $B(n)$ is the set of nodes that can reach node n with their transmission. As said, the objective in solving routing problem is to maximize the fraction of total commodities that is routed through the network from source to destination. This can be expressed introducing a parameter α that represents for all commodities the fraction of F^k admitted in the network. Therefore objective is to find the optimum α, denoted as α^*, such that for each one of the given K commodities it is possible to route in the network exactly $\alpha^* \cdot F^K$ units of flow from the source node s^k to destination t^k. The optimum α^* is that particular value of α that satisfy the following objective function:

$$Max\left\{ \sum_{k=1}^{K} \alpha \cdot F^k \right\} \tag{97}$$

Further, conservation equations and non negativity must be satisfied by every flow:

$$\sum_{j \in A(n)} f^k_{n,j} - \sum_{j \in B(n)} f^k_{j,n} = \begin{cases} \alpha \cdot F^k \; if \; s^k = n \\ -\alpha \cdot F^k \; if \; t^k = n \\ 0 \quad otherwise \end{cases} \tag{98}$$

$$f^k_{i,j} \geq 0 \text{ for all } (i,j) \in E \tag{99}$$

$$\alpha \in [0,1] \tag{100}$$

The equation (98) simply states for each commodity k that the difference between incoming flow and outgoing flow on a given node n is positive if node n is the source of the commodity, is negative if node n is the destination and is equal to zero if node n is an intermediate node of k ‘s path. Equation (99) instead states, obviously, that cannot exist negative commodities in the network.

To model the capacity and interference constraints, new sets of nodes are introduced: the set S^1_i made up by all links that are adjacent to node i:

$$S^1_i = \{(i,j), j \in V\} \tag{101}$$

and for each node $j \in A(i)$ (e.g. nodes u,v,z in Figure 7-10), the set $S^2_{i,j}$ including all links that have one of their end in $A(i)$ and meanwhile do not belong to S^1_i:

$$S^2_{i,j} = \{(j,k) \mid k \neq i\} \tag{102}$$

Obviously for the given node i, exactly $|A(i)|$ set of this type exist.

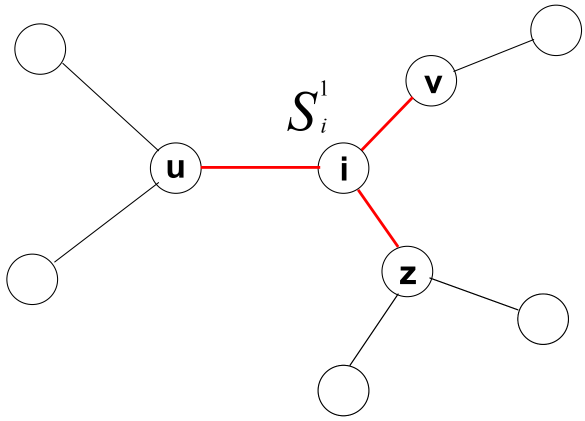


Figure 7-9: One hop contraints of node i

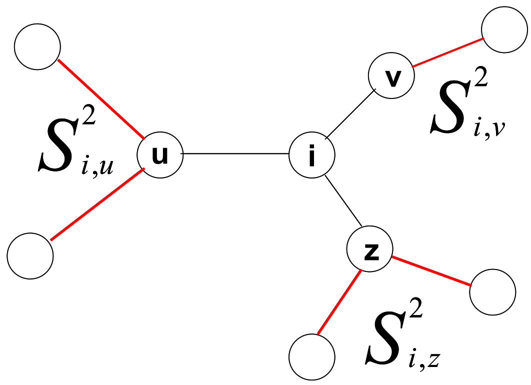


Figure 7-10: two hops contraints of node i

Starting from these sets, it is possible to build up the set G_i formed by all relays within two hops from relay i, this set can be expressed as:

$$G_i = \left\{ S_i^1 \cup \bigcup_{j \in A(i)} S_{i,j}^2 \right\} \tag{103}$$

It is possible to characterized each set G_i by a theoretical capacity C_i that represents the maximum aggregate flow that can be routed over the whole set of links belonging to that group. All flows routed on links belonging to set S_i^1 contributes in consuming common resources associated to the whole group G_i and to bound possibilities for new transmissions/receptions of relay i. About the various set $S_{i,j}^2$ build on each neighbour of node i, only the heaviest loaded one should be considered for writing the new QoS constraint, since it is the one that gives the more restrictive condition about G_i's resources consumption and about limitation for further transmissions/receptions of node i. It is hence possible to write the constraint to be inserted in the mathematical model as follow:

$$\sum_{k=1}^{K} \sum_{(i,j) \in S_i^1} f_{i,j}^k + \max_{j \in A(n_i)} \left\{ \sum_{k=1}^{K} \sum_{(j,l) \in S_{i,j}^2} f_{j,l}^k \right\} \le C_i \, \forall G_i \tag{104}$$

It is possible to transpose the non linear constraint (104) to a set of linear constraints without any approximation obtaining in this way a linear model for our problem. This is done eliminating the max operator that gives the non linearity and extending the constraint to all set $S_{i,j}^2$ related to relay i comprising in this way also the heaviest loaded one that gives the more restrictive condition. This set can be compacted in the following expression:

$$\sum_{k=1}^{K}\sum_{(i,j)\in S_i^1} f_{i,j}^{k} + \sum_{k=1}^{K}\sum_{(j,l)\in S_{i,j}^2} f_{j,l}^{k} \le C_i \, \forall S_{i,j}^2 \in G_i, \forall G_i \tag{105}$$

It is now possible to write the mathematical model used to describe problem of QoS routing in a network composed by fixed/movable relays equipped with omni-directional antennas and a unique radio interface:

$$Max\left\{\sum_{k=1}^{K}\alpha\cdot F^{k}\right\} \tag{106}$$

so that

$$\sum_{j\in A(n)} f_{n,j}^{k} - \sum_{j\in B(n)} f_{j,n}^{k} = \begin{cases} \alpha\cdot F^{k} \; if \; s^{k}=n \\ -\alpha\cdot F^{k} \; if \; t^{k}=n \\ 0 \quad otherwise \end{cases} \tag{107}$$

$$f_{i,j}^{k} \ge 0 \text{ for all } (i,j)\in E \tag{108}$$

$$\alpha \in [0,1] \tag{109}$$

$$\sum_{k=1}^{K}\sum_{(i,j)\in S_i^1} f_{i,j}^{k} + \sum_{k=1}^{K}\sum_{(j,l)\in S_{i,j}^2} f_{j,l}^{k} \le C_i \, \forall S_{i,j}^2 \in G_i, \forall G_i \tag{110}$$

7.2.4 The generalized Dijkstra's algorithm

This pargraph refers to the multi-constrained Quality of Service routing algorithm issue (see 4.2.5.5.2) and contains the detailed description of the proposed generalized version of the Dijkstra's algorithm. For the purposes of unambiguous identification of the nodes that are used repeatedly, the algorithm uses a function *node:* $N \rightarrow V$. The aim of this function is to assign the next natural number to each node that was accepted as an element of the path leading from *s* to *t*. In consequence many numbers may be associated with the nodes that are elements of distinct paths. The generalized Dijkstra's algorithm is presented in Figure 7-11. The following notation is used:

- *V* – the set of nodes,
- *V** – the extended set of nodes,
- *X* – the subset of *V**, used for storing information about the nodes that will be used for building the set of the *k* shortest paths,
- $count_i$ – the number of paths leading from the source node *s* to the node *i*,
- *elm* – the number of elements in the set *V**, whereas the *{elm}* is the element of index *elm*,
- d_i – the distance of the *i-th* element (representing the *node(i)*) from the source node *s*,
- π_i – the predecessor (father) of the node *i* on the path from the source node *s*,
- P_k – the number of the shortest paths between nodes *s* and *t*.

```
for ∀i ∈ V − {s} do
    count_i ← 0
end for
count_s ← −1
elm ← 1
h(elm) ← s
d_elm ← 0
X ← {elm}
h^{-1}(s) ← {elm}
P_k ← ∅
while (count_t < k) and (X ≠ ∅) do
    l ← element of X | d_l ≤ d_x ∀x ∈ X
    X ← X − {l}
    i ← h(l)
    count_i ← count_i + 1
    if (i = t) then
        p ← path from 1 to l
        P_k ← P_k ∪ {h(p)}
    end if
    if (count_i ≤ k) then
        for ∀arc(i, j) ∈ E do
            elm ← elm + 1
            π_elm ← l
            h(elm) ← j
            d_elm ← d_l + w(i, j)
            X ← X ∪ {elm}
            h^{-1}(j) ← h^{-1}(j) ∪ {elm}
        end for
    end if
end while
```

Figure 7-11: The generalized Dijkstra's algorithm

7.3 Cooperative Relaying

7.3.1 Introductory Example

Figure 7-12 depicts the most simple cooperative relay network. Two mobile stations cooperate with each other in a symmetric fashion in order to communicate to third node. In a cellular or any other infrastructure-based network, this third node could be a base station, while in an ad-hoc context, the third node would just be the common destination node or the next hop for the two source nodes.

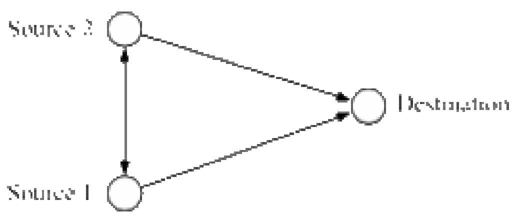


Figure 7-12: A simple cooperative relaying scenario with two source nodes that cooperate to communicate to a common destination node.

In a conventional, direct system, both source nodes would transmit independently and directly to the destination. Assuming a TDMA protocol with K symbol periods per slot, node 1 would transmit in times *[0,...,K]* to the destination node, which is followed by node 2's transmission in the second slot *[K+1,...,2K]* as illustrated in Figure 7-13.

In a cooperative relaying scenario, the two nodes may decide to relay parts or all of their information via the other node in order to take advantage of the spatial diversity. This cooperative transmission is restricted by the so-called *orthogonality constraint*: a terminal cannot transmit and receive simultaneously at the same frequency. This requires the stations to receive and transmit at orthogonal subchannels, which could in practice be achieved by splitting the available resources in the frequency or time domain. While noting that for theoretic investigations it would suffice to consider *orthogonal channels* in general, we illustrate the example by assuming *time-division*.

To allow for a fair comparison, we demand that the total time for sending data is equal for direct systems and our exemplary cooperative scheme. Thus, a *simple static cooperative relaying protocol* that yields diversity for both sources splits the two frames of total duration *2K* symbols into *four* sub-slots of duration *K/2*. Consider Figure 7-13. In the first sub-slot *[0, ..., K/2]*, source 1 broadcasts its information to both source 2 and the destination. Then, in the second sub-slot *[K/2+1, ..., K]*, source 2 relays the signal it has received from source 1 to the destination. In the remaining two sub-slots, source 1 and source 2 switch their roles. Now, source 2 first broadcasts its information to source 1 and the destination during *[K+1, ..., 3K/2]*, and subsequently source 1 assists source 2 by sending the relayed version during *[3K/2+1, ..., 2K]*.

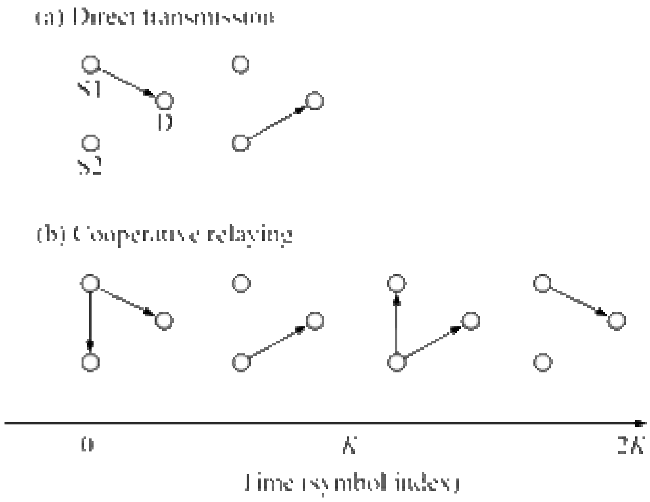


Figure 7-13: A simple cooperative relaying scheme. Two nodes may either send consecutively and directly to the destination (top), or the slots are subdivided to allow for cooperation in time-division manner (bottom).

It is obvious that this scheme suffers from the shorter time available for data transmission: while for direct transmission *K* symbol periods are allocated for a message, the same amount has to be transmitted in *K/2* symbol periods in this *repetition-coded* cooperative relay scheme, thereby calling for a double spectral efficiency on each of the individual links. Put equivalently, the capacity of the relay scheme is half that of direct transmission.

The simple example clearly demonstrates the fundamental tradeoff of cooperation and relay protocols: the diversity gains are challenged by the constraints imposed by orthogonality restrictions and the repetition-coded nature of the protocols. We will see that this issue arises at both physical and system level.

7.3.2 Possible Parameters and Assumptions for Cooperative Relaying for Mobile Relays

The following list – without being exhaustive – indicates possible parameters and areas of work on mobility aspects of cooperative relaying:

- **Types of relays→** We will assume both user terminals and pure relays to be deployed, but not at the same time. They will be evaluated in separate cases. However, some priority might be given to pure relays. At a further stage, it might be possible to simultaneously consider them in one simulation scenario.
- **Scenarios→** Cooperative relaying is more attractive for urban cases of relatively low/medium mobility scenarios where due to the plethora of relays we have more chances of selecting favourable relays compared to rural areas (e.g. train scenario) where we expect to have very few mobile relays of possibly high mobility. Thus, initially we will assume urban/suburban scenarios.
- **Number of hops→** We are still dealing with a cellular environment and allowing many hops (like AdHoc networks) may complicate things. The higher the number of hops the more potential gain we might have (although not linearly), but provided that complexity is maintained at low levels. Again, due to the (sometimes high) mobility, a large number of hops might induce large delays in a

frequently changing environment. We feel that a maximum of 3 hops should be targeted, although in our initial evaluations we will assume a two-hop strategy. To further extend work on more than 3-hop approaches, we need to see what the implications and the incremental gain is. We need to investigate the trade off between complexity and gain. The criteria for choosing also 2-3 or more hops should also be investigated along with the delays introduced by the relays, related to their impact on QoS for certain applications/traffic e.g. voice communication.

- **Symmetry** →Asymmetric: The initial assumption of cooperative relaying follows the conventional cellular environment (hierarchical structure), where a user terminal will be able to "listen" to a number of relay(s) and/or the AP. Thus, we will assume a "asymmetric" approach and not the case of two elements (relays) exchanging roles of AP and relay. Additionally, the gain from the path loss reduction is much more in the asymmetric case. Signalling might also be required between the two relays (either direct or through a BS) for the symmetric case, which might complicate even more the processes.
- **Static protocols** → Static protocols, as explained in the relevant section, are attractive when an environment is either static or very slowly changing, with a small standard deviation. The gain is the reduced complexity in terms of algorithms and also the "zero" signalling. However, in dynamic systems, as the one that mobile relays form, having static protocols may not be such an attractive approach. Conditions change all the time e.g. channel condition, position of relays and it would be much more interesting (although complex) to investigate cooperative adaptive algorithms (either based on intra-relay measurements or from feedback signalling from UEs or the AP). Protocols which can adapt to fast or slow changing conditions should be in place to cope with new needs or different RRM strategies. The typical example is the case of propagation conditions of paths in a relay-based environment which change often. Of course, the price to pay is signalling overhead, delays and complexity in algorithms. However, we have to make sure those are minimised by using e.g. broadcast and not dedicated signalling.
- **Forwarding strategies**→Simple Amplify-and-Forward ("repeater, non-regenerative relaying") strategies have already been investigated, to some extend in other fora e.g. 3GPP for repeaters [95][96][97] However, limitations exist. More gain from relays can be extracted from more complicated approaches. An interesting approach would be the Decode-and-Forward strategy ("bridge, router, regenerative relaying"). Performance could be monitored based on the different types of layer 1/2 encoding/decoding strategies (FEC, ARQ, etc). Additional requirements could also be studied e.g. feedback signalling for ARQ schemes.
- **Combining**→Either Maximum ratio combining or Selection combining (SC) can be applied. Selection combining can be employed if the relay decides to decode or not (either by itself or based on some feedback).
- **Coding strategy**→ Repetition-based ("relay repeats") or incremental redundancy ("relay codes") Different schemes can be employed here. For the time being the initial assumption could be just repetition based. Further, we could consider more complicated schemes e.g. incremental redundancy ARQ schemes.
- **Diversity available/Type of receivers**→ We will assume one antenna per user equipment and per relay although, for the sake of completion, some results related to 2-3 antennas (at the UE) may be included. This issue will be directly affected by the MA technique selected.
- **Power**→ A very important issue for cooperative mobile relaying is power control/allocation. The issues of Tx power from mobile relays and the interference that might be induced considering multiple links from multiple relays for cooperative relaying should be considered. The initial approach will be to assume that the transmitted power of the relay is a percentage of the Tx power of the AP. To be strictly correct for reasons of comparison the total power levels for the conventional and relaying topologies should be the same. However, in a realistic scenario that might not be applicable in the sense that by reducing the Tx power of the AP we limit coverage in areas where we do not intend to have relays.
- **Interference** → Due to the large number of relays and possible links, as stated in the previous section, interference might be induced in a cell. Thus, we should investigate what is the interference rise for a number of cooperative links. Moreover, cooperative mobile relaying should not only be seen from a UE point of view, but also its side effects at a system level should be investigated. What we should measure is the interference from multiple links in one cell, from multiple UEs within one cell and from the surrounding cells. Figure 7-14 portrays that. System level simulations need to be prepared. The initial assumption will be to evaluate the interference on a cell level. At a further stage, we will consider interference from multiple cells. We also need to consider the issues of always new

relays "coming in" the target are and old relays coming out of the target area, which will further complicate our algorithms.

- **Rooting algorithms** → Finding the best branch to route packets to a UE should be a relative straightforward technique in fixed-relay deployments. However, in the case of cooperative mobile relaying more information should be taken into account. Additionally this information should be updated regularly. Some of the parameters include the following
  - Location of relay
  - Velocity
  - Trajectory
  - Dynamic Capabilities e.g. power availability

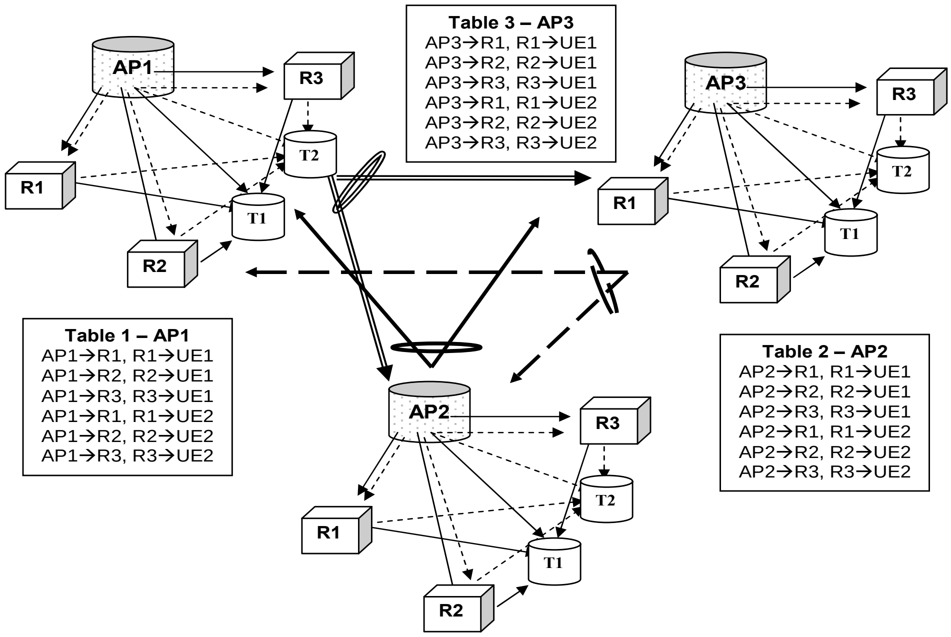


Figure 7-14 Intra-cell and inter-cell interference among 3 cells

- **Channel state information (CSI)**→ As stated above in two occasions, feedback mechanisms might be required either in the relay, in the UE or in the BS. These can help update frequently changing parameters. What would be interesting is to investigate the trade off of signalling and delays to the performance enhancement for certain schemes e.g. ARQ.

7.3.3 Additional Concepts for Cooperative Relaying

In this subsection, we continue the presentation of concepts as started in 4.2.6.2.3.

D- Virtual Antenna Arrays

In a different context, Dohler et al. investigate the possibility of forming *virtual antenna arrays* [167][166][165]. By examining various amplify-and-forward concepts ranging from three-terminal cooperation to the use of multiple groups of relays, each representing a virtual antenna array, they show improvements induced by the resulting diversity gains. However, some of the demonstrated results would deserve clarification of the underlying assumptions, justification of these, and a discussion of the feasibility of the concepts. For example, the implications of the orthogonality constraint are not taken into account, and diversity gains are modelled in a simplistic manner.

E- Large-scale Relaying

Many relays assisting a single source-destination-pair Wittneben and Rankov study large-scale relaying [205], where *many* relays act as *active scatterers* in an amplify-and-forward manner to bring uncorrelated channels and hence full-rank properties to otherwise correlated MIMO scenarios; see Figure 7-15.

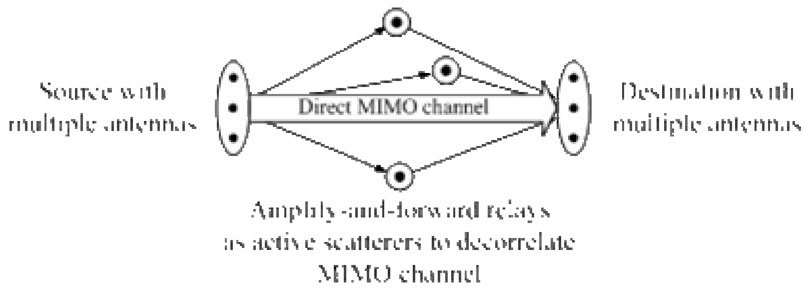


Figure 7-15: Relays as active scatterers to help decorrelating a MIMO channel [205].

Larsson et al. [187] likewise study the case of a large number of amplify-and-forward relays. It is, however, assumed that phase and amplitudes are adjusted by complex multiplication so that coherent superposition occurs at the destination as illustrated in Figure 7-16. The resulting SNR is maximized under a total relay power constraint by determining optimum complex gains c_i at each relay. This leads to significant SNR improvements, for which channel state information and information on the power allocation must be available at the transmitting relays.

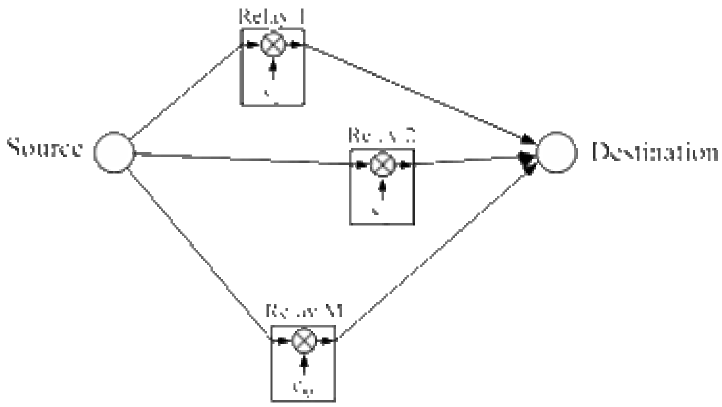


Figure 7-16: Large-scale relaying with complex gains at relays to achieve constructive superposition at the destination [187].

F- Network Aspects

There are few works addressing the potential impact of cooperative relaying at the network level, where multiple source-destination pairs operate. A complete *network* of nodes is considered by Barbarossa et al. [162]. An ODMA scenario is examined, where terminals communicate with the base station in dedicated time slots. The use of *one time slot* of the cell is envisaged for all inter-mobile communication in this cell; see Figure 7-17. Dedicating just a single slot for all cooperating links provides a *"balance between the waste due to sharing and the gain due to the increase of diversity and capacity"* [162]. In other words, the cooperation gains and the limited required rate increase come at the cost of mutual interferences.

The applicability of the results suffers to some extent from the crucial assumptions on error-free peer-to-peer links, negligible inter-user interference, and perfect synchronization. Likewise, channel state information is assumed to be available at transmitters.

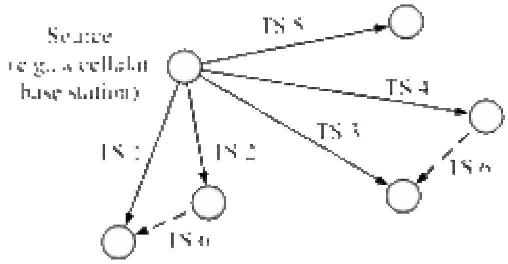


Figure 7-17: Cellular cooperative ODMA scenario (downlink) [162][194]. Transmission from the base station to the mobiles (solid arrows) take place in orthogonal time slots (TS), while all cooperative transmissions among mobiles (dashed arrows) use the same time slot (TS 6 in this example).

The work is extended by Scutari [194] for a cellular ODMA system. Under the assumption that *all* channels and noises between *all* terminals are known *at all terminals*, it is shown by means of a game-theoretic approach that a useful/feasible power allocation requires a minimum distance between cooperating pairs, as otherwise strong interferences severely degrade system performance.

Similar approaches are taken by Munoz, Agustin et al. [189][159], who study a cellular TDMA system (HSDPA) for amplify-and-forward as well as for decode-and-forward systems. Resources for relaying are again provided by a single timeslot per cell, that is shared by many relays.

Transmit powers are assigned by a game-theoretic approach [159], whereby a utility function determines the cases in which relaying outperforms direct transmission. This approach is interesting, as it allows for making decisions on routing and scheduling.

G- Aspects of Cooperative Relaying for Mobile Relays

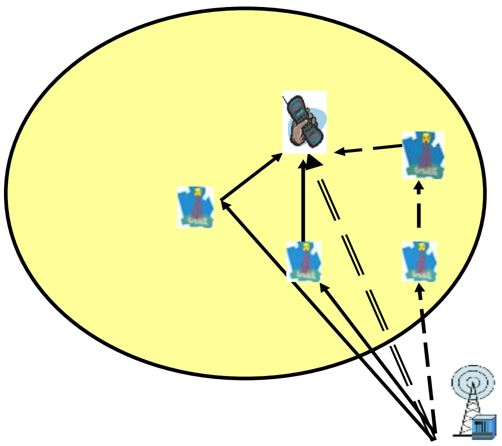


Figure 7-18: Cooperative relaying concept. Relayed and direct paths combined

Due to the complexity of the mobile cooperative relaying schemes and also the different algorithms / techniques we can introduce, we will initially define some relatively simple approaches which will be used to evaluate cooperative relaying for mobile relays. Some of the main issues (and at the same time some assumptions) that we will take into account are summarized in the appendix.

We turn to a discussion of a simple analytical method for evaluating a basic cooperative mobile relay approach. We assume one AP, two hop strategy and one UE which can receive 3 paths (2 relayed and one direct form the AP). The whole processes could be split into three main areas of

- Measurements
- Signalling
- Actions

These need to be considered for each of the elements (BS/AP, Relays, and UEs). As we have stated above, mobility posses some constraints, so all the above need to be considered.

Figure 7-19 shows a simple representation of the minimum elements for cooperative mobile relaying, applicable also for conventional cooperative relaying. What is also shown is all the paths between those elements denoted as d_i with i=1, 5.

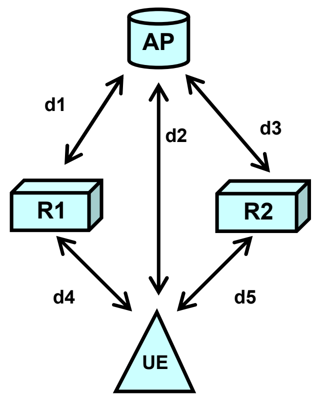


Figure 7-19 Simple case of cooperative relaying links.

What Table 7-2 shows is the link (in UL/DL) that any of the elements could potentially be measuring for reasons of power control, channel quality, location etc.

Table 7-2 Measurement of UL/DL links

	d1	d2	d3	d4	d5
AP	YES(UL)	YES(UL)	YES(UL)		
Relay1	YES(DL)			YES(UL)	
Relay2			YES(DL)		YES(UL)
UE		YES(DL)		YES(DL)	YES(DL)

As we see, this process posses some constrains especially for mobile relays which, in the most complicated should monitor both DL/UL links form the BS/UE.

Table 7-3 shows the combinations of signalling among the elements through specific links. For the Relay1-Relay2 combination it is for further study (FFS) if we assume direct communication among the relays. (This applies for the Symmetric strategy as stated in the previous section).

Table 7-3 Signalling of measurements

	AP	Relay1	Relay2	UE
AP to→	X	D1(DL)	D3 (DL)	D2(DL)
Relay1 to →	D1 (UL)	X	FFS	D4(DL)
Relay2 to →	D3(UL)	FFS	X	D5(DL)
UE to →	D2 (UL)	D4(UL)	D5(UL)	X

Thus, what the intention is to study which of those measurements/links and actions related to the above should be evaluated or used under the perspective of mobility. For instance, the less “mobile” a relay is, the less frequent feedback mechanisms from it to the AP are required. This is very applicable to Adaptive Decode & Forward schemes where a number of options exist e.g. adaptive/simple/complex decode & forward schemes. Additionally we have to bear in mind that d1/d2 links might be considerably impacted due to signalling not only from the relay to AP, but also due to transparent signalling from the UEs (through the relay) to the AP.

Another issue we need to take into account is the location of the coordination mechanisms and algorithms. If for instance, algorithms are located in the AP, this mean that measurements from the UE have to pass through the relay and be directed to the AP which means additional signalling for the (Relay-AP) UL/DL links.

Finally, we have to take into account that the target for cooperative relaying is the enhancement of reception at the UE. Thus, in general terms, it is only the UE that can "know"/measure the quality at its side which puts additional burden and signalling in the UE. However, what the other elements (e.g. relays) can do is estimate and based on those estimations/measurements (without any feedback from the UE) take certain decisions instead of relaying to UE UL signalling.

In this section, we have assumed two relays and two hops. The higher the number of relays we have and especially the more hops we assume, the more complicated the system becomes and the more difficult to deal with. It would be interesting though, to evaluate the amount of additional resources the network needs to allocate, based on the number of relays or the number of hops. For instance, for each additional relay we assume (still for two hops) the paths are increased by a factor of two.

There are a number of issues that could be considered for initial simulations.Some of those could be related to interference measurements (Algorithms for evaluating all possible links and combinations of them and selecting the ones with the least possible interference and maximum gain (gain in terms of QoS) at the UE side), general Intra-cell and inter-cell interference rise for one/multiple UEs, Statistical results for connectivity issues e.g. number of relays that a UE/AP can see, average (Relay-UE) distances, number of times within a time window that connectivity is adequate etc, criteria for evaluating among a number of paths and selecting those for cooperative relaying etc

In general, simulations should aim to be simple, realistic, use common parameters as in the case of fixed relays, and be comparable to simple fixed-relay simulations.

As we have seen in the general section, cooperative mobile relaying is a very interesting and promising area for work. As stated also in the previous section, several are the protocols, algorithms and strategies that we can incorporate. These need to be evaluated under the mobility factor through, preferably, link/system level simulations. Evaluation should be done not only specifically for those techniques as "stand-alone" solutions, but also under the perspective of the side effects they might induce. However, as in the general case of mobile relaying, we have first to evaluate simple schemes and define the incremental gain that mobile cooperative relaying has to offer in conventional relaying approaches, before moving to more complex and demanding schemes.

7.3.4 Usage Region

What is of interest for the design of routing protocols and their corresponding link metrics is the area in which a relay should be located. Relaying becomes preferable over direct transmission in regions in which it achieves SNR gains.

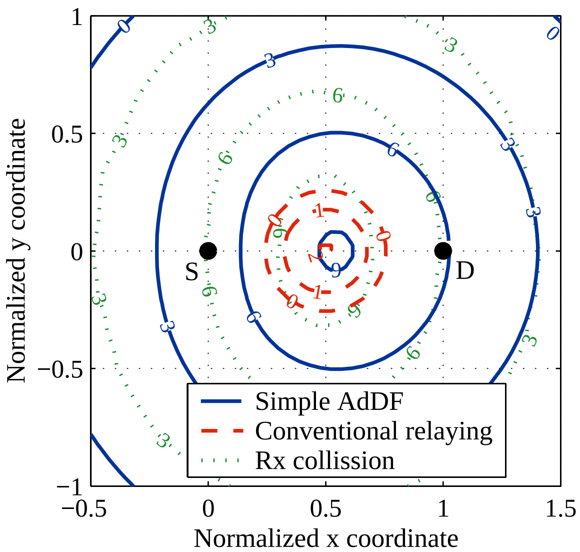


Figure 7-20: Usage region, i.e., the contours of the SNR gain for three protocols as a function the relay position. Source and destination are located at (0,0) and (1,0), respectively. Dashed: conventional relaying, solid: cooperative simple AdDF, dotted: receive collision. The contour labels represent the SNR gain over direct transmission (dB).

The corresponding area, also referred to as a *usage region*, is illustrated in Figure 7-20. It depicts the contours of the SNR gain, measured in dB, for $R=2$ bit/s/Hz and a pathloss exponent 4.0. Conventional relaying requires the relay to be located strictly between source and destination, while the cooperative

AdDF and receive collision schemes exhibit a considerably larger usage region. Conventional relaying achieves at most 2 dB, the cooperative AdDF scheme yields up to 9 dB potential energy savings. The receive collision scheme exhibits the largest usage region and SNR gains. All protocols perform best if the relay is located approximately halfway between source and destination. Moreover, we note that the general shape and the centers of these regions are very similar, thus indicating that routing algorithms with link metrics that are based on propagation losses work well for all of the studied relaying protocols. In other words, network protocols for routing, relay selection, and resource assignment that have been designed for conventional relaying are equally applicable for cooperative relaying.

7.3.5 Examples for System Connectivity Models

Figure 7-21 to Figure 7-29 show examples for each system connectivity model with the chosen channel allocation in brackets. Other channel allocations are possible, in some cases with better performance. The given channel allocations have been chosen since they can be consistently applied across all system connectivity models to illustrate the connectivity differences between the models. It is noted in the following discussion where an alternate channel allocation for a model would correspond to one of the distributed spatial diversity techniques presented in the literature.

Single Relay Single Destination (1R1D): This system connectivity model occurs when each relay is connected to one transmitter and the destination is connected to one transmitter. The minimum cost constraint set is 2CA+NRC+NDC+SCR+SCT.

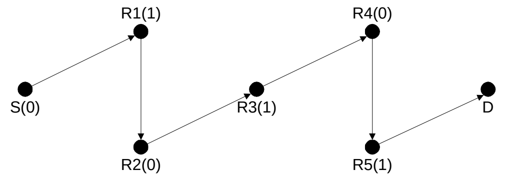


Figure 7-21: Example 1R1D System Connectivity Model

The 1R1D model is minimally connected, and corresponds to the multi-hop models presented in [230][239].

Single Relay 2Chnl Destination (1R2D): This system connectivity model occurs when each relay is connected to one transmitter and the destination is connected to a subset of transmitters on one channel. The minimum cost constraint set is 2CA+NRC+DC+SCR+SCT.

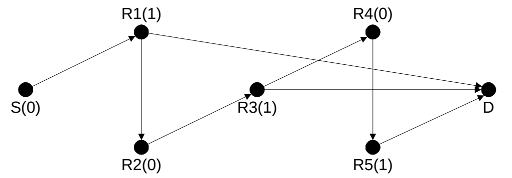


Figure 7-22: Example 1R2D System Connectivity Model

The 1R2D model corresponds to the multi-user diversity models presented in [227][229] under an alternate channel allocation where the source transmits on channel C0 and all relays transmit on channel C1. The source and destination are connected to all relays.

Single Relay 2Chnl Destination 2Chnl Source (1R2D2S): This system connectivity model occurs when each relay is connected one transmitter, the destination is connected to a subset of transmitters on one channel, and the source is connected to a subset of relays on one channel and the destination on the other. The minimum cost constraint set is 2CA+NRC+DC+SCR+MCT.

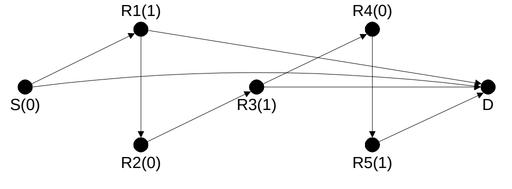


Figure 7-23: Example 1R2D2S System Connectivity Model

The 1R2D2S model corresponds to the cooperative diversity models presented in [232]-[235],[237][238] under an alternate channel allocation where the source transmits on channels C0 and C1 and all relays transmit on channel C1. The source and destination are connected to all relays and to each other.

Single Relay Full Destination (1RFD): This system connectivity model occurs when each relay is connected to one transmitter and the destination is connected to all transmitters. The minimum cost constraint set is 2CA+NRC+DC+MCR+SCT.

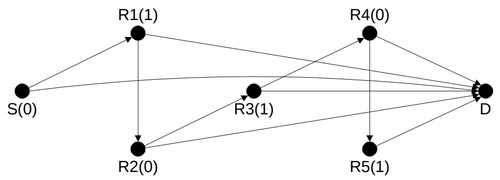


Figure 7-24: Example 1RFD System Connectivity Model

The 1RFD model corresponds to the cooperative diversity models presented in [232]-[235],[237][238] under an alternate channel allocation where the source transmits on channel C0 and all relays transmit on channel C1. The source and destination are connected to all relays and to each other.

2Chnl Relay 2Chnl Destination (2R2D): This system connectivity model occurs when each relay is connected to a subset of transmitters on one channel and the destination is connected to a subset of transmitters on one channel. The minimum cost constraint set is 2CA+RC+DC+SCR+SCT.

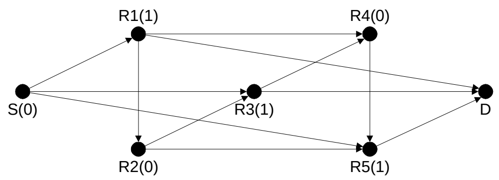


Figure 7-25: Example 2R2D System Connectivity Model

2Chnl Relay 2Chnl Destination Full Source (2R2DFS): This system connectivity model occurs when each relay is connected to a subset of transmitters on one channel, the destination is connected to a subset of transmitters on one channel, and the source is connected to all receivers. The minimum cost constraint set is 2CA+RC+DC+SCR+MCT.

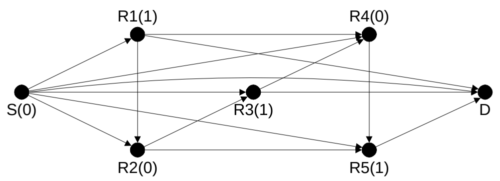


Figure 7-26: Example 2R2DFS System Connectivity Model

2Chnl Relay Full Destination (2RFD): This system connectivity model occurs when each relay is connected to a subset of transmitters on one channel and the destination is connected to all transmitters. The minimum cost constraint set is 2CA+RC+DC+MCR+SCT.

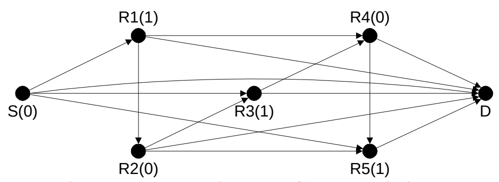


Figure 7-27: Example 2RFD Connectivity Model

The 2RFD model effectively extends the cooperative diversity models presented in [232]-[235],[237][238] to the case where the relays belong to different cooperation groups in multiple tiers between the source and destination.

2Chnl Relay Full Destination Full Source (2RFDFS): This system connectivity model occurs when each relay is connected to a subset of transmitters on one channel, the destination is connected to all transmitters, and the source is connected to all receivers. The minimum cost constraint set is 2CA+RC+DC+MCR+MCT.

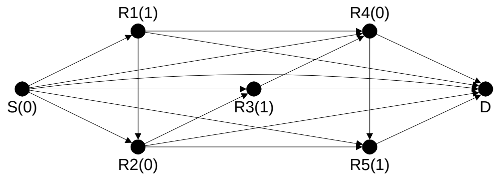


Figure 7-28: Example 2RFDFS System Connectivity Model

Full Relay Full Destination (FRFD): Each relay is connected to all transmitters previous along the transmission path and the destination is connected to all transmitters. The minimum cost constraint set is NCA+RC+DC+MCR+SCT.

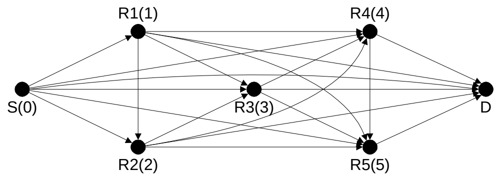


Figure 7-29: Example FRFD System Connectivity Model

The FRFD model is fully connected, and corresponds to the multi-hop diversity models presented in [160][161].

8. Annex III: Acronyms

ACK	ACKnowledgment
ACS	Access Control Server
AE_{WX}	WINNER Access Equipment for Mode X
AMC	Adaptive Modulation and Coding
AN	Ambient Networks
AP	Access Point
AR	Access Router
ARQ	Automatic Repeat reQuest
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BLER	Block Error Rate
BS	Base Station
BTS	Base Transceiver Station
CCK	Complementary Code Keying
CDMA	Code-Division Multiple Access
CFP	Contention Free Period
CSMA/CA	Carrier-Sense Multiple Access/Collision Avoidance
CS_W	WINNER Connection Service
CQI	Channel Quality Indicator
DCF	Distributed Coordination Function
DIFS	Distributed Inter-Frame Space
DL	Downlink
DSSS	Direct Sequence Spread Spectrum
EDCA	Enhanced Distributed Channel Access
FBRN	Fixed Bridging Relay Node
FCS	Fast Cell Selection
FDD	Frequency Division Duplex
FFT	Fast Fourier Transform
FHSS	Frequency Hopping Spread Spectrum
FoM	Figures of Merit
FPLRN	Fixed Physical Layer Relay Node
FRRN	Fixed Routing Relay Node
FRS	Fixed Relay Station
GLL	Generic Link Layer
GPSR	Greedy Perimeter Stateless Routing
HARQ	Hybrid Automatic Repeat reQuest
HCF	Hybrid Coordination Function
HERS	HEterogeneous Relay Station
HORS	HOmogeneous Relay Station
HSDPA	High Speed Downlink Packet Access
HS-DSCH	High Speed Downlink Shared Channel
HS-PDSCH	High Speed Physical Downlink Shared Channel

IBSS	Independent Basic Service Set
IFFT	Inverse FFT
IP	Internet Protocol
L2T	Layer 2 Tunnel
LOS	Line of Sight
LUT	Look-Up Table
MAC	Medium Access Control
MAC-g	Generic Medium Access Control
MAC-r	Mode-Specific Medium Access Control
MBRN	Mobile Bridging Relay Node
MCP	Modes Convergence Protocol
MIMO	Multiple-Input Multiple-Output
MPLRN	Mobile Physical Layer Relay Node
MRA	Multi-Radio Access
MRRN	Mobile Routing Relay Node
NLOS	None LOS
ODMA	Opportunity Driven Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PCF	Point Coordination Function
PDU	Protocol Data Unit
PER	Packet Error Rate
PHY	Physical Layer
PS	Packet Switch
QAM	Quadrature Amplitude Modulation
QoS	Quality-of-Service
QPSK	Quadrature Phase Shif Keying
RAB	Radio Access Bearer
RAN	Radio Access Network
RANG	Radio Access Network Gateway
RAT	Radio Access Technology
RLC	Radio Link Control
RN	Relay Node
RNC	Radio Network Controller
RRC	Radio Resource Control
RRC-g	Generic Radio Resource Control
RRC-r	Mode-Specific Radio Resource Control
RRM	Radio Resource Management
SDU	Service Data Unit
SIFS	Short Inter-Frame Space
SNIR	Signal-to-Noise-Interference Ratio
SS	Secondary Station
STTD	Space-Time Transmit Diversity
TDMA	Time-Division Multiple Access
TNL	Transport Network Layer

TTI	Transmission Time Internal
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UT	User Terminal
UTRA	Universal Terrestrial Radio Access
WCDMA	Wideband Code Division Multiple Access
WiFR	Wireless Fixed Relays Routing
WINNER	WWI New Radio IP
WLAN	Wireless Local Area Network
WSE	Weighted Spectral Efficiency
WWI	Wireless World Initiative

9. Annex IV: References

[1] Jörn von Häfen, et al.: ”WINNER D7.1 System requirements“, IST-2003-507581 WINNER

[2] Joachim Sachs, “A Generic Link Layer for Future Generation Wireless Networking,” *in proceedings IEEE ICC*, May 2003.

[3] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Overall Description, 3GPP TS 25.401.

[4] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Radio Interface Protocol Architecture (Release 6), 3GPP TS 25.301.

[5] B. Walke, R. Pabst, D. Schultz: A Mobile Broadband System based on Fixed Wireless Routers, Proceedings of International Conference on Communication Technology 2003 (pp. 1310-1317), Beijing, China

[6] D. Schultz, B. Walke, R. Pabst, T. Irnich, Fixed and Planned Relay Based Radio Network Deployment Concepts, Proc. of WWRF10, Oct. 2003, New York, USA

[7] M. Siebert, B. Walke “Design of Generic and Adaptive Protocol Software (DGAPS),” *Proc. of the Third Generation Wireless and Beyond (3Gwireless '01)*, San Francisco USA, June 2001.

[8] L. Berlemann, M. Siebert, B. Walke “Software Defined Protocols Based on Generic Protocol Functions for Wired and Wireless Networks,” in Proc. of Software Defined Radio Technical Conference 2003, Orlando, USA

[9] L. Berlemann, B. Walke and S. Mangold, "Behaviour Based Strategies in Radio Resource Sharing Games," to appear in Proc. of PIMRC 2004, Barcelona Spain, 5-8 September 2004.

[10] L. Berlemann, G. R. Hiertz, B. Walke and S. Mangold, "Cooperation in Radio Resource Sharing Games of Adaptive Strategies," to appear in Proc. Of VTC Fall 2004, Los Angeles USA, 26-29 September 2004.

[11] B. Braden et.al., “RFC 2309 - Recommendations on Queue Management and Congestion Avoidance in the Internet,” http://www.faqs.org/rfcs/rfc2309.html.

[12] http://www.breitband-isdn.ch

[13] http://www.nokia.com

[14] G. Foschini and M. Gans. On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas, Wireless Personal Communications, Mar 1998, vol. 6, pp 311-335

[15] H. Bolckei, D. Gesbert and A. J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems IEEE Trans. on Communications, 50(2), Feb. 2002, pages 225 -234

[16] S.M. Alamouti. A Simple Transmit Diversity Technique for Wireless Communications. IEEE Journal on Selected Areas in Communications, 16(8): 1451–1458, October 1998.

[17] D. Lacroix and D. Castelain. A study of OFDM parameters for high data rate radio LANs. In Proceedings of the IEEE Vehicular Technology Conference, volume 2, pages 1075–1079, 2000.

[18] WINNER Deliverable, D7.2: “System Assessment Criteria Specification”, 2004.

[19] WINNER Internal Document, “Calibration of System Simulator Tools”, 2004.

[20] WINNER Annex I: “Description of Work”, 2003.

[21] W. Mohr, R. Lueder, and K.-H. Moehrmann, “Data Rate Estimates, Range Calculations and Spectrum Demand for New Elements of Systems Beyond IMT-2000,” in Proc. of WPMC’02, Honolulu, Hawaii, USA, Oct 2002.

[22] J. Khun-Jush, P. Schramm, U. Wachsmann, and F. Wenger, “Structure and Performance of the HiperLAN/2 Physical Layer,” in Proc. of the VTC Fall-1999, Amsterdam, The Netherlands, Sep 1999, pp. 2667–2671.

[23] 3GPP, “Selection Proc. for the Choice of Radio Transm. Techn. of the UMTS, (UMTS 30.03),” ETSI, Sophia Antipolis, France, Report TR 101 112, V3.2.0, Apr. 1998.

[24] L. M. Correia(Editor), “Wireless Flexible Personalised Communications, COST 259: European Co-operation in Mobile Radio Research”, COST Secretariat, European Commission, Brussels, Belgium, Mar 2001.

[25] E. Damosso and L. M. C. (Editors), “COST 231 Final Report - Digital Mobile Radio: Evolution Towards Future Generation Systems”, COST Secretariat, European Commission, Brussels, Belgium, Aug 1999.

[26] BRAIN, “D 3.1: Technical Requirements and Identification of Necessary Enhancements for HIPERLAN Type 2 Air Interface”, IST-1999-10050 BRAIN WP3 - Deliverable, Sept. 2000.

[27] ETSI. Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; System Overview, Feb. 2000. ETSI, TR 101 683.

[28] ETSI. Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions, Dec. 2001. ETSI, TS 101 761.

[29] ETSI. Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer, Dec. 2001. ETSI, TS 101 475.

[30] ETSI. Broadband Radio Access Networks (BRAN); HIPERACCESS; System Overview, Mar. 2002. ETSI, TR 102 003.

[31] B. Walke. Mobile Radio Networks: Networking, Protocols and Traffic Performance, 2nd Edition; ISBN: 0-471-49902-1, Wiley, November 2001

[32] ITU-T. Information Technology - Open Systems Interconnection - Basic Reference Model: The Basic Model. Recommendation X.200, ITU, Jul. 1994.

[33] IST-STRIKE, D5.2 - DLC Implementation Specifications for MTMR Systems, June 2004

[34] Hoymann, C. and Forkel, I.: “MAC Layer Concepts to Support Space Division Multiple Access in IEEE 802.16a,” In Proceedings of The Seventh International Symposium on Wireless Personal Multimedia Communications (WPMC 2004), p.p. V3, 300-304, Abano Terme, Italy, Sep. 2004

[35] IEEE. Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE P802.16-REVd/D5-2004, May 2004.

[36] IEEE. Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE P802.16e/D5, September 2004.

[37] www.wimaxforum.org

[38] ITU-R Rec. M.1390, Methodology for the Caluculation of IMT-2000 Terrestrial Spectrum Requirements, 1999

[39] T. Irnich, B. Walke, Spectrum Estimation Methodology for Next Generation Systems, to appear in Proc. of PIMRC 2004, Oct 2004, Barcelona, Spain

[40] N. Esseling, E. Weiss, A. Krämling, W. Zirwas, Supporting Cost Efficient Public 5GHz-W-LAN Roll Out with a Multi Hop HiperLAN/2 Concept, Proc. of 55th VTC 2002 Spring, May 2002, Birmingham, Al, USA

[41] B. Walke, R. Pabst, D. Schultz: A Mobile Broadband System based on Fixed Wireless Routers, Proceedings of International Conference on Communication Technology 2003 (pp. 1310-1317), Beijing, China

[42] D. Schultz, B. Walke, R. Pabst, T. Irnich, Fixed and Planned Relay Based Radio Network Deployment Concepts, Proc. of WWRF10, Oct. 2003, New York, USA

[43] N. Esseling, B. Walke, R. Pabst, Performance Evaluation of a Fixed Relay Concept for Next Generation Wireless Systems, to appear in Proc. of PIMRC2004, October 2004, Barcelona, Spain

[44] T. Cover and J. Thomas, *Elements of Information Theory.* Wiley, 1991

[45] L.R. Ford and D.R. Fulkerson, *Flows in networks*. Princeton University press, 1962

[46] R. Koetter and M. Medard, *An algebraic approach to network coding*. IEEE Transactions on Information Theory, 2(5): 782-795, March 2000

[47] Piyush Gupta and P.R. Kumar, *The capacity of wireless networks*. IEEE Transactions on Information Theory, 46(2): 388-404, March 2000

[48] L.-L. Xie and P.R. Kumar, *A network information theory for wireless communication: scaling and optimal operation*. Submitted to IEEE Transaction on Information Theory, April 2002

[49] Piyush Gupta and P.R. Kumar, *Towards an information theory of large networks: an achievable rate region*. IEEE Transactions on Information Theory, 49(8)1877-1894, August 2003

[50] M. Gastpar, *To code or not to code*. PhD Thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), March 1996

[51] Thomas M. Cove rand Abbas A. El Gamal, *Capacity Theorem for the relay channels*. IEEE Transactions on Information Theory, 25(5): 572-584, September 1979

[52] Edward C. van der Meulen, *A survey on multi-way channels in information theory*. IEEE International Symposium on information theory, 23(1): 1-37, July 2002

[53] B. Schein and R. Gallager, *The Gaussian parallel relay network*. IEEE International Symposium on information theory, page 22, June 2000

[54] M. Gastpar and M. Vitterli, *On the capacity of wireless networks: the relay case*. INFOCOM, 21st annual joint conference of the IEEE computer and communications societies, 3:1577-1586, June 2002

[55] L.-L. Xie and P.R. Kumar, *An achievable rate for the multiple-level relay channel*. Submitted to IEEE Transactions on Information Theory, November 2003

[56] G. Kramer and M. Gastpar, *Capcity theorems for wireless relay channels*. 41st Allerton conference on communication, control, and computer, October 2003

[57] A. Host-Madsen, *On the capacity of wireless relaying*. Vehicular Technology Conference VTC Spring, 2002

[58] J. Laneman and Gregory W. Wornell, *Energy-efficient antenna sharing and relaying for wireless networks*. Wireless communications and networking IEEE conference, 1:7-12, September 2000

[59] FITNESS Consortium, IST-2000-30116, Deliverable D4.2, UMTS/WLAN Interoperability Scenarios and Mechanisms, July 2003

[60] Mattias Lampe, Hermann Rohling, and Wolfgand Zirwas. Misunderstandings about Link Adaptation for frequency selective fading channels. 2002.

[61] S. Nanda and K. Rege, Frame error rates for convolutional codes on fading channels and the concept of effective Eb/N0, IEEE Transaction on Veh. Techn., 47(4), November 1998

[62] Klaus Witrisal, Yong-Ho Kim, and Ramjee Prasad, A novel approach for performance evaulation of OFDM with error correction coding and interleaving, Proceedings of the IEEE VTC, Fall, vol. 1 (294-299), September 1999

[63] J. Medbo and H. Hallenberg J. Berg, Propagation characteristics at 5GHz in typical radio LAN scenarios, IEEE Transaction on Veh. Techn., 1(185-189), 1999

[64] FITNESS Consortium, IST-2000-30116, Deliverable D3.3.1, MTMR Baseband transceivers needs for intra-system and inter-system (UMTS/WLAN) reconfigurability, November 2003

[65] N. Esseling, B. Walke and R. Pabst, "Performance Evaluation of a Fixed Relay Concept for Next Generation Wireless Systems", Invited Paper at PIMRC 2004, Barcelona, 5.-8. Sept. 2004 also to be published in "EMERGING LOCATION AWARE BROADBAND WIRELESS ADHOC NETWORKS", Kluwer 2004

[66] J. Habetha, R. Dutar, and J. Wiegert, “Performance Evaluation of HiperLAN/2 Multi-hop Ad Hoc Networks,” in *Proc. European Wireless*, Florence, Italy, Feb 2002, pp. 25–31.

[67] T. Irnich, D. Schultz, R. Pabst, and P. Wienert, “Capacity of a Relaying Infrastructure for Broadband Radio Coverage of Urban Areas,” in *Proc. 10th WWRF Meeting,* New York, USA, Oct 2003, http://www.comnets.rwth-aachen.de/

[68] N. Esseling, H. Vandra, and B. Walke, “A Forwarding Concept for HiperLAN/2,” in *Proc. European Wireless 2000,* Dresden, Germany, Sep 2000, pp. 13–17.

[69] S. Mangold, “Analysis of IEEE802.11e and Application of Game Models for Support of Quality-of-Service in Coexisting Wireless Networks”, PhD Thesis, Aachen University (RWTH), Aachen, 2003, http://www.comnets.rwth-aachen.de/ .

[70] N. Esseling, E. Weiss, A. Kraemling, and W. Zirwas, “A Multi Hop Concept for HiperLAN/2: Capacity and Interference,” in *Proc. European Wireless 2002*, vol. 1, Florence, Italy, Feb 2002, pp. 1–7.

[71] D. Schultz, B. Walke, R. Pabst, and T. Irnich, “Fixed and Planned Relay Based Radio Network Deployment Concepts,” in *Proc. 10th WWRF Meeting,* New York, USA, Oct 2003, http://www.comnets.rwth-aachen.de/ .

[72] W. Mohr, R. Lueder, and K.-H. Moehrmann, “Data Rate Estimates, Range Calculations and Spectrum Demand for New Elements of Systems Beyond IMT-2000,” in *Proc. of WPMC’02,* Honolulu, Hawaii, USA, Oct 2002.

[73] J. Khun-Jush, P. Schramm, U. Wachsmann, and F. Wenger, “Structure and Performance of the HiperLAN/2 Physical Layer,” in *Proc. of the VTC Fall-1999*, Amsterdam, The Netherlands, Sep 1999, pp. 2667–2671.

[74] 3GPP, “Selection Proc. for the Choice of Radio Transm. Techn. of the UMTS, (UMTS 30.03),” ETSI, Sophia Antipolis, France, Report TR 101 112, V3.2.0, Apr. 1998.

[75] L. M. Correia(Editor), “Wireless Flexible Personalised Communications, COST 259: European Co-operation in Mobile Radio Research”, *COST Secretariat, European Commission*, Brussels, Belgium, Mar 2001.

[76] E. Damosso and L. M. C. (Editors), “COST 231 Final Report - Digital Mobile Radio: Evolution Towards Future Generation Systems”*, COST Secretariat, European Commission,* Brussels, Belgium, Aug 1999.

[77] BRAIN, “D 3.1: Technical Requirements and Identification of Necessary Enhancements for HIPERLAN Type 2 Air Interface”, *IST-1999-10050 BRAIN WP3 - Deliverable,* Sept. 2000.

[78] D. Schultz, R. Pabst, and T. Irnich, “Multi-hop based Radio Network Deployment for efficient Broadband Radio Coverage,” in Proc. WPMC, Yokosuka, Japan, October 2003.

[79] UMTS 30.03 version 3.2.0, Technical Report TR 101 112, ETSI, April 1998.

[80] N. Esseling, E. Weiss, A. Krämling, W. Zirwas, “A Multi Hop Concept for HiperLAN/2: Capacity and Interference”, in *Proc. European Wireless 2002*, Vol.1, pp. 1-7, Florence, Italy, February 2002.

[81] M. Hassan-Ali, K. Pahlavan, "A new statistical model for site-specific indoor radio propagation prediction based on geometric optics and geometric probability", IEEE Transactions on Wireless Communications, vol. 1, pp. 112 -124, 2002.

[82] 802.16a-2003, “IEEE Standard for local and metropolitan area networks --- Part 16: Air Interface for Fixed Wireless Access Systems --- Amendment 2: Medium Access Control Modification and Additional Physical Layer Specifications for 2-11 GHz”, June 2003.

[83] 802.16-2001, “IEEE Standard for local and metropolitan area networks --- Part 16: Air Interface for Fixed Wireless Access Systems”, April 2001.

[84] Carl Eklund, Roger B. Marks, Kenneth L. Stanwood, Stanley Wang, “IEEE Standard 802.16: A Technical Overview of the WirelessMANTM Air Interface for Wireless Access Systems”, IEEE Communication Magazine, June 2002.

[85] Simone Redana, Matthias Lott, “Performance Analysis of IEEE 802.16a in Mesh Operation Mode”, in Proc. of the 13th IST SUMMIT, Lyon, France, June 2004.

[86] N. Esseling, E. Weiss, A. Krämling, W. Zirwas, “A Multi Hop Concept for HiperLAN/2: Capacity and Interference”, in *Proc. European Wireless 2002*, Vol.1, pp. 1-7, Florence, Italy, February 2002.

[87] B. Walke, R. Pabst, D. Schultz: A Mobile Broadband System based on Fixed Wireless Routers, Proceedings of International Conference on Communication Technology 2003 (pp. 1310-1317), Beijing, China

[88] N. Esseling, E. Weiss, A. Krämling, W. Zirwas: A Multi Hop Concept for HiperLAN/2: Capacity and Interference. In Proc. European Wireless 2002, Vol. 1, pp. 1-7, Florence, Italy, February 2002. available from www.comnets.rwth-aachen.de/cnroot_engl.html/~publications .

[89] K. Fall, K. Varadhan, “ns notes and documentation,” The VINT Project, UC Berkley, LBL, USC/ISI, and Xerox PARC, 1999. Avail. from http://www.isi.edu/nsnam/ns:

[90] 802.16a-2003, "IEEE Standard for local and metropolitan area networks --- Part 16: Air Interface for Fixed Wireless Access Systems --- Amendment 2: Medium Access Control Modification and Additional Physical Layer Specifications for 2-11 GHz", June 2003.

[91] 802.16-2001, "IEEE Standard for local and metropolitan area networks --- Part 16: Air Interface for Fixed Wireless Access Systems", April 2001.

[92] 802.16-REVd/D5-2004, "IEEE Standard for local and metropolitan area networks --- Part 16: Air Interface for Fixed Wireless Access Systems, May 2004.

[93] Carl Eklund, Roger B. Marks, Kenneth L. Stanwood, Stanley Wang, "IEEE Standard 802.16: A Technical Overview of the WirelessMANTM Air Interface for Wireless Access Systems", IEEE Communication Magazine, June 2002.

[94] http://www.isi.edu/nsnam/ns/

[95] 33GPP 25.106 – "UTRA repeater radio transmission and reception".

[96] 3GPP 25.143 – "UTRA repeater conformance testing".

[97] 3GPP 25.956 – "UTRA Repeater Planning Guidelines and System Analysis".

[98] www.etsi.org/mesa

[99] 3GPP 25.924 – "Opportunity Driven Multiple Access".

[100] WINNER WP3 T3.2 "Terminology proposal" - BSCW server

[101] WINNER WP3, Internal Report 3.1 (IR3.1)

[102] WINNER WP3, "Calibration of simulation tools"

[103] WINNER WP7, "System Assessment Criteria Specification" - D7.2"

[104] White paper "Relay based Deployment concepts for Wireless and Mobile Broadband Cellular Radio", WWRF/WG4 Relaying subgroup, WWRF#9, Zurich, Switzerland, July 2003.

[105] Halim Yanikomeroglu: Fixed and Mobile Relaying Technologies for Cellular Networks, Second Workshop on Applications and Services in Wireless Networks, July 2004, Paris France.

[106] D. Schultz, R. Pabst, T. Irnich, "Multi-hop based Radio Network Deployment for efficient Broadband Radio Coverage", Proc. of WPMC, October 2003, Yokosuka, Japan

[107] 3GPP 25.305, State 2 functional specification of user equipment (UE) positioning in UTRAN

[108] 3GPP, R2-010301, (R4-010193 to TSG RAN WG2) LS on effect of repeaters on positioning accuracy

[109] Byron Alex. Bakaimis, "A solution on the degradation of accuracy levels for mobile positioning in the presence of repeaters for FDD in UMTS", First International Workshop in Location Based Services, IEE, September 2002

[110] Byron Alex. Bakaimis, "Reconfigurable mobile relays", (to be submitted at WWRF#12)

[111] WINNER WP3 Workplan: "WP3_T3.3_Workplan", https://bscw.eurescom.de.

[112] C. Perkins, Ad Hoc on Demand Distance Vector (AODV) Routing, Internet draft, 1997.

[113] D. B. Johnson and D. A. Maltz, Dynamic Source Routing (DSR) in Ad Hoc Wireless Networks, Internet draft, 1997.

[114] C. Perkins, E. Royer, S. R. Das and M. K. Marina, Performance Comparison of Two On-Demand Routing Protocols for Ad Hoc Networks, IEEE Personal Communications, vol. 8, no. 1, Feb. 2001, pp. 16-28.

[115] Z. J. Haas and M. R. Pearlman, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks," Internet draft, 1997.

[116] S. Chakrabarti and A. Mishra, QoS Issues in ad hoc wireless network, IEEE Communications Magazine, vol. 39, no. 2, Feb 2001, pp. 142-148.

[117] E. Crawley et al., "A Framework for QoS-Based Routing in the Internet," RFC 2386, Aug. 1998.

[118] J.L. Sobrinho and A.S. Krishnakumar, Quality-of-Service in ad hoc Carrier Sense Multiple Access Wireless Networks, IEEE JSAC, vol. 17, no. 8, Aug. 1999, pp. 1353-1414.

[119] A. Iwata et al., Scalable Routing Strategies for Ad Hoc Wireless Networks, IEEE JSAC, vol. 17, no. 8, Aug. 1999, pp. 1369–79.

[120] C.R. Lin and J.-S. Liu, QoS Routing in Ad Hoc Wireless Networks, IEEE JSAC, vol. 17, no. 8, Aug. 1999, pp. 1426–38.

[121] S. Chen and K. Nahrstedt, Distributed Quality-of-Service Routing in Ad Hoc Networks, IEEE JSAC, vol. 17, no. 8, Aug. 1999, pp. 1488-1505.

[122] C.R. Lin, On-demand QoS routing in multi-hop mobile networks, IEEE INFOCOM 2001. April 2001, vol. 3, pp. 1735-1744.

[123] Chenxi Zhu and M.S. Corson, QoS routing for mobile ad hoc networks, IEEE INFOCOM 2002, June 2002, vol. 2, pp. 958-967.

[124] Hongxia Sun, H.D. Hughes, Adaptive QoS routing based on prediction of local performance in ad hoc networks, IEEE WCNC 2003, March 2003, vol. 2, pp. 1191- 1195.

[125] M. Garey and D. Johnson. Computers and Intractability: a Guide to the Theory of NP-Completeness. W.H. Freeman, 1979.

[126] E. M. Royer and C.-K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," IEEE Personal Communications, April 1999, pp.46-55.

[127] S.-J. Lee, M. Gerla and C.-K. Toh, "A Simulation Study of Table-Driven and On-Demand Routing Protocols for Mobile Ad Hoc Networks," IEEE Network, July/Aug. 1999, pp.48-54.

[128] M. Mauve, J. Widmer, H. Hartenstein, "A Survey on Position-Based Routing in Mobile Ad-Hoc Networks." In: IEEE Network. Vol. 15 (2001), Nr. 6, S. 30-39.

[129] C.E: Perkins, "Ad hoc Networks," Addison-Wesley, 2001.

[130] D. Chr. Schultz, R. Pabst, M. Schmidt, M. Lott, A. Pollard, J. Ojala, P. Pellati, J. Gosteau, P. Slanina, A. Santos, and Q Perez, IST-2003-507581, WINNER Internal Report 3.1.

[131] H. Bolukbasi, *Capacity and Scalability Analysis in Mesh, Cellular and Relay Networks*, M.A.Sc. Thesis, Dec. 2004, Carleton University; supervisors: H. Yanikomeroglu and D. Falconer.

[132] G. Cair, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Info. Theory,* vol. 44, no. 3, pp. 927-946, May 1998.

[133] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[134] C.E. Perkins and E.M. Royer, "Ad Hoc On-Demand Distance Vector (AODV) Routing", IETF Internet Draft, draft-ietf-manet-aodv-13.txt

[135] David B. Johnson, Davis A. Maltz, "Dynamic Source Routing in Ad Hoc Networks", Mobile Computing, T.Imielinski and H.Korth, Eds., Kulwer, 1996, pp.152-181.

[136] David B. Johnson, Davis A. Maltz, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks", October 1999 IETF Draft, http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-06.txt

[137] J.Li, J. Jannotti, D.De Couto, D.Karger and R.Morris, "A scalable location service for geographic ad-hoc routing", In MobiCom '00, pages 120 130, August 2000.

[138] S.Basagni, I.Chlamatac, V.Syrotiuk and B.Woodward, "A distance routing effect algorithm for mobility (DREAM)", In Proc. of the 4th Annual ACM/IEEE Int. Conf. on Mobile Computing and Networking (MOBICOM) '98, pages 76-84, Dallas, TX, USA, 1998.

[139] Y.-B. Ko, N.H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," in Proc. of ACM/SIGMOBILE 4th Annual Int. Conf. on Mobile Comupting and Networking (MobiCom´98), Dallas, Oct. 1998.

[140] T. Kosch, C. Schwingenschlögl, L. Ai "Information Dissemination in Multi-hop Inter-Vehicle Networks - Adapting the Ad-hoc On-demand Distance Vector Routing Protocol (AODV)," in Proc. of IEEE Intern. Conf. on *Intelligent Transportation Systems*, Singapore, Sept. 3-6, 2002.

[141] M. Lott, M. Weckerle, W. Zirwas, H. Li, E. Schulz, "Hierarchical cellular multi-hop networks," in Proc. of 5th European Personal Mobile Communications Conference (EPMCC 2003), Glasgow, Scotland, 22-25 April, 2003.

[142] M. Lott, M. Weckerle, M. Siebert, "Scheduling in wireless multi-hop networks," in Proc. of Int. Symposium of Perf. Eval. of Comp. Telecom. Systems (SPECTS´04), San Jose, 25.-29. July 2004.

[143] H. Li, M. Lott, W. Zirwas, M. Weckler, E. Schulz, Multi-hop Communications in Future Mobile Radio Networks, IEEE PIMRC 2002, Sept. 2002.

[144] H. Li, M. Lott, W. Zirwas, M. Weckler, E. Schulz, Hierarchical Cellular Multi-hop Networks, EPMCC 2003, March 2003.

[145] A.A. Assad, "Multicommodity network flow – a survey". Networks, vol. 8 pagg. 37-91, John Wiley & Sons, Inc.

[146] http://www.isi.edu/nsnam/ns

[147] Daniel Chr. Schultz, Bernhard Walke, et al. „Workplan for WINNER WP3", , 27.02.04

[148] H. J. Chao, X. Guo, „Quality of Service Control in High-Speed Networks", John Wiley & Sons, New York, 2002

[149] Z. Wang, J. Crowcroft, „Quality-of-service routing for supporting multimedia applications", IEEE Journal on Selected Areas in Communications, vol. 14, no. 7, pp. 1228-1234, September 1996

[150] S. Chen, K. Nahrstedt, „An Overwiev of Quality-of-Service Routing for the Next Generation High-Speed Networks: Problems and Solutions", IEEE Network, pp. 64-79, November/December 1998

[151] M. Benzaid, P. Minet, K. Al Agha, „A framework for Integrating Mobile-IP and OLSR Ad-Hoc Networking for Future Wireless Mobile Systems", Wireless Networks, vol. 10, issue 4, pp.377-388, July 2004

[152] T. Clausen, P. Jacquet, „Optimized Link State Routing Protocol (OLSR)", RFC 3626, IETF, October 2003, http://www.ietf.org

[153] E. Martins, M. Pascoal, J. Santos, „Labeling Algorithms for Ranking Shortest Paths", 1999, http://citeseer.ist.psu.edu

[154] T. H. Cormen, C. E. Leiserson, R. L. Rivest, „Introduction to algorithms", The MIT Press, 2nd edition, September 1, 2001.

[155] T. Janevski, „Traffic Analysis and Design of Wireless IP Networks", Artech House, 2003.

[156] S.M. Alamouti. A Simple Transmit Diversity Technique for Wireless Communications. IEEE J. Select. Areas Comm., 16(8):1451-1458, October 1998.

[157] P. A. Anghel, G. Leus, and M. Kaveh. Distributed Space-Time Coding in Cooperative Networks. In Proc. of the Nordic Signal Processing Symp., Norway, October 2002.

[158] P. A. Anghel, G. Leus, and M. Kaveh. Multi-User Space-Time Coding in Cooperative Networks. In Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, April 2003.

[159] A. Agustin, O. Mu noz, and J. Vidal. A Game-theoretic Approach for Cooperative MIMO Systems with Cellular Reuse of the Relay Slot. In Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Montreal, Quebec Canada, May 2004.

[160] J. Boyer, D. Falconer, and H. Yanikomeroglu. A Theoretical Characterization of the Multi-hop Wireless Communications Channel with Diversity. In Proc. IEEE Globecom 2001, volume 2, pages 841 -845, San Antonio, Texas, November 2001.

[161] J. Boyer, D. D. Falconer, and H. Yanikomeroglu. Multi-hop Diversity in Wireless Relaying Channels. IEEE Trans. Comm., 2004. Submitted, revised Jan. 2004, available at http://www.sce.carleton.ca/faculty/yanikomeroglu/cv/ publications.shtml.

[162] S. Barbarossa and G. Scutari. Cooperative Diversity through Virtual Antenna Arrays in Multi-hop Networks. In Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, April 2003.

[163] T. M. Cover and A. A. El Gamal. Capacity Theorems for the Relay Channel. IEEE Trans. Inform. Theory, 25(5):572-584, September 1979.

[164] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley & Sons, New York, 1991.

[165] M. Dohler, A. Gkelias, and H. Aghvami. 2-Hop Distributed MIMO Communication System. IEEE Electronics Letters, June 2003.

[166] M. Dohler, E. Lefranc, and H. Aghvami. Space-Time Block Codes for Virtual Antenna Arrays. In Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun., Lisbon, Portugal, September 2002.

[167] M. Dohler, F. Said, and H. Aghvami. Concept of Virtual Antenna Arrays. In IEEE Globecom 2002, Taipei, Taiwan, 2002.

[168] G. J. Foschini and M. J. Gans. On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas. Wireless Personal Communications, 6(3):311, March 1998.

[169] P. Gupta and P. R. Kumar. The Capacity of Wireless Networks. IEEE Trans. Inform. Theory, 46(2):388-404, March 2000.

[170] M. Gastpar, G. Kramer, and Gupta P. The Multiple-relay Channel: Coding and Antenna-clustering Capacity. In Proc. IEEE Int. Symp. Inform. Theory, Lausanne, Switzerland, July 2002.

[171] M. Gastpar and M. Vetterli. On the Capacity of Wireless Networks: The Relay Case. In Proc. IEEE INFOCOM 2002, volume 3, pages 1577-1586, New York, 2002.

[172] Joachim Hagenauer. Rate-Compatible Punctured Convolutional Code (RCPC Codes) and their Applications. IEEE Trans. Commun., 36(4):389-400, April 1988.

[173] P. Herhold. Cooperative Relaying - Protocols and Performance. PhD thesis in preparation, Technische Universität Dresden, Dresden, Germany, 2004.

[174] P. Herhold and G. Fettweis. CAIRO Project, Work Package 1. Technical report, TU Dresden, 2003.

[175] P. Herhold and G. Fettweis. CAIRO Project, Work Package 2. Technical report, TU Dresden, 2003.

[176] P. Herhold and G. Fettweis. CAIRO Project, Work Package 3. Technical report, TU Dresden, 2004.

[177] T. E. Hunter, A. Hedayat, M. Janani, and A. Nosratinia. Coded Cooperation with Space-Time Transmission and Iterative Decoding. In WNCG Wireless Networking Symposium, Austin, Texas, October 2003.

[178] T. E. Hunter and A. Nosratinia. Coded Cooperation under Slow Fading, Fast Fading, and Power Control. In Proc. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, California, November 2002.

[179] T. E. Hunter and A. Nosratinia. Cooperation Diversity through Coding. In Proc. IEEE Int. Symp. Inform. Theory, page 220, Lausanne, Switzerland, July 2002.

[180] T. E. Hunter and A. Nosratinia. Performance Analysis of Coded Cooperation Diversity. In IEEE Int. Conf. Commun. (ICC), Anchorage, Alaska, 2003.

[181] P. Herhold, W. Rave, and G. Fettweis. Relaying in CDMA Networks: Pathloss Reduction and Transmit Power Savings. In Proc. 57th IEEE Vehic. Technol. Conf., Jeju, Korea, April 2003.

[182] P. Herhold, E. Zimmermann, and G. Fettweis. Relaying Reconsidered: The Impact of Cooperation. IEEE J. Select. Areas Comm. Submitted Oct. 2003.

[183] P. Herhold, E. Zimmermann, and G. Fettweis. A Simple Cooperative Extension to Wireless Relaying. In Int. Zurich Seminar on Comm. (IZS), Zurich, Switzerland, February 2004. (Appendix available at www.ifn.et.tu-dresden.de/~herhold/publications).

[184] P. Herhold, E. Zimmermann, and G. Fettweis. Cooperative Relaying - Benefits and Challenges. In 13th Joint Workshop on Comm. and Coding (JWCC), Galtur, Austria, March 2004.

[185] P. Herhold, E. Zimmermann, and G. Fettweis. On the Performance of Cooperative Amplify-and-Forward Relay Networks. In 5th Int. ITG Conf. on Source and Channel Coding, Erlangen, Germany, January 2004.

[186] M. Janani, A. Hedayat, T. Hunter, and A. Nosratinia. Coded Cooperation in Wireless Communications: Space-time Transmission and Iterative Decoding. IEEE Trans. Signal Proc., 2003. Submitted.

[187] P. Larsson and H. Rong. Large-Scale Cooperative Relaying Network with Optimal Coherent Combining under Aggregate Relay Power Constraints. In Proc. Wireless World Research Forum, Meeting #8 (WWRF), Beijing, China, February 2004.

[188] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour. IEEE Trans. Inform. Theory, April 2003. (accepted for publication).

[189] O. Mu noz, A. Agustin, and J. Vidal. Cellular Capacity Gains of Cooperative MIMO Transmission in the Downlink. In Int. Zurich Seminar on Comm. (IZS), pages 22-26, Zurich, Switzerland, February 2004.

[190] R. U. Nabar, H. Bölcskei, and F. W. Kneub□uhler. Fading relay channels: Performance limits and space-time signal design. IEEE Journal on Selected Areas in Com- munications, June 2004. to appear, preprint available at http://www.nari.ee.ethz.ch/commth/pubs/viewpub.php?ident=jsac03.

[191] Ioannis Oikonomidis. Cooperative Coding Using Turbo Codes. In 13th Joint Workshop on Comm. and Coding (JWCC), Galt□ur, Austria, March 2004.

[192] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Visvanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis. Relay-based Deployment Concepts for Wireless and Mobile Broadband Cellular Radio. IEEE Comm. Magazine, September 2004.

[193] T. S. Rappaport. Wireless Communications: Principles and Practice. Prentice Hall, Inc., New Jersey, 2nd edition, 2002.

[194] G. Scutari, S. Barbarossa, and D. Ludovici. Cooperation Diversity in Multi-hop Wireless Networks Using Opportunistic Driven Multiple Access. In IEEE Signal Processing Workshop on Sig- nal Processing Adv. in Wireless Comm. (SPAWC), Rome, Italy, June 2003.

[195] A. Stefanov and E. Erkip. Cooperative Coding for Wireless Networks. In Proc. IEEE Conf. on Mobile and Wireless Communi- cation Networks, Stockholm, Sweden, September 2002.

[196] A. Stefanov and E. Erkip. Cooperative Information Transmission in Wireless Networks. In Proc. 2002 Asian-European Infor- mation Theory Workshop, Breisach, Germany, June 2002.

[197] A. Sendonaris, E. Erkip, and B. Aazhang. Increasing Uplink Capacity via User Cooperation Diversity. In IEEE Int. Symp. Inform. Theory (ISIT), Cambridge, MA, August 1998.

[198] A. Sendonaris, E. Erkip, and B. Aazhang. User Cooperation Diversity - Part I: System Description. IEEE Trans. Commun., 51(11):1927-1938, November 2003.

[199] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity, Part II: Implementation aspects and performance analysis. IEEE Trans. Commun., 51(11):1927-1938, November 2003.

[200] I. E. Telatar. Capacity of Multi-Antenna Gaussian Channels. Techn. report, Bell Labs, Lucent Technologies, October 1995.

[201] P. K. Vitthaladevuni and M. S. Alouini. BER computation of 4/M-QAM Hierarchical Constellations. IEEE Trans. Broadcast- ing, 47(3):228-239, September 2001.

[202] E. C. van der Meulen. Transmission of Information in a TTerminal Discrete Memoryless Channel. Technical report, Dept. of Statistics, Univ. of California, Berkeley, CA, 1968.

[203] E. C. van der Meulen. Three-Terminal Communication Channels. Adv. Appl. Prob., 3:120-154, 1971.

[204] M. C. Valenti and B. Zhao. Distributed turbo codes: Towards the capacity of the relay channel. In Proc. IEEE Vehic. Techn. Conf., October 2003.

[205] A. Wittneben and B. Rankov. Impact of Cooperative Relays on the Capacity of Rank Defficient MIMO Channels. In IST Mobile & Wireless Communications Summit 2003, Aveiro, Portugal, June 2003.

[206] Y. Xue and T. Kaiser. Cooperated TDMA Uplink Based on Hierarchically Modulated Alamouti Code. In Int. Zurich Seminar on Comm. (IZS), pages 200-203, Zurich, Switzerland, February 2004.

[207] E. Zimmermann, P. Herhold, and G. Fettweis. On the Performance of Cooperative Diversity Protocols in Practical Wireless Systems. In Proc. 58th IEEE Vehic. Technol. Conf., Orlando, FL, October 2003.

[208] E. Zimmermann, P. Herhold, and G. Fettweis. A Novel Protocol for Cooperative Diversity in Wireless Networks. In European Wireless Conf. 2004, Barcelona, Spain, February 2004.

[209] E. Zimmermann, P. Herhold, and G. Fettweis. Distributed Incremental Redundancy. In 13th Joint Workshop on Comm. and Coding (JWCC), Galt□ur, Austria, March 2004.

[210] E. Zimmermann, P. Herhold, and G. Fettweis. The Impact of Cooperation on Diversity-Exploiting Protocols. In Proc. 59th IEEE Vehic. Technol. Conf., Genoa, Italy, May 2004.

[211] Ernesto Zimmermann. Verfahren zur kooperativen Nutzung r□aumlicher Diversit□at im Mobilfunk. Diploma thesis, TU Dresden, April 2003.

[212] B. Zhao and M. C. Valenti. Cooperative Diversity Using Distributed Turbo Codes. In Virginia Tech Symp. on Wireless Per- sonal Commun., Blacksburg, VA, USA, June 2003.

[213] B. Zhao and M. C. Valenti. Distributed turbo coded diversity for the relay channel. IEE Electronics Letters, 39(10):786-787, May 2003.

[214] B. Zhao and M. C. Valenti. Some New Adaptive Protocols for the Wireless Relay Channel. In Proc. Allerton Conf. on Commun., Control, and Comp., Monticello, IL, USA, October 2003. From www.csee.wvu.edu/~mvalenti/documents/allerton prep.pdf.

[215] M. Dohler, Virtual Antenna Arrays, Ph.D. thesis, King's College, London, UK, 2003.

[216] A. Adinoyi, H. Yanikomeroglu, and S. Loyka, "Hybrid macro- and generalized selection combining microdiversity in lognormal shadowed Rayleigh fading channels", IEEE Int.'l Conf. on Communications (ICC'04), Paris, France, June 2004.

[217] A. Adinoyi and H. Yanikomeroglu, "On the performance of hybrid macro/microdiversity in the reverse-link microcellular networks", submitted to IEEE Wireless Communications and Networking Conf. 2005 (WCNC'05).

[218] P. Herhold, E. Zimmermann, and G. Fettweis, "A simple cooperative extension to wireless relaying", Int.'l Zurich Seminar on Communications, Zurich, Switzerland, 2004.

[219] M. Yuksel and E. Erkip, "Diversity in relaying protocols with amplify and forward", IEEE Globecom 2003, pp. 2025-2029.

[220] P. A. Anghel, G. Leus, M. Kaveh, "Multi-User Space-Time Coding in Cooperative Networks," in Proc. of IEEE ICASSP, Hong-Kong, April 6-10,2003.

[221] P. A. Anghel, G. Leus, and M. Kaveh, "Distributed Space-Time Coding in Cooperative Networks," In Proc. of the Nordic Signal Processing Symposium, Norway, October 4 -7, 2002.

[222] Yung-Szu Tu and Gregory Pottie, "Coherent Cooperative Transmission from Multiple Adjacent Antennas To a Distant Stationary Antenna Through AWGN Channels", in Proc. of IEEE VTC 2002, Birmingham, Alabama, USA, May 6-9, 2002.

[223] Z. Zeng, M. Dohler, H. Aghvami, "System Performance of a W-CDMA based Network with deployed Virtual Antenna Arrays," in Proc. of IEEE ICT2002, Beijing, June 23 - 26, 2002.

[224] M. Dohler, E. Lefranc, H. Aghvami, "Space-Time Block Codes for Virtual Antenna Arrays," in Proc. of IEEE PIMRC2002, Lisbon, Portugal, September 15 – 18, 2002.

[225] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications", IEEE JSAC, Vol. 16, No. 8, pp 1451-1458, Oct.1998.

[226] V. Tarokh, H. Jafarkhani, A. R. Calderbank, "Space–Time Block Codes from Orthogonal Designs", IEEE Trans. Inform. Theory, vol. 45, no. 5, pp.1456-1467, July

[227] P. Anghel, G. Leus, and M. Kaveh, "Multi-User Space-Time Coding in Cooperative Networks," Proc. of *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.

[228] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "On the Aggregate SNR of Amplified Relaying Channels," accepted to *IEEE GLOBECOM*, 2004.

[229] V. Emamian, P. Anghel, and M. Kaveh, "Multi-User Spatial Diversity in an Shadow-Fading Environment," Proc. of *IEEE Vehicular Technology Conference*, 2002.

[230] M. Hasna and M. Alouini, "End-to-End Performance of Transmission Systems with Relays over Rayleigh Fading Channels," *IEEE Trans. on Wireless Communications*, vol. 2, no. 6, 2003.

[231] P. Herhold, E. Zimmermann, G. Fettweis, "A Simple Cooperative Extension to Wireless Relaying," Proc. of *Int. Zurich Seminar on Communications*, 2004.

[232] J. Laneman and G. Wornell, “Energy-Efficient Antenna Sharing and Relaying for Wireless Networks,” Proc. of *IEEE Wireless Communications and Networking Conference*, 2000.

[233] J. Laneman and G. Wornell, “Exploiting Distributed Spatial Diversity in Wireless Networks,” Proc. of *Allerton Conference on Communications, Control, Computing*, 2000.

[234] J. Laneman, D. Tse, and G. Wornell, “Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour,” submitted to *IEEE Trans. on Information Theory*, 2002.

[235] J. Laneman, D. Tse, and G. Wornell, “Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour,” submitted to *IEEE Transactions on Information Theory*.

[236] A. Ribeiro, X. Cai, and G. Giannakis, “Symbol Error Probabilities for General Cooperative Links,” Proc. of *IEEE International Conference on Communications*, 2004.

[237] A. Sendonaris, E. Erkip, and B. Aazhang, “User Cooperation Diversity – Part I: System Description,” *IEEE Trans. on Communications*, vol. 51, no. 11, 2003.

[238] A. Sendonaris, E. Erkip, and B. Aazhang, “User Cooperation Diversity – Part II: Implementation Aspects and Performance Analysis,” *IEEE Trans. on Communications*, vol. 51, no. 11, 2003.

[239] L. Yang, H. Hasna, and M. Alouini, “Average Outage Duration of Multi-hop Communication Systems with Regenerative Relays,” Proc. of *IEEE Vehicular Technology Conference*, 2003.

[240] C. Hoymann, et al. “STRIKE D3.1.1: HIPERMAN/HIPERLAN/2 interworking methods”, Deliverable D3.1.1. of the IST STRIKE project

[241] 3GPP, TS 22.146 Stage 1 / TR 23.846 Stage 2, “MBMS Architectural and functional description”

[242] B. Walke, R. Pabst, D. Schultz, S. Mukherjee, H. Viswanathan, H. Yanikomeroglu et al., F. Fitzek, P. Herhold, N. Akhtar, R. Tafazolli. White paper "Relay-based Deployment Concepts for Wireless and Mobile Broadband Cellular Radio". WWRF/WG4 Relaying Subgroup, June 2003.

[243] IST-2001-38354 STRIKE European project within the Framework Program 5. D3.1.1 “HiperMAN-HiperLAN/2 interworking methods”, June 2003.

[244] http://www.ctr.kcl.ac.uk/Private/Mischa/publications.htm

[245] Mischa Dohler, Fatin Said, and Hamid Aghvami. Concept of Virtual Antenna Arrays. Proc. IEEE Globecom 2002, November 2002.

[246] Hongyi Wu, Chunming Qiao, Swades De, and Ozan Tonguz. Integrated cellular and ad hoc relaying systems: iCAR. IEEE Journal on Select Areas in Communications, 19(10):2105-2115, October 2001.

[247] Charles S. Leong, B.S. Lee, A.R. Nix, P. Strauch, J.P. McGeeHan. A 3D Pyramid Network for Improved Communications at 60 GHz. COST 273 TD(04)082 Gothenburg, Sweden 2004/June/9-10.

[248] Yozo Shoji and Hiroyo Ogawa ‘Experimental Demonstration of 622 Mbps Millimeter-Wave over Fiber Link for Broadband Fixed Wireless Access System,’ IEICE Trans.Electron., Vol. E86-C, No.7, July 2003.

[249] H. Al Raweshidi, S. Komaki (ed): Radio over Fiber, Artech House, London, Boston, 2002.

[250] A. Ng’oma. Broadband Radio@Hand, Deliverable D6.1. Design of a Radio-over-fibre System for Wireless LANs. May 2002.

[251] Gliese U., Nielsen T. N., Norskov S., and Stubkjaer K. E. ; “Multifunction Fibre-Optic Microwave Links Based on Remote Heterodyne Detection”; IEEE Trans.On Microwave theory and Techniques, Vol. 46, No. 5, May 1998.

[252] Fuster J. M., Marti J., Candelas P., Martinez F.J., Sempere L., “Optical Generation of Electrical Modulation Formats”, in Proceedings of the 27th European Conference on Optical Communication 2001 (ECOC’01), Amsterdam, Sept. 30 -Oct. 4, 2001, pp 536-537.

[253] http://www.fiber-optics.info/glossary-uv.htm.

[254] Efrein Gago, Luis Cucala, Víctor Iglesias and Manuel Gil (Telefonica Moviles España and Telefonica I+D), “Radio on Fibre systems for practical 2G and 3G deployment”. COST 273 TD(04)001 Athens, Greece 2004/January/26-28.