ESPOO 2004
VTT RESEARCH NOTES 2248

Efficient wireless networking with advanced services and negotiated QoS

VTT TIEDOTTEITA – RESEARCH NOTES 2248

Efficient wireless networking with advanced services and negotiated QoS

Tapio Frantti, Marko Jurvansuu & Aarne Mämmelä (eds.)

VTT Electronics

ISBN 951–38–6473–1 (URL: http://www.vtt.fi/inf/pdf/)
ISSN 1455–0865 (URL: http://www.vtt.fi/inf/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 4374

VTT, Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (09) 4561, fax (09) 456 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 4374

VTT Elektroniikka, Kaitoväylä 1, PL 1100, 90571 OULU
puh. vaihde (08) 551 2111, faksi (08) 551 2320

VTT Elektronik, Kaitoväylä 1, PB 1100, 90571 ULEÅBORG
tel. växel (08) 551 2111, fax (08) 551 2320

VTT Electronics, Kaitoväylä 1, P.O.Box 1100, FIN–90571 OULU, Finland
phone internat. + 358 8 551 2111, fax + 358 8 551 2320

Technical editing Marja Kettunen

Efficient wireless networking with advanced services and negotiated QoS. Ed. by Frantti, Tapio, Jurvansuu, Marko & Mämmelä, Aarne. Espoo 2004. VTT Tiedotteita – Research Notes 2248. 110 p.

Keywords Quality of Service (QoS), wireless network routing, positioning technologies, spectral efficiency, cellular systems

Abstract

In the AuRa project an efficient wireless networking scheme for mobile terminals with advanced services and negotiated quality of service (QoS) level was researched and developed. The project consisted of various topics in the various communication layers of the OSI model. In the physical and data link layers area spectral efficiency was researched. It was observed that log-normal shadowing is very crucial for the spectral efficiency. In outdoor positioning, advantages of wideband channel delay profiles were utilised so that database correlation method (DCM) algorithm can locate the terminal using only one base station hereby avoiding the problem common to observed time difference of arrival algorithm (OTDOA). For the indoor positioning a configuration tool was developed to make signal chart measurement faster with more dense and accurate grid. Hence, the typical accuracy of positioning in indoor of 2 m was achieved. In vertical handover topics signalling schemes as well as key parameters mapping and classification for the related QoS issues were developed in order to make vertical handover possible between the WLAN and UTMS networks. In the QoS research an end-to-end QoS measurement tool was also developed. From the measurements significant variations in end-to-end latency time of General Packet Radio Systems (GPRS) were observed between different operators and terminal types.

Network layer topics included power efficiency research on ad hoc and sensor networks taking into consideration mobility of nodes and the number of sequential hops in signalling and data delivery. Furthermore, methods and tools for parallel networking simulation were researched and the graphical user interface tool for the NS2 simulation tool was built with the aim of hiding the complexity of NS2 tool's scripts and distribution of simulation runs.

Preface

Project publication covers the research topics and results of the strategic research project AuRa of the Technical Research Centre of Finland (VTT) that spanned two years from the beginning of 2002 to the end of 2003. The AuRa project was one of the special Networking theme projects of VTT. The project was fully performed with the strategic funding and with VTT's own resources.

The main subjects in the AuRa project were *spectral and power efficiency of air interface techniques, location techniques of the broadband radio communication systems, vertical handovers and quality of service (QoS) between different kind of networks as well as power efficiency and routing in ad hoc networks.* The project consisted of several parts performed in several units and groups of VTT. VTT Electronics researched and developed all the above mentioned topics. VTT Information Technology researched outdoor positioning techniques in WCDMA based cellular networks. VTT Industrial Systems unit participated to the project via developing Bluetooth solutions to personal area networks (PAN). The purpose was to collect know-how and cumulated information from various units of VTT.

The research results published here were made possible thanks to the financial resources allocated by the general management of VTT. The effort of project group consisting of Janne Alasalmi, Dr. Tapio Frantti, M.Sc. Ville Haataja, Dr. Marko Jurvansuu, M.Sc. Suvi Juurakko, Lic. Tech. Pekka Koskela, M.Sc. Adrian Kotelba, M.Sc. Heikki Laitinen, M.Sc. Mikko Majanen, M.Sc. Marja Matinmikko, Prof. Aarne Mämmelä, M.Sc. Petri Määttä, Lic. Tech. Risto Nordman, M.Sc. Marko Palola, M.Sc. Seppo Rantala, M.Sc. Timo Sukuvaara, and M.Sc. Ville Typpö is especially acknowledged.

At Oulu, Finland, 30th January, 2004

Dr. Tapio Frantti, Dr. Marko Jurvansuu, Prof. Aarne Mämmelä

Editors

Contents

List of symbols

3GPP	Third Generation Partnership Project (ETSI)
AAA	Authentication, Authorisation and Accounting
ABR	Available Bit Rate
Ameba	ad hoc network simulation environment based on Beowulf cluster platform
ATM	Asynchronous Transfer Mode
AURA	autoconfigurative radiocommunication networking
BER	Bit Error Ration
BLUE	Name of active queue management algorithm
BTS	base station
CAC	Connection Admission Control
CBR	Constant Bit Rate
CDR	Connection Dropping Rate
CMMBCR	Conditional Max-Min Battery Capacity Routing
CN	Core Network
CPU	central processing unit
CS	coding scheme
Diffserv, DS	Differentiated services
DES	discrete event simulation
DNS	Domain Name Server
E	event
EAP	Extensible Authentication Protocol
EOF	End Of File
ETSI	European Telecommunications Standards Institute
FIFO	First In First Out
FTP	File Transfer Protocol
GFR	Guaranteed Frame Rate
GGSN	Gateway GPRS Support Node
GloMoSim	global mobile information systems simulation library
GPS	Global Positioning System

GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
GUI	graphical user interface
HTTP	HyperText Transport Protocol
IEEE	Institute of Electrical and Electronics Engineers
ICMP	Internet Control Message Protocol (IETF)
IntServ	Integrated Services
IP	Internet Protocol (IETF)
J2ME	Java 2 MicroEdition
LAM	local area multicomputer
LP	logical process
MAC	Medium Access Control
MPLS	Multi-Protocol Label Switching (IETF)
MOOSE	Maisie-based object-oriented simulation environment
MPI	message passing interface
MPICH	MPI Chameleon
NIC	Network Interface Card
Node B	UTRAN Base station
nrt-VBR	non-real-time Variable Bit Rate
ns-2	the Network Simulator, version 2
OPNET	optimized network engineering tools
OS	Operating System
OSI	Open System Integration
OTA	over-the-air provisioning
Parsec	parallel simulation environment for complex systems
PDCH	Packet Data Channel
PDES	parallel discrete event simulation
PDNS	parallel/distributed network simulator
PDP	Packet data Protocol
QoE	Quality of Experience

QoS	Quality of service
RAB	Radio Access Bearer
RED	Random Early Detection queue management
RNC	Radio Network Controller
RSVP	Resource ReSerVation Protocol (IETF)
rt -VBR	real-time Variable Bit Rate
RTS/CTS	Request To Send/Clear To Send
RRC	Resource Reservation Control
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SIM	Subscriber Identity Module
SIR	Signal Interference Ration
SLA	Service Level Agreement
SSF	scalable simulation framework
SSV	shared state variable
SV	state variable
SW	Software
TCL	tool command language
TCP	transmission control protocol (IETF)
TDMA	Time division multiple access
UBR	Unspecified Bit Rate
UMTS	Universal Mobile Telecommunications System
USIM	Universal Subscriber Identity Module (3GPP Working Group T3)
UTRAN	UMTS Radio Access Network
WEP	Wired Equivalent Privacy algorithm
Wi-Fi	Wireless Fidelity
WLAN	Wireless LAN
WPA	Wi-Fi Protected Access
WPAN	Wireless Personal Area Network

1. Introduction

This publication puts together state-of-art reports and achieved research results in the AuRa project of the Technical Research Centre of Finland (VTT) during the years 2002-2003. The AuRa project was focused on several important topics of wireless communications: *spectral efficiency of cellular systems*, *quality of service issues (QoS) in vertical handovers between the different networking technologies*, *positioning technologies in indoor and outdoor wireless networks*, *routing issues in ad hoc networks and graphical user interface (UI) development for wireless networks simulation environment*.

The purpose of this publication is to help researchers and designers to direct their activities towards better competitiveness in wireless communication area by utilising the expertise of VTT and by increasing their own view on these selected fields. Earlier research work is summarized in the roadmap publication "Communications Technologies – The VTT Roadmaps". The significance of the publication is based on the carefully chosen research topics that are in line with current international research of wireless community.

The radio spectrum is a limited resource and the basic resources and limitations to the QoS fulfilment are bandwidth, transmission power, delay, and the complexity of the system. Therefore, measures for spectral efficiency are important for the comparison of systems. Moreover, if different systems are compared with spectral efficiency, also offered QoS must be included. Section 2 of this publication covers the topic in detail.

Services of the network can be characterized by a QoS. Some of the services are reliable (i.e., they never loose data) and mostly they are implemented by having the receiver to acknowledge the receipt of data or by using redundancy or usually by combining these two methods in data transmission. However, this produces undesirable overhead and delays. Hence, quality of service issues can be enhanced in all the communication layers from the physical layer to the applications. QoS issues and vertical handovers with QoS parameter mapping are discussed in Section 3.

Traditionally, in telecommunication networks QoS measurement devices have been used for several years to measure network's quality of service. However, the user experience measurements have steadily increased their interests and the user experience measurements via use of connected terminals to services are nowadays required for the advanced QoS clarifications. Section 3 also covers the developed novel tool for the user experience measurement in QoS clarification.

Positioning techniques both in indoor and outdoor environment have also gained extensive interest among operators, product producers, and service providers to offer a new kind of services and to enhance existing services. Section 4 of the this publication offers positioning technologies overview in wireless indoor and outdoor networks.

The mobility of wireless hosts is one of the most hot research topic in the field. Usually, the coverage area is divided into centrally or base station controlled cells providing the terminals connectivity to the central control, which can be switched to the other cells if required (i.e., handovers are used). The other fundamental approach to the mobility of the hosts is independent, self-organising network composed of several terminals equipped with traffic routing properties. These kind of networks are called ad hoc networks. Section 5 concentrates on the wireless network routing issues, i.e., ad hoc routing. Section 5 also introduced the developed graphical user interface, especially, for the ad hoc network simulations.

The target of the project was to *specify a spectrally efficient networking scheme to support high mobility in wireless networks for advanced services, like location based functions, with a negotiated quality of service level*. Therefore, the project topics consist of physical layer, medium access layer, link layer, network layer, transport layer and application specific parts. In the physical layer area spectral efficiency, location definition algorithms and vertical handover topics were researched. On the medium access layer the focus was on the scheduling and related QoS issues. On link layer the aim was in the automatic repeat request research for the error correction and related QoS issues. Network layer topics includes efficient routing research, especially in the ad hoc networks, and on the transport layer were researched QoS issues. From the application point of view, the utilisation of location based services and other advanced services were researched and analysed. Section 6 (discussions) concludes the topics for the specification.

The purpose of the publication is to offer information from the main topics via embedded state-of-art review and analysis, proceed deeper to the fields with research results and finally form a view of the future on the fields. Hence, the organization of the rest of the report is following. In section 2 various area spectral efficiency related topics as fundamental limits for wireless networks are described and analysed. Section 3 concentrates on vertical handover and quality of service issues between the different networking technologies and Section 4 focus on indoor and outdoor positioning techniques in wireless networks. The efficient routing issues, especially on wireless ad hoc networks, are discussed in Section 5. Section 6 concludes the achieved results and present the view of the near future on wireless networking.

2. Spectral efficiency of cellular systems

Adrian Kotelba, Marja Matinmikko, Janne Alasalmi, Risto Nordman

The main issues in the development of future wireless communication systems will be high data rates and mobility of terminals. A major challenge is the efficient use of available resources to achieve the quality of service (QoS) requirements. The basic resources are bandwidth, transmission power, delay, and complexity. Radio spectrum and transmission power should be used efficiently while taking the complexity and delay constraints into account. Spectral and power efficiencies are included in the criteria for evaluating the radio transmission technologies for third generation (3G) wireless systems in International Telecommunication Union (ITU) recommendations (Rec. ITU-R M.1225). There are complicated trade-offs between the spectral and power efficiency of the system, complexity, and performance, for example bit error rate (BER). Fading and different forms of crosstalk, such as multiple-access interference (MAI), intersymbol interference, and intrasymbol interference, reduce the performance.

The radio spectrum is a limited resource, which limits the number of radio channels that can be available for users. In cellular systems, the same frequencies are reused in different cells, thus causing co-channel interference, which lowers the capacity of the channels. Measures for spectral efficiency are important as they allow comparison of existing and proposed systems, and estimate the ultimate capacity of the system. One such measure is *area spectral efficiency* (ASE) (Hatfield 1977, Alouini & Goldsmith 1999a). Without the ASE concept, conventional link level performance tests can be misleading since MAI is not properly taken into account. A literature review on ASE done in the AuRa project is given by Matinmikko (2002). If different cellular systems are compared with ASE, QoS offered by the systems must be included. Typically, quality requirements include the degree of coverage in terms of traffic or area, the grade of service in terms of waiting times or blocking probability, and interference levels.

2.1 Fundamental limits in wireless communications

The performance of wireless communication systems is limited by fundamental limits of nature. Available frequency spectrum is a limited resource, which gives a fundamental limit on the number of radio channels that can be made available for users. Propagation conditions constrict the available frequencies and limit the reliability of wireless communication systems. The fundamental limitations caused by radio propagation are considered in Crane (1981). Harmuth (1981) discusses the fundamental limits on the bandwidth of line-of-sight (LOS) radio systems. Fundamental limits in antennas are considered by Hansen 1981.

Atmospheric molecules, mainly oxygen and water vapor, cause molecular absorption of radio waves (Crane 1981). Attenuation increases as a function of frequency and experiences peaks at certain frequencies. Atmospheric conditions, such as rain, snow, and clouds, influence the attenuation of the radio waves. The attenuation increases in rain and fog as a function of the frequency. Attenuation by snow and other frozen particles is of little importance at frequencies less than 60 GHz. Hail can cause attenuation at frequencies as low as 1 GHz.

Fundamental limits also arise from natural and man made electronic *noise*. Noise from natural phenomena, such as atmospheric and thermal noise, is usually Gaussian while man made noise may be impulsive in nature. Man made noise arises from devices not designed to radiate electromagnetic energy, such as electrical motors, powerlines, ignition systems, and computers. Previously, the ignition system of motor vehicles was the major source of interference in mobile radio systems in urban areas due to lack of suppression methods. (Parsons 2000, pp. 263–265)

Information theory sets fundamental limits on the data rate that can be achieved in a channel. Fundamental limits caused by information theory are considered by Wyner 1981. *Shannon capacity* in bits per second gives the theoretical upper bound for the maximum transmission data rate at an arbitrarily small BER without any delay or complexity constraints. The Shannon capacity of the channel is limited by the transmission power and bandwidth. In information theory, the source entropy is the fundamental limit on the number of binary digits per symbol necessary to represent a source. Error-free encoding is possible if and only if the average number of bits used per source symbol is larger than or equal to the source entropy. Information theory has generally ignored the role of *delay* as a fundamental parameter in wireless communications.

Fundamental limits also arise from the time-variant nature of the wireless channel. As the terminal velocities and data rates keep growing, the measurability issues of the time-variant channel become important. There are fundamental limits on the estimation accuracy in estimating the channel conditions such as the Cramer-Rao bound. In addition, the non-linear phenomena in the parts of transmitter (TX) and receiver (RX) limit the performance.

2.2 Channel models

2.2.1 Radio wave propagation

Radio propagation channel is the main reason for many impairments in the wireless system’s performance. Characterization of the channel is therefore important in assessing the system performance. The basic radio propagation mechanisms both in indoor and outdoor environment are reflection, refraction, diffraction, and scattering. Indoor channel modelling includes propagation inside buildings as well as into buildings. Radio propagation models are usually empirical based on curve fitting on measurement data, or physical based on physical propagation mechanisms. In practice, all radio communication channels can be characterized as linear time-variant filters. For an overview of cellular radio channels see Mämmelä & Järvensivu (2002), Parsons (2000), Rappaport (1996), Saunders (1999).

Channel induced temporal and spatial variations of the received signal are divided into three types of fading: *path loss*, slow fading or *shadowing*, and fast fading or *multipath fading*. Path loss is the overall decrease in the signal strength with distance between TX and RX caused by spreading of the electromagnetic wave and obstructive effects of surrounding objects. Path loss causes large and slow signal variations. Shadowing is superimposed on the path loss resulting in slow random variations in the signal amplitude. Multipath fading causes fast random variations of the signal amplitude and phase due to mutual interference of multipath signal components.

The indoor radio propagation channel is surveyed in Hashemi (1993) and Molkdar (1991). In indoor environment, distances are much shorter and the variability of the environment is greater for a smaller range of the distance between TX and RX. Propagation inside buildings is strongly influenced by the layout of the building, the construction materials, and the building type. Indoor propagation suffers from interference caused by electronic equipment. The motion of people and equipment around the antennas causes temporal variations in the statistics of the indoor channel. Doppler shifts are often negligible because rapid motions and high velocities are absent inside buildings.

Signal transmission over the radio channel is often assumed to be *reciprocal*, i.e., the locations of TX and RX can be interchanged without changing the received signal strength. For a narrowband cellular radio channel the reciprocity requires that the uplink and downlink signals have the same carrier frequency and the physical environment and the locations of the base station and user terminals are static between uplink and downlink transmission (Kasapi et al. 2000, p. 297). However, the signal-to-noise ratio (SNR) is not reciprocal because noise is not reciprocal.

2.2.2 Link level models

Path loss is the ratio of the received signal power to the transmitted signal power, usually expressed in dB. The simplest path loss model is the free space loss in which attenuation increases exponentially with distance from TX to RX. Another fundamental path loss model is the *plane earth loss*, which assumes a direct path and a reflection from the ground (Saunders 1999, p. 96). Outdoor path loss models are summarized in Saunders (1999, pp. 151–175) and Rappaport (1996, pp. 110–122). Empirical models for *microcells* include the *dual-slope model* in which two separate path loss exponents are used to characterize the propagation (Saunders 1999, pp. 252–253). Physical path loss models for macrocells are presented in Saunders (1999, pp. 160–171), and for microcells in Saunders (1999, pp. 253–263). ITU has recommended certain path loss models for 3G systems in Rec. ITU-R M.1225. These models include pedestrian environment, vehicular environment, indoor office environment, and outdoor to indoor propagation.

Indoor channel experiences higher path losses and sharper changes in the mean signal level compared to the outdoor channel. Empirical path loss models for indoor channels are presented in Saunders (1999, pp. 278–282) and physical path loss models in Saunders (1999, pp. 278–282). Indoor channel modeling includes also propagation into buildings, which is summarized in Parsons (2000, pp. 191–195, 278–282) and Saunders (1999, pp. 274–278). Propagation into buildings is characterised with penetration loss, which depends on construction materials, building layout, and floor height. At lower floors, attenuation is larger while at higher floors a LOS path may exist (Rappaport 1996, p. 132).

The path loss predicted from the models is a constant for a given distance between TX and RX. Due to different propagation conditions along each path, the total path loss is a random variable and the channel causes *shadowing*. Shadowing is due to large terrain features, such as buildings and hills in macrocells and vehicles in microcells (Saunders 1999, pp. 180-182, 198). The path loss expressed in absolute units at a particular location is usually log-normally distributed around the distance dependent mean value. Log-normal distribution is used in oudoor, indoor, and outdoor to indoor environments to model shadowing (Saunders 1999, p. 264, p. 282). The signal levels in dB at a given distance have a Gaussian distribution about the mean. Log-normal distribution approximates the product of several random variables which represent the different attenuation factors (Mämmelä & Järvensivu 2002).

The multipath propagation of radio waves results in amplitude and phase variations in the received signal, i.e. *multipath fading*. The most important physical factors affecting multipath fading are reflecting objects in the channel, speed of the mobile, speed of surrounding objects, and the transmission bandwidth (Rappaport 1996, p. 139). The magnitude of the received signal envelope follows the Rayleigh distribution, if a large number of independent identically distributed random signal components arrive at RX. If the received signal is made up of multiple reflective rays and a significant LOS component, the channel is Rician fading. Nakagami distribution includes the Rayleigh distribution as a special case and can approximate the Ricean distribution. It is shown to give good fit to empirical data from mobile radio channels. Rayleigh and Rician distributions are used to model the received signal in both outdoor and indoor environment. In addition, Weibull distrubution and Suzuki distribution are sometimes applicable to indoor (Hashemi 1993).

The first multipath models were so called *classical models*, which focused on the time-domain aspects and did not properly model the spatial domain. Widely accepted classical models for outdoor, indoor, and outdoor to indoor environment proposed by ITU for 3G systems are presented in Rec. ITU-R M.1225 (1997). The multipath fading is characterized by a channel impulse response model, which is defined with the number of channel taps, time delay relative to the first tap, average power relative to the strongest tap, and Doppler spectrum of each tap. The statistical model for indoor multipath propagation by Saleh and Valenzuela (1987) has been widely accepted for indoor propagation modeling.

The classical models were applicable until quite recently when the need for directional models arose. *Spatial channel models* were developed to take into account the direction of the radio waves (see Ertel et al. 1998). Recently, the growing interest in multiple TX and RX antennas has led to the development of multiple-input multiple-output (MIMO) channel models. MIMO channels are appealing in rich multipath environment since much higher throughput can be achieved when many channels with lower gain are used instead of one channel with a high gain. In the early studies of the flat fading MIMO channels, the entries of the channel matrix, i.e. the fading coefficients between different antenna pairs, were assumed to be independent complex Gaussian random variables. For more realistic modeling, the correlation between the antennas has also been taken into account. The capacity of a MIMO channel is also sensitive to the rank behavior of the channel. Based on this idea, a non-line-of-sight (NLOS) MIMO channel model for outdoor environment was presented in Gesbert et al. (2000). This theoretical MIMO channel model includes an uncorrelated high rank model, an uncorrelated low rank model, and a correlated low rank model.

Multi-Element Transmit and Receive Antennas (METRA) project has developed a widely used MIMO channel model (Kermoal et al. 2002). The measurement data for validating the channel model was collected in indoor and indoor-to-outdoor MIMO channels, but the METRA model is applicable to outdoor propagation, too. Only a small set of parameters is needed to fully characterize the communication scenario in the proposed channel model. MIMO radio channels are presented as tapped delay lines whose taps are complex valued matrices. The work is continued in Intelligent METRA (I-METRA) project (http://www.ist-imetra.org/). Other MIMO channel modeling efforts include the COST 273 activity carried out in the framework of European Cooperation in the field of Scientific and Technical Research (COST).

2.3 Measures of spectral efficiency

2.3.1 Link spectral efficiency

The performance of a communications link can be evaluated with the link spectral efficiency (LSE), which is the information rate over the channel divided by the channel bandwidth for a given BER. An upper bound for the link spectral efficiency is obtained from the channel capacity equation by normalizing the capacity with the channel bandwidth. The Shannon capacity of a bandlimited additive white Gaussian noise (AWGN) channel in bits/s is

$$C = B\log_2\left(1 + P / BN_0\right) \tag{2.1}$$

where B is the channel bandwidth, P is the average transmission power, and N_0 is the noise power density. The Shannon capacity defines the theoretical upper bound for the data rate over the AWGN channel at an arbitrarily small BER without any delay or complexity constraints.

If the transmission rate can be varied, LSE is defined as the average transmitted data rate per unit bandwidth for a specified average transmit power and BER. In Lee (1990) the capacity of a Rayleigh fading channel with constant power policy was derived. The channel capacity is averaged over signal-to-noise ratio variations. The fading channel capacity per bandwidth is always lower than the capacity of an ideal AWGN channel. Günther (1996) extended Lee’s results to Ricean fading channels. The capacity of the Rayleigh fading channel was found to be very close to the capacity of the ideal AWGN channel.

The capacity of a single-user flat fading channel with ideal channel measurement information at TX and RX for various adaptive transmission policies was derived in Goldsmith & Varaiya (1997). The results were extended and closed form solutions for the Rayleigh fading channel capacity with different adaptive transmission and diversity-combining techniques were derived in Alouini & Goldsmith (1999b). In a multiuser channel, the capacity is evaluated with the *Shannon capacity region* which gives the maximum data rates at which the users can simultaneously receive data with an arbitrarily small error probability (El Gamal & Cover 1980). The capacity region gives an upper bound for the capacity of a single cell in a multicell environment. In Goldsmith (1997) the Shannon capacity region of the downlink channel of a single cell was considered.

Communications with multiple TX and RX antennas through MIMO channels appears as a spectrally efficient communications method in rich multipath environment. In the case of a single user with independent fading between antenna pairs and channel state information only at RX, the capacity increases linearly with the smaller of the number TX and RX antennas. With no channel state information in a piece-wise constant channel model, where the channel is assumed constant for T symbol intervals and then changes to an independent value, there is no use making the number of TX antennas greater than T (Marzetta & Hochwald 1999). With multiple users, different users must transmit independent signals. When channel state information is available at RX only, users can collectively achieve the throughput of a single user. With channel state information at both TX and RX, maximum throughput is obtained by allowing the user with the strongest propagation coefficient transmits during each coherence interval. Laboratory demonstrations of a MIMO link have shown high link spectral efficiencies in the order of 20–40 bits/s/Hz (Golden et al. 1999). In realistic propagation environments, fades between different antenna pairs are correlated which leads to capacity reduction.

2.3.2 Area spectral efficiency

The link level performance studies can be misleading when the total system performance is concerned. The Shannon capacity of a multicell system is an open problem because the frequency reuse and the corresponding interference models are difficult to include in Shannon's work (Pottie 1995, Goldsmith 1997). Frequency-reuse in cellular systems introduces co-channel interference, which ultimately determines the data rates and bit error rates available to the users. While the frequency-reuse improves the channel efficiency in channels per Hz, increased co-channel interference decreases the capacity of these channels. The effects of co-channel interference are commonly assed with the *outage probability*, i.e. the probability of co-channel interference (Cox 1982). Outage occurs if signal-to-interference ratio (SIR) is below some acceptable level, which causes the failure to obtain satisfactory reception.

The area spectral efficiency (ASE) is an extension of LSE to measure the system performance of cellular systems. Hatfield (1977) defined ASE as the carried traffic for the unit sizes of the radio bandwidth and service area (Erlangs/Hz/m^2). ASE was studied in analog speech systems in the 1970's, in digital speech systems in 1990's, and digital data systems from late 1990's. Single antenna systems are well covered, while the system performance of multiple antenna systems is currently under study. Coverage efficiency is also an important design consideration because the cost of cellular systems depends heavily on the number of base stations. Coverage efficiency is defined as the inverse of the number of base stations per km^2 needed to meet the coverage requirements in (Rec. ITU-R M.1225).

Early studies of ASE assumed equal resources for all users. Alouini and Goldsmith (1999a) considered ASE of cellular data systems in which the mobiles adapt their transmission rates according to fading and interference conditions. Adaptive transmission improves the throughput in fading channels, but requires accurate channel estimation and reliable feedback between RX and TX. ASE of cellular data systems is defined as the sum of the maximum average user data rates per unit bandwidth and unit area (bits/s/Hz/m^2) for a specified BER. Alouini and Goldsmith (1999a) considered the uplink of a cellular frequency division multiple access (FDMA) or time division multiple access (TDMA) system. The system model is shown in Figure 2.1. Users were randomly located in their cells and the signal propagates through a slowly varying flat-fading channel which is characterized with two-slope path loss model, log-normal shadowing, and multipath fading following the Nakagami distribution. The system is assumed to be interference-limited so that the thermal noise power is negligible relative to the co-channel interference power. Co-channel interferers outside the first dominant tier of interfering cells and interchannel interference are neglected. The fluctuations in the SIR are assumed to be perfectly tracked at the base station RX and fed back to the transmitting mobiles via an error-free feedback channel with negligible time delay. Perfect adaptation to the actual SIR is assumed so that the data rates can be presented with the Shannon capacity. Therefore, ASE is not parameterized by BER.

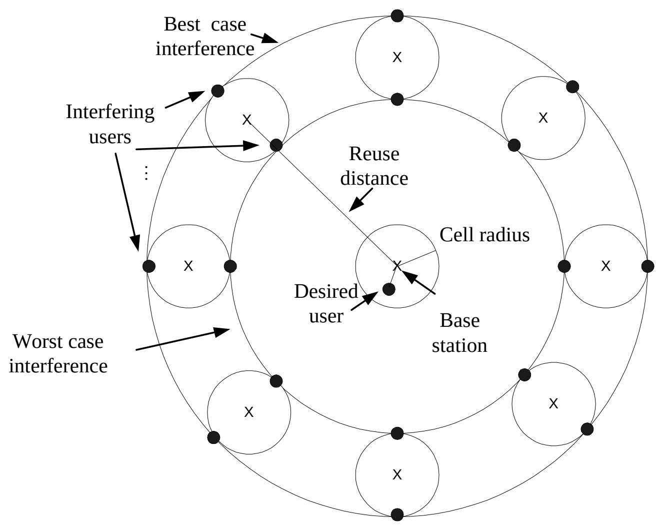

Figure 2.1. System model for ASE evaluation.

ASE was first derived for fully loaded systems in which all the channels in the cells are fully used and the number of interferers is constant. The results were also extended to partially loaded systems. ASE is approximated by

$$A = \frac{\sum_{k=1}^{K} C_k}{B\pi\left(D/2\right)^2} \tag{2.2}$$

where K is the number of active serviced channels per cell, C_k is the maximum data rate of the kth user in the cell, B is the bandwidth allocated to the cell, and D is the reuse distance between co-channel cells. The sum of the bit rates is divided by the cell bandwidth and area covered with one set of frequencies. In the fading channel, capacity is averaged. If all the users are assigned the same bandwidth, capacities are equal and ASE as the sum of the maximum average data rates per unit bandwidth and unit area becomes

$$A = \frac{4}{\pi D^2} \int \log_2 (1+\gamma) p(\gamma) d\gamma \, . \tag{2.3}$$

Now, ASE is not parameterized by the number of users. The analysis is simplified by computing SIR and the corresponding ASE for only the worst case and best case interference conditions. In the worst case, all co-channel interferers are on the near boundary of their respective cells. In the best case, co-channel interferers are on the far boundary of their cells. In practice, the ASE is between the two values, which as shown with Monte Carlo simulations (Alouini and Goldsmith 1999a, Alasalmi 2002).

Alouini and Goldsmith (1999a) define ASE as the sum of the maximum data rates divided by the bandwidth and the coverage area of the cell's base station which does not give fair comparison of the spectral efficiency if the cell sizes are different. As the cost of cellular systems depends heavily on the number of base stations, instead of expressing the ASE per m^2, ASE per cell is more appropriate. The aim in Alouini & Goldsmith (1999a) is to maximize the sum of the data rates while no attention is paid to the distribution of the data rates between different users. In fact, one user closest to the base station could transmit with the highest data rate while others would not transmit at all. The distribution of data rates for a given user is not included in ASE definition although it is important in practise. The Shannon capacity equation is used to determine the data rates of the users, which implies that the ASE calculations are not parameterized by delay or BER.

The spectral efficiency of single antenna systems is well covered. The spectral efficiency of systems with one transmitter antenna and multiple receiver antennas is also covered (Winters et al. 1994). Most of the previous work on spectral efficiency of MIMO systems has concentrated on a single link with no external interference. However, studies of system performance of MIMO have appeared recently. Catreux et al. (2002) studied ASE of a MIMO system with adaptive transmission in noise-limited cellular environment which corresponds to either an isolated cell or a multicell system in which the interference from outside the cell is small compared to thermal noise. The MIMO channels increased the average ASE significantly. In Catreux et al. (2001) and Farrokhi et al. (2002) the spectral efficiency of MIMO systems with adaptive transmission in interference limited cellular environment was studied. The system performance is measured in terms of average data throughput normalized by the bandwidth when the averaging is done over user location, shadowing, and fast fading. The performance of the MIMO system is compared to SISO systems and SIMO systems and the results indicate that the average spectral efficiency in bits/s/Hz is higher in MIMO systems.

Power efficiency is also an important design criterion in future wireless systems. The aim is to conserve battery power at mobile stations and to minimize interference to other users by minimizing both transmission power and signal processing power while satisfying the quality of service requirements (Bambos 1998). Many techniques in the link layer that reduce transmission power require significant amount of signal processing and thus may not lead to any power savings. Therefore, energy-constrained systems should develop energy efficient processing techniques that minimize power requirements across all levels of the protocol stack and minimize messages for network control (Goldsmith & Wicker 2002). Minimizing the average transmission power per user is studied in Bambos (1998) and Zorzi and Rao (1997). Zorzi and Rao (1997) defined the *energy efficiency* in wireless packet data systems as the total amount of data delivered divided by the total energy consumed under quality of service constraints. Goodman and Mandayam (2000) defined a similar measure, the *utility*, as the number of information bits received per Joule of energy expended. Goldsmith and Wicker (2002) considered the design challenges of energy-constrained wireless ad hoc networks. Power conscious design of wireless circuits and systems is thoroughly discussed in Abidi et al. (2000).

2.4 Methods to improve spectral efficiency

ASE of cellular systems can be increased by using either optimal or experimental methods. Optimization of a cellular network is a very complicated multidimensional and nonlinear problem with certain constraints, which implies the use of differential calculus. Analytic maximization can be done in very simple cases only (Qiu & Chawla 1999), and therefore some approximate and iterative methods are needed. Another and more traditional approach is to use experimental methods. The basic idea is to develop a tool with which we can estimate ASE for a selected set of signals, which are then compared in terms of ASE. The global optimum is not found since not all possible signals can be tested. To estimate ASE for a certain signal set, we must use approximate semianalytical methods comprising both analysis and simulations.

Pottie (1995) outlines several communications techniques in fading channels. The approaches are broadly categorized into fixed design for the worst case, interference averaging, interference avoidance, and interference suppression. Design for the *worst case* implies the use of techniques that provide margin against noise. These techniques include higher power, lower bit rate, error control coding, and clever modulation. They are potentially expensive and result in limited gain. Combined time, frequency, and antenna diversity are possible approaches for interference *averaging*. Interference averaging is useful when accurate SIR estimation is not possible. Instead of designing for the worst case, information can be spread over time, frequency or space so that channel looks like an average SIR case. Examples of this include spread spectrum systems, channel coding with interleaving, automatic repeat request schemes, and equal gain combining.

When RX can estimate SIR, coherent combining techniques, interference cancellation techniques, and improved channel coding schemes can be used. Interference *avoidance* is possible if the channel fading is slow enough to allow accurate estimation and feedback of channel conditions. Interference avoidance methods include dynamic channel and power allocation, adaptive transmission, and adaptive TX and RX antenna arrays. Adaptive transmission provides performance improvement in fading channels. Transmission power level, transmission rate, constellation size, coding rate or scheme, or any combination of these parameters can be varied to match the channel conditions. Interference *suppression* is obtained with antenna arrays, cell sectoring, joint estimation of signals, and macro diversity. (Pottie 1995, Abidi et al. 2000)

Communications in wireless MIMO channels is very attractive due to the large capacity gains promised by the technique in rich scattering. Therefore, the key issue in improving ASE of cellular systems will be the use of multiple TX and RX antennas as shown in Catreux et al. (2001) and Farrokhi et al. (2002).

2.5 Analytical and simulation results

2.5.1 Link level performance

2.5.1.1 Multi-antenna link model

Let us consider a single user narrowband Gaussian channel with t transmit and r receive antennas. The complex channel coefficients between the ith transmit and the jth receive antenna, denoted by g_{ij}, are assembled into a channel matrix

$$\mathbf{G} = \left[g_{ij}\right]_{i,j=1}^{r,t} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1t} \\ g_{21} & g_{22} & \cdots & g_{2t} \\ \vdots & \vdots & & \vdots \\ g_{r1} & g_{r2} & \cdots & g_{rt} \end{bmatrix}. \tag{2.4}$$

The matrix entries g_{ij} represent joint effect of shadowing and multipath fading. Since the decorrelation distance of shadows is usually large compared to antenna array dimensions, we assume that the shadowing effect is spatially deterministic from one antenna to another. On the other hand, multipath components are very likely to be uncorrelated between antennas. Consequently, the effects of shadowing and multipath fading can be separated and the channel matrix $\mathbf{G}$ becomes

$$\mathbf{G} = \sqrt{s}\left[h_{ij}\right]_{i,j=1}^{r,t} \tag{2.5}$$

where h_{ij} encompass the effect of random multipath fading and $\sqrt{s}$ represents the shadowing effect.

Let $\mathbf{x} \in \mathrm{C}^t$ denote the transmitted complex signal vector and $\mathbf{y} \in \mathrm{C}^r$ denote the received complex signal vector. The received $\mathbf{y}$ and transmitted $\mathbf{x}$ signal vectors are related by

$$\mathbf{y} = \sqrt{s}\mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.6}$$

where $\mathbf{n} \in \mathrm{C}^r$ is vector of complex additive white Gaussian noise samples. The spatial covariance matrix of the transmitted signal $\mathbf{x}$ is defined as $\mathbf{P}_x = E\left[\mathbf{x}\mathbf{x}^*\right]$. Furthermore, the total transmit power is limited to P_{av} irrespective of the number of transmit antennas, that is, $\mathrm{tr}\left(\mathbf{P}_x\right) = P_{\mathrm{av}}$. The symbol $(\cdot)^*$ denotes the conjugate transpose operation. We assume that the noise is spatially white and therefore the spatial covariance matrix of the noise is a diagonal matrix with noise variance σ_n^2 per receiver antenna, i.e., $E\left[\mathbf{n}\mathbf{n}^*\right] = \sigma_n^2\mathbf{I}_r$.

Shadowing

We assume that the random shadowing gain follows a log-normal distribution. The probability density function of shadowing gain s is

$$p(s) = \frac{\xi}{\sqrt{2\pi}\sigma_s s}\exp\left[-\left(\frac{\xi \ln s - \mu}{\sqrt{2}\sigma_s}\right)^2\right] \tag{2.7}$$

where $\mu = -\sigma_s^2 / 2\xi$ to ensure that $E[s] = 1$. The shadowing standard deviation is denoted by σ_s and $\xi = 10/\ln 10$.

A multipath propagation scenario is obtained by reducing the shadowing standard deviation σ_s to zero. Therefore, the probability density function of shadowing gain

$p(s)$ becomes strongly peaked around its mean value $E[s]$ and in the limit as $\sigma_s \to 0$ it becomes delta function

$$p(s)\big|_{\sigma_s=0} = \delta(s - E[s]) = \delta(s-1). \tag{2.8}$$

Multipath fading

Let us consider a multi-antenna channel with uncorrelated Rayleigh fading. In this case, the channel matrix $\mathbf{H}$ is a random matrix with independent and identically distributed (i.i.d.) complex zero-mean Gaussian entries of unit variance.

Let $m = \min(r,t), n = \max(r,t), \alpha = n - m$, and

$$\mathbf{W} = \begin{cases} \mathbf{HH}^* & : \quad r \le t \\ \mathbf{H}^*\mathbf{H} & : \quad r > t \end{cases}, \tag{2.9}$$

then the eigenvalues of $m \times m$ positive definite matrix $\mathbf{W}$ are distributed according to the probability density function

$$p(\lambda_1, \ldots, \lambda_m) = \frac{c}{m!}\left\{\det\left[\left(\lambda_j^{\alpha} e^{-\lambda_j}\right)^{1/2} \lambda_j^{i-1}\right]_{i,j=1}^{m}\right\}^2 \tag{2.10}$$

where c is a normalization constant

$$c = \left[\prod_{k=1}^{m} \Gamma(n-k+1) \prod_{k=1}^{m} \Gamma(m-k+1)\right]^{-1}. \tag{2.11}$$

The probability density function $p(\lambda)$ of a single eigenvalue λ of $\mathbf{W}$ is (Telatar 1999)

$$p(\lambda) = \frac{1}{m} \sum_{k=0}^{m-1} \frac{k!}{(k+\alpha)!} \left[L_k^{\alpha}(\lambda)\right]^2 \lambda^{\alpha} e^{-\lambda} \tag{2.12}$$

where $L_k^{\alpha}(\lambda)$ is the associated Laguerre polynomial of order *k*.

Let us consider a multi-antenna channel with correlated Rayleigh fading. In this case, the channel matrix $\mathbf{H}$ is a random matrix with complex zero-mean Gaussian entries with covariance matrix

$$\operatorname{cov}\mathbf{H} = \mathbf{M}_t \otimes \mathbf{M}_r. \tag{2.13}$$

Matrix $\mathbf{M}_r \in \mathrm{C}^{r \times r}$ is correlation matrix of any pair of columns of $\mathbf{H}$ and thus represents correlation between receiver antennas. Similarly, matrix $\mathbf{M}_t \in \mathrm{C}^{t \times t}$ is correlation matrix of any two rows of $\mathbf{H}$ and represents correlation between transmitter antennas.

In general, the joint probability density function of the positive eigenvalues $\lambda_1, \ldots, \lambda_m$ of matrix $\mathbf{W}$ is unknown except for special cases where one of the covariance matrices is an identity matrix.

Let us consider a multi-antenna channel with semi-correlated Rayleigh fading. In this scenario, we assume that either transmit or receive antennas are correlated. Consequently, the channel matrix $\mathbf{H}$ is a complex Gaussian random matrix with covariance matrix $\operatorname{cov}\mathbf{H} = \mathbf{I}_t \otimes \mathbf{M}_r$, if the correlation exists between receiver antennas, or $\operatorname{cov}\mathbf{H} = \mathbf{M}_t \otimes \mathbf{I}_r$ when transmitter antennas are correlated.

The joint probability density function of unordered eigenvalues $\lambda_1, \ldots, \lambda_m$ of matrix $\mathbf{W}$ is given in a simple form only when the covariance matrix of the correlated side is $m \times m$ matrix. Namely, the joint probability density function of the positive eigenvalues $\lambda_1, \ldots, \lambda_m$ of $\mathbf{W}$ is

$$p(\lambda_1, \ldots, \lambda_m) = \frac{(-1)^{m(m-1)/2}}{m! \prod_{k=1}^{m} \Gamma(n-k+1)} \frac{\det\left[e^{-\lambda_j/\phi_i}\right]_{i,j=1}^{m} \det\left[\lambda_j^{n-i}\right]_{i,j=1}^{m}}{\det\left[\phi_j^{\alpha+i}\right]_{i,j=1}^{m}}. \tag{2.14}$$

where $\phi_j, j = 1,2,\ldots,m$ are distinct eigenvalues of covariance matrix of the correlated side.

2.5.1.2 Average link spectral efficiency

Channel capacity in Shannon’s sense is defined as the maximum value of mutual information between channel input and channel output. This capacity concept is based on assumption that the error probability can be driven to zero by increasing the transmission blocklength. This capacity notion was widely used to determine the maximum possible transmission rate in time-invariant AWGN channels. However, Shannon introduced similar definition for a channel with time-varying state (Shannon 1958). Capacity of such a channel is expressed in terms of the ordinary means for memoryless channels. Observe that channel capacity cannot be determined unless we assume that channel fading process is ergodic. Ergodicity means that statistical properties of a fading process can be found based on a single realisation. In particular, the channel capacity is obtained by averaging the instantaneous channel capacity over all possible channel states.

Single-antenna channel

In a composite multipath/shadowing environment the probability density function of received signal-to-noise ratio γ has the form

$$p(\gamma)=\frac{\xi}{\sqrt{2\pi}\sigma_s}\int_0^\infty \frac{1}{s^2}\exp\left[-\left(\frac{\xi\ln s-\mu}{\sqrt{2}\sigma_s}\right)^2-\frac{\gamma}{s}\right]ds\,. \tag{2.15}$$

The average link spectral efficiency of the transmission through single-antenna channel with channel state information at the receiver only is Kotelba (2004)

$$\frac{C_{rx}}{B}=\frac{\xi}{\sqrt{2\pi}\sigma_s\ln 2}\int_0^\infty \frac{\Gamma(0,1/s)}{s}\exp\left[-\left(\frac{\xi\ln s-\mu}{\sqrt{2}\sigma_s}\right)^2+\frac{1}{s}\right]ds\,. \tag{2.16}$$

The average link spectral efficiency of the transmission through single-antenna channel with channel state information at the transmitter and receiver is (Kotelba 2004)

$$\frac{C_{rx}}{B}=\frac{\xi}{\sqrt{2\pi}\sigma_s\ln 2}\int_0^\infty \frac{\Gamma(0,\nu/s)}{s}\exp\left[-\left(\frac{\xi\ln s-\mu}{\sqrt{2}\sigma_s}\right)^2\right]ds \tag{2.17}$$

where ν is the optimal cut-off value that satisfies the average power constraint (Kotelba 2004)

$$\int_\nu^\infty \left(\nu^{-1}-\gamma^{-1}\right)p(\gamma)d\gamma=1\,. \tag{2.18}$$

Multi-antenna channel

The average link spectral efficiency of the transmission through multiple-antenna channel is (Telatar 1999)

$$\begin{aligned}\frac{C}{B} &= E\left[\log_2\det\left(\mathbf{I}_r+\frac{s}{\sigma_n^2}\mathbf{HP}_x\mathbf{H}^*\right)\right]\\ &= \int_0^\infty\cdots\int_0^\infty \log_2\det\left(\mathbf{I}_r+\frac{s}{\sigma_n^2}\mathbf{HP}_x\mathbf{H}^*\right)p(\lambda_1,\ldots,\lambda_m,s)d\lambda_1\cdots d\lambda_m ds\end{aligned} \tag{2.19}$$

Since shadowing and multipath fading are independent, the joint probability density function of eigenvalues and shadowing gain is the product of respective probability density functions, i.e., $p(\lambda_1,\ldots,\lambda_m,s)=p(\lambda_1,\ldots,\lambda_m)p(s)$. Consequently,

$$\frac{C}{B}=\int_0^\infty\cdots\int_0^\infty \log_2 \det\left(\mathbf{I}_r+\frac{s}{\sigma_n^2}\mathbf{H}\mathbf{P}_x\mathbf{H}^*\right)p(\lambda_1,\ldots,\lambda_m)d\lambda_1\cdots d\lambda_m\, p(s)ds\,. \tag{2.20}$$

Let us assume that only the receiver has perfect knowledge about the multi-antenna channel. Since the transmitter has no access to the channel state information, the optimal power allocation strategy is to equally distribute power among all transmitter antennas (Telatar 1999):

$$\mathbf{P}_x=\frac{P_{\text{av}}}{t}\mathbf{I}_t\,. \tag{2.21}$$

Thus, the average link spectral efficiency is

$$\frac{C_{\text{rx}}}{B}=E\left[\log_2\det\left(\mathbf{I}_r+\frac{sP_{\text{av}}}{t\sigma_n^2}\mathbf{H}\mathbf{H}^*\right)\right]=E\left[\sum_{k=1}^m\log_2\left(1+\frac{sP_{\text{av}}}{t\sigma_n^2}\lambda_k\right)\right] \tag{2.22}$$

where we used the fact that matrix determinant is the product of matrix eigenvalues.

The average link spectral efficiency with uncorrelated multipath fading is

$$\frac{C_{rx}}{B}=\int_0^\infty\cdots\int_0^\infty\sum_{k=1}^m\log_2\left(1+\frac{sP_{\text{av}}}{t\sigma_n^2}\lambda_k\right)p(\lambda_1,\ldots,\lambda_m)d\lambda_1\cdots d\lambda_m\, p(s)ds \tag{2.23}$$

where the joint probability density function of eigenvalues is given by (2.10). By using Haagerup & Thorbjørnsen (2003) Proposition 5.4 we obtain

$$\frac{C_{rx}}{B}=m\int_0^\infty\int_0^\infty\log_2\left(1+\frac{sP_{\text{av}}}{t\sigma_n^2}\lambda\right)p(\lambda)d\lambda p(s)ds \tag{2.24}$$

where $p(\lambda)$ is the probability density function of a single eigenvalue given by (2.12). Note that in a multipath fading environment (i.e., when $\sigma_s \to 0$) the average link spectral efficiency is (cf. Telatar 1999)

$$\frac{C_{rx}}{B}=m\int_0^\infty\log_2\left(1+\frac{sP_{\text{av}}}{t\sigma_n^2}\lambda\right)p(\lambda)d\lambda\,. \tag{2.25}$$

The double integral in (2.24) can be evaluated numerically using Laguerre-Gauss and Hermite-Gauss quadratures for inner and outer integral, respectively.

The joint probability density function of positive eigenvalues of matrix **W** when both antenna arrays are correlated is generally unknown and thus precludes the derivation of closed-form analytical expression for average link spectral efficiency. The average link spectral efficiency can only be estimated using Monte Carlo methods.

Let us assume that the correlation between antennas exists on one side of the link. Furthermore, let the correlated side be equipped with *m* antennas. The average link spectral efficiency with receiver channel state information is given by (2.23) where the joint probability density function of eigenvalues is defined in (2.14). By using Chiani et al. (2003) Theorem 3 we obtain

$$\frac{C_{rx}}{B} = \frac{(-1)^{m(m-1)/2}}{\det\left[\phi_j^{\alpha+i}\right]_{i,j=1}^{m} \prod_{k=1}^{m} \Gamma(n-k+1)} \int_0^\infty \sum_{k=1}^{m} \det \Upsilon_k(s) p(s) ds \tag{2.26}$$

where the entries of matrix $\Upsilon_k(s)$ are

$$\upsilon_{ij}(k,s) = \begin{cases} \int_0^\infty \log_2\left(1 + \dfrac{sP_{\mathrm{av}}}{t\sigma_n^2}\lambda\right)\lambda^{n-j} e^{-\lambda/\phi_i} d\lambda & : \quad i = k \\ \phi_i^{n-j+1}\Gamma(n-j+1) & : \quad i \neq k \end{cases}. \tag{2.27}$$

In a multipath Rayleigh fading scenario, the average link spectral efficiency reduces to (cf. Kang & Alouini 2003)

$$\frac{C_{rx}}{B} = \frac{(-1)^{m(m-1)/2}}{\det\left[\phi_j^{\alpha+i}\right]_{i,j=1}^{m} \prod_{k=1}^{m} \Gamma(n-k+1)} \sum_{k=1}^{m} \det \Upsilon_k(1) \tag{2.28}$$

The integrals in (2.27) and (2.26) can be evaluated numerically using Laguerre-Gauss and Hermite-Gauss quadratures.

Let us assume that the channel state information is available at transmitter and receiver side, then the average link spectral efficiency is

$$\frac{C_{rxtx}}{B} = \int_0^\infty \cdots \int_0^\infty \sum_{k=1}^{m} \log_2\left[1 + \frac{s\lambda_k}{\sigma_n^2} P(\lambda_k)\right] p(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m\, p(s) ds \tag{2.29}$$

where $\mathbf{P}_x = \mathrm{diag}\left[P(\lambda_1), P(\lambda_2), \ldots, P(\lambda_m)\right]$ is a diagonal matrix with entries

$$P(\lambda_k) = \left(\frac{1}{\nu} - \frac{1}{\lambda_k s}\right)^+ = \begin{cases} 0 & : \quad s\lambda_k \leq \nu \\ \nu^{-1} - (\lambda_k s)^{-1} & : \quad s\lambda_k > \nu \end{cases}. \tag{2.30}$$

The available transmission power is divided among transmit antennas with respect to the average power constraint

$$\begin{aligned} P_{\mathrm{av}} &= E\left[tr\left(\mathbf{P}_x\right)\right] \\ &= \int_0^{\infty}\cdots\int_0^{\infty}\sum_{k=1}^{m}\left(\frac{1}{\nu}-\frac{1}{\lambda_k s}\right)^{+} p\left(\lambda_1,\ldots,\lambda_m\right)p\left(s\right)ds \end{aligned} \tag{2.31}$$

Let us assume that the multipath fading is uncorrelated. The optimal cut-off value ν is found by solving the average power constraint (2.31). By using Haagerup-Thorbjørnsen (2003) Proposition 5.4 we obtain

$$E\left[\sum_{k=1}^{m}\left(\frac{1}{\nu}-\frac{1}{\lambda_k s}\right)^{+}\right] = m\int_0^{\infty}\int_{\nu/s}^{\infty}\left(\frac{1}{\nu}-\frac{1}{\lambda s}\right)p(\lambda)d\lambda p(s)ds \tag{2.32}$$

where $p(\lambda)$ is the probability density function of a single eigenvalue given by (2.12). There is no closed-form analytical expression for optimal cut-off value ν and it has to be found by numerical search techniques.

The average link spectral efficiency with transmitter and receiver channel state information is obtained by using Haagerup-Thorbjørnsen (2003) Proposition 5.4. Namely,

$$\frac{C_{rxtx}}{B} = m\int_0^{\infty}\int_{\nu/s}^{\infty}\log_2\left(\frac{s\lambda}{\nu\sigma_n^2}\right)p(\lambda)d\lambda p(s)ds\,. \tag{2.33}$$

In a multipath Rayleigh fading environment the average link spectral efficiency is (cf. Jayaweera & Poor 2003)

$$\frac{C_{rxtx}}{B} = m\int_{\nu/s}^{\infty}\log_2\left(\frac{s\lambda}{\nu\sigma_n^2}\right)p(\lambda)d\lambda\,. \tag{2.34}$$

We plot the average link spectral efficiency in a multi-antenna channel with uncorrelated Rayleigh fading and shadowing in Figs 2.2 and 2.3. The shadowing standard deviation was set to 8 dB, which is a typical value for macrocells. The results suggest that using channel state information at the transmitter is always advantageous in a sense that it allows higher average link spectral efficiencies. However, the advantage of using channel state information at the transmitter is less pronounced when the signal-to-noise ratio is sufficiently high. In general, random shadowing significantly reduces the average link spectral efficiency. The impact of correlation between antennas on the average link spectral efficiency is shown in Fig. 2.4. The numerical results show that the average link spectral efficiency gets smaller when the correlation between antennas increases. Furthermore, systems operating at high signal-to-noise ratio are more sensitive to correlation than those operating at low signal-to-noise ratios, i.e., the capacity loss due to correlation is significantly smaller for latter systems.

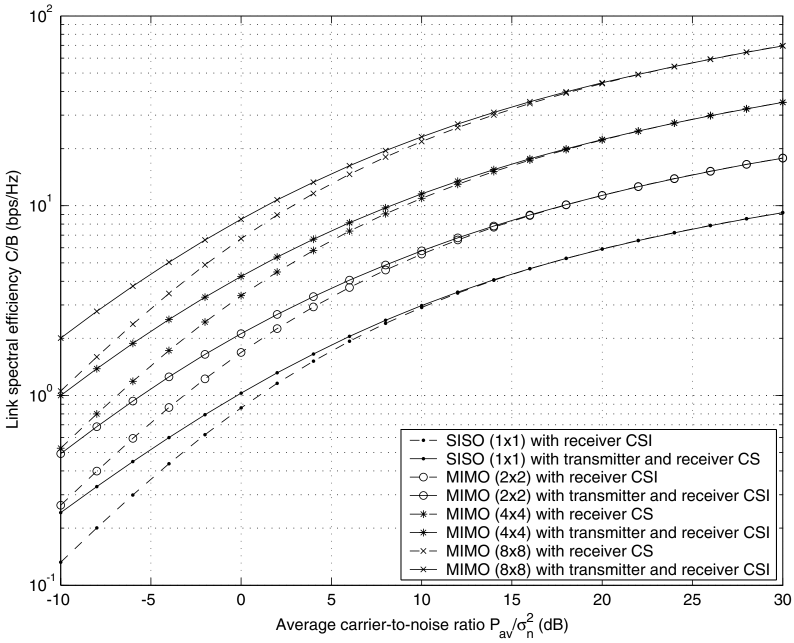

Figure 2.2. Average link spectral efficiency in a multi-antenna channel with uncorrelated multipath Rayleigh fading.

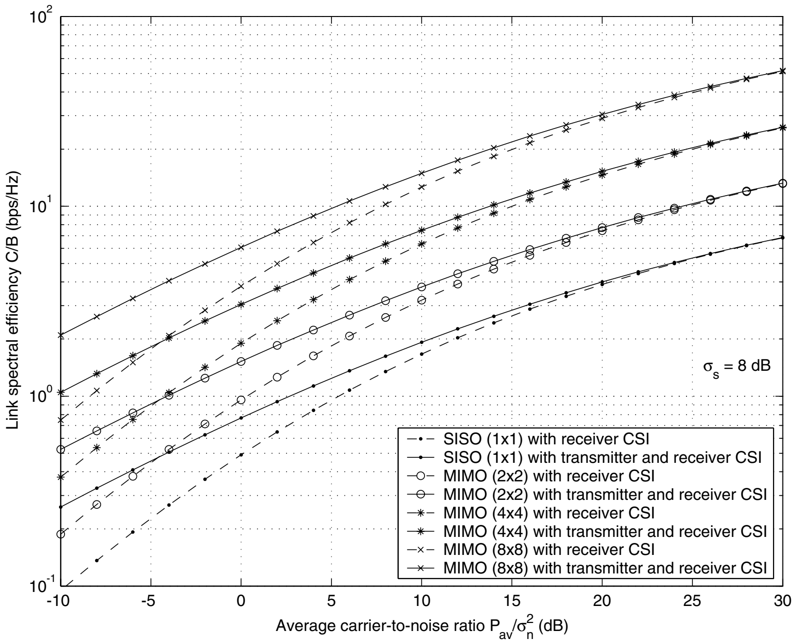

Figure 2.3. Average link spectral efficiency in a multi-antenna channel with uncorrelated Rayleigh fading and shadowing ($\sigma_s = 8\,\mathrm{dB}$).

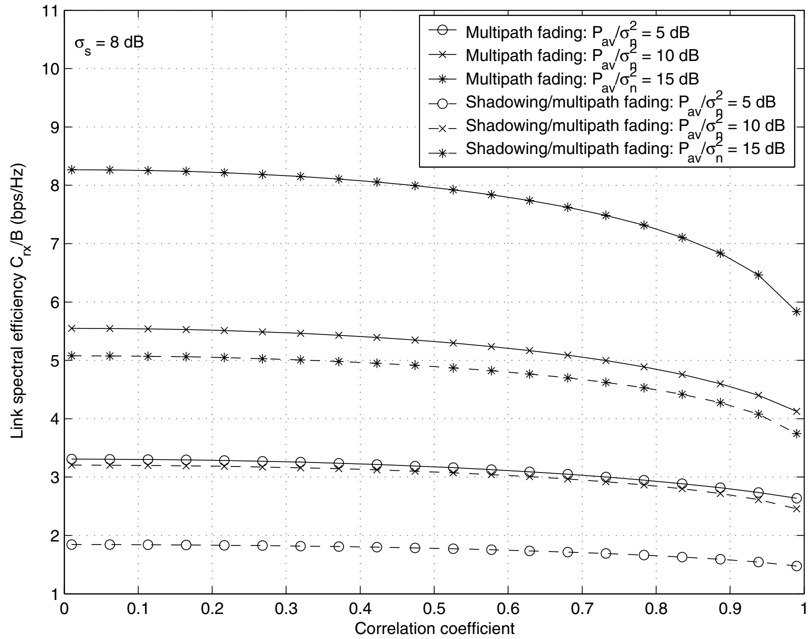

Figure 2.4. Impact of correlation between antennas in a multi-antenna channel with semi-correlated Rayleigh fading and shadowing ($\sigma_s = 8\,\mathrm{dB}$ *).*

2.5.1.3 Capacity versus outage performance

The ergodic assumption is not necessarily satisfied in practical communication systems operating on fading channels. In fact, if tight delay constraints are demanded, as is the case of speech transmission over wireless channels, the ergodicity requirement cannot be satisfied. In these circumstances, channel capacity is viewed as a random variable, as it depends on the instantaneous channel parameters. The suitable performance measure is the capacity versus outage, which is determined by the probability that the channel cannot support a given rate: i.e., an outage probability is assigned to any given rate (Biglieri et al. 1998):

$$F(R) = \Pr(C_{rx} < R). \tag{2.35}$$

Note that the capacity versus outage performance is described by the cumulative distribution function of the channel capacity distribution $F(R)$. In practice, the complementary cumulative distribution function, i.e., $\Pr(R < C_{rx})$, is often used.

Single-antenna systems

Calculation of capacity versus outage performance is relatively simple for single-antenna flat fading channels. This is because single-dimensional distribution of the random fading process is usually known. The probability density function of the link spectral efficiency is

$$p(C_{rx}) = 2^{C_{rx}/B} \ln 2\, p\left(2^{C_{rx}/B} - 1\right) \tag{2.36}$$

where $p(z)$ is the probability density function of signal-to-noise ratio distribution at z. The cumulative distribution function of the link spectral efficiency R/B is

$$F(R/B) = \Pr\left(\gamma < 2^{R/B} - 1\right) = \int_0^{2^{R/B}-1} p(\gamma)\, d\gamma\ . \tag{2.37}$$

Multi-antenna systems

Calculation of capacity versus outage performance for multi-antenna channels is more difficult. The probability density function is found by calculating the inverse Laplace transform of moment generating function $M(z)$ of the link spectral efficiency distribution, i.e.,

$$p(C_{rx}/B) = \frac{1}{2\pi i} \int_{\sigma - i\omega}^{\sigma + i\omega} M(-z) e^{zC_{rx}/B} dz \tag{2.38}$$

where σ is a vertical contour in the complex plane chosen so that all singularities of $M(-z)$ are to the left of it. Similarly, by the properties of Laplace transform, the cumulative distribution function of the link spectral efficiency, or equivalently the outage probability, is

$$\Pr\left(\frac{C_{rx}}{B} < \frac{R}{B}\right) = \frac{1}{2\pi i} \int_{\sigma - i\omega}^{\sigma + i\omega} z^{-1} M(-z) e^{zR/B} dz \tag{2.39}$$

The inverse Laplace transforms of moment generating functions are calculated numerically using a modified version of de Hoog algorithm (Hollenbeck 1998, Hoog et al. 1982).

Let us consider a multi-antenna channel with uncorrelated multipath fading where the channel state information is available only on the receiver side. The moment generating function of the link spectral efficiency is

$$M(z) = E\left[e^{zC_{rx}/B}\right] = E\left[\prod_{k=1}^{m}\left(1+\frac{sP_{\mathrm{av}}}{t\sigma_n^2}\lambda_k\right)^{z/\ln 2}\right]$$
$$= \int_0^\infty \cdots \int_0^\infty \prod_{k=1}^{m}\left(1+\frac{sP_{\mathrm{av}}}{t\sigma_n^2}\lambda_k\right)^{z/\ln 2} p(\lambda_1,\ldots,\lambda_m)p(s)ds \tag{2.40}$$

where the joint probability density function of eigenvalues is given by (2.10). By using Chiani et al. (2003), Theorem 2 we obtain

$$M(z) = c\int_0^\infty \det \Psi(s,z)p(s)ds \tag{2.41}$$

where

$$\Psi(s,z) = \left[\int_0^\infty \left(1+\frac{sP_{\mathrm{av}}}{t\sigma_n^2}\lambda\right)^{z/\ln 2} \lambda^{\alpha+i+j-2}e^{-\lambda}d\lambda\right]_{i,j=1}^{m}. \tag{2.42}$$

In a multipath environment, the moment generating function is (cf. Wang & Giannakis 2002)

$$M(z) = c\det\Psi(1,z) \tag{2.43}$$

The integrals in (2.42), that is, the entries of matrix $\Psi(s,z)$, can be expressed in terms of confluent hypergeometric functions of the second kind $U(a,b,z)$ (Gradshteyn & Ryzhik 2000, Eq. 9.211.4, p.1013) or evaluated numerically using Laguerre-Gauss quadratures.

Let us consider a multi-antenna channel with semi-correlated fading where the covariance matrix of the correlated side is m x m matrix. Furthermore, let channel state information be available to the receiver only. The moment generating function of channel capacity is given by (2.40) with the joint probability density function of eigenvalues defined in (2.14). By using Chiani et al. (2003) Theorem 2 we obtain

$$M(z) = \frac{(-1)^{m(m-1)/2}}{\det\left[\phi_j^{\alpha+i}\right]_{i,j=1}^{m}\prod_{k=1}^{m}\Gamma(n-k+1)}\int_0^\infty \det\Psi(s,z)p(s)ds \tag{2.44}$$

where

$$\Psi(s,z) = \left[\int_0^\infty \left(1+\frac{sP_{\mathrm{av}}}{t\sigma_n^2}\lambda\right)^{z/\ln 2} \lambda^{n-j}e^{-\lambda/\phi_i}d\lambda\right]_{i,j=1}^{m}. \tag{2.45}$$

In a multipath environment, the moment generating function is (cf. Kang & Alouini 2003)

$$M(z) = \frac{(-1)^{m(m-1)/2}}{\det\left[\phi_j^{\alpha+i}\right]_{i,j=1}^{m} \prod_{k=1}^{m} \Gamma(n-k+1)} \det \Psi(1, z). \tag{2.46}$$

We show the capacity versus outage performance of different multi-antenna systems operating in various propagation environments in Figs. 2.5, 2.6, and 2.7. The results show that adding more transmit and more receive antennas generally improves the capacity versus outage performance. In other words, the multi-antenna system can support a given transmission rate with lower outage probability or, equivalently, support higher transmission rate for the same outage probability. In general, the shadowing worsens the capacity versus outage performance for low required data rates and improves it slightly for high data rates (cf. Fig. 2.7). The results in Fig. 2.7 show that the capacity versus outage performance deteriorates when antennas are correlated. In particular, the outage probability for a given transmission rate increases when the correlation coefficient gets larger.

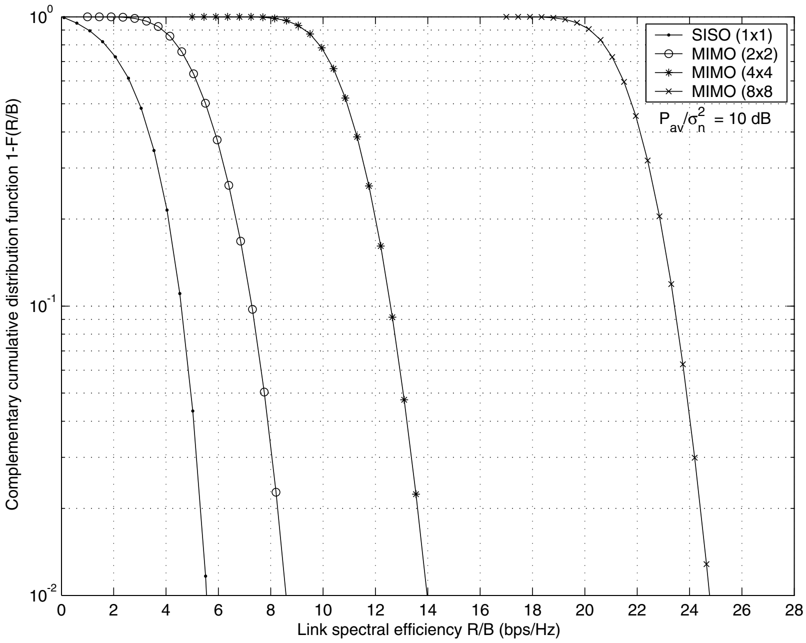

Figure 2.5. Complementary cumulative distribution function of the link spectral efficiency in a multi-antenna channel with uncorrelated multipath Rayleigh fading ($P_{av} / \sigma_n^2 = 10\,\mathrm{dB}$).

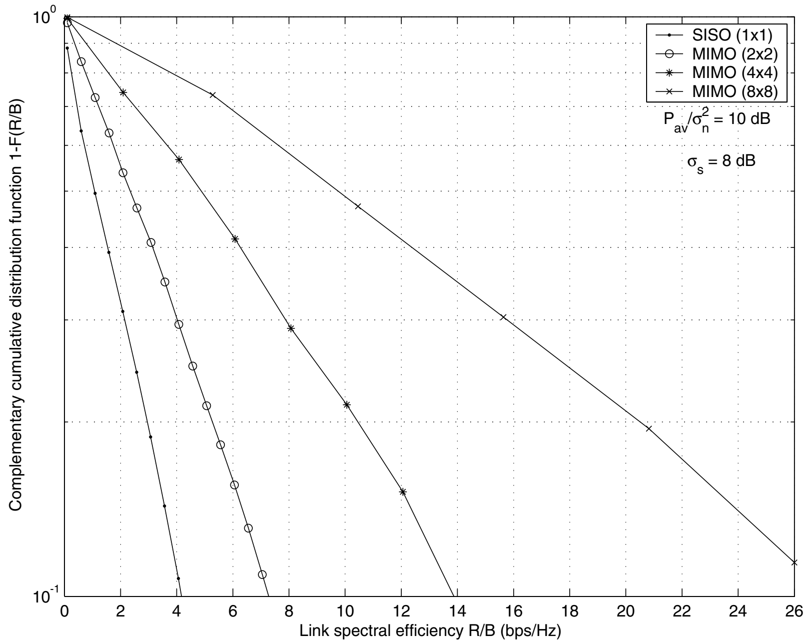

Figure 2.6. Complementary cumulative distribution function of the link spectral efficiency in a multi-antenna channel with uncorrelated multipath Rayleigh fading and shadowing ($P_{av}/\sigma_n^2 = 10\,\mathrm{dB}, \sigma_s = 8\,\mathrm{dB}$).

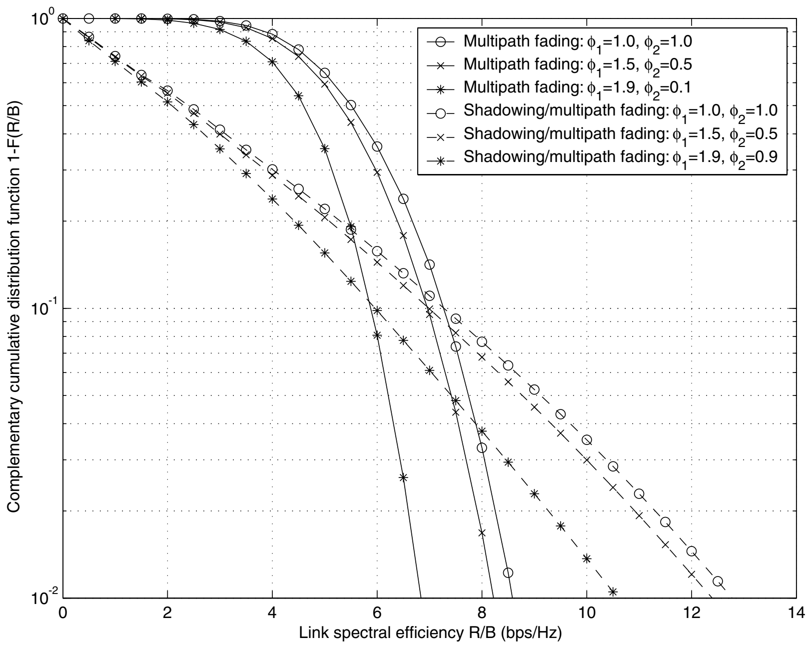

Figure 2.7. Impact of correlation between antennas on the capacity versus outage performance in a multi-antenna channel with semi-correlated multipath Rayleigh fading and shadowing ($P_{av}/\sigma_n^2 = 10\,\mathrm{dB}, \sigma_s = 8\,\mathrm{dB}$).

2.5.2 Modulation and coding

An important parameter in any signal constellation is the Euclidean distance between two signal points d. With gaussian distributed disturbance that has zero mean and variance $N_0/2$, the probability of an M-ary PSK symbol of being in error is a monotonically decreasing function on d^2E/N_0. Additionally we have $E = \log_2 M \cdot E_b$. This means that the PSK scheme with the highest $d^2 \cdot \log_2 M$ gives the lowest symbol error probability provided E_b is kept constant. In this terms, M values 2 and 4 are known to be good, as well it known that $M < 4$ leads to weaker power-efficiency.

The ternary phase shift keying (3-PSK) is an interesting modulation scheme, which, at least in theory, can beat the power efficiency of BPSK and QPSK. It is not at all any new contrivance; such authors as Marvin Simon (Simon et al. 1995, p. 130) and John Proakis (Proakis 1995, p. 185) have given a brief example on it. In fact, the ternary-PSK signals with 120° distinction from each other are simplex signals.

The concept of simplex signal or simplex code says that taking any of the signal points or code words, the Euclidean distance to any other point or code word is always the same. Hence, all other signals of code words lie always at the minimum distance d_{min} from the transmitted one. There are not any other distances in the constellation or in the set of the code words. Theoretically, the simplex signals or simplex codes are worth of pursuing, since they maximise the minimum distance. Unfortunately, the number of signal points is limited so that simplex signals are not normally used in applications. The same is even more true with simplex codes. One just cannot find any large set of code words having just one distance between them.

To transmit binary data with a ternary modulation, a binary-ternary transform is required, which inevitably brings some redundancy to the system. In general with short words, this redundancy deteriorates the performance. By using 11 or 19 bit words, this redundancy remains so small its impact remains negligible. In the case of ternary symbol error, the ternary-binary transform will generate several bit errors. This means that although the word error probability is lower with 3-PSK than with QPSK, the bit error probability tends to be higher, especially at a low E_b/N_0.

The power spectral density of 3-PSK has somewhat wider main lobes that QPSK. Hence, the ternary phase modulation may not be the most efficient if the narrow spectrum is the main issue to be considered.

Concatenation of 3-PSK with a hard decision decoded Reed-Solomon code is feasible. By applying 11 bit words in binary-ternary transform and assigning those words to RS code symbols, one obtains the coding scheme that works well with medium block sizes.

It outperforms QPSK and RS code with symbol length 8 bits when the block size exceeds 2000 bits. It attains the best performance with block sizes near to 20000 bits, where a coding with 8 symbols must resort to several code words.

The perfect ternary Golay codes can be used to error correction of the ternary words. They match well to 19 bit words that are mapped to 12 ternary symbols. Thus, there are only few short block lengths available. The performance with hard decision decoding is not the most impressive, but there is much more potential if soft decision decoding is applied. It is not, however, studied in this document, for which reason the uttermost performance of the Golay codes remain unsettled.

2.5.3 Network level performance

The network level performance of cellular systems was evaluated in the AuRa project with Matlab simulations using the ASE concept from Alouini & Goldsmith (1999a). We consider the uplink of fully loaded non-sectorised cellular TDMA/FDMA systems. The system model is shown in Section 2.3.2. The simulation method and the results are summarized in the report (Alasalmi 2002). The report also contains analysis and simulations of BER and theoretical link spectral efficiencies from the Shannon capacity equation for AWGN, log-normal shadowing, Rayleigh fading, Ricean fading, and Nakagami-*m* fading channels with diversity.

The simulation of network performance in terms of ASE goes in short as following. First, we generate the random positions of the desired user and interfering co-channel users in their respective cells. Then, we calculate the distance from the desired and interfering users to the base station of the desired user. From the selected channel model, we compute the received signal powers at the base station and obtain SIR for the desired user by dividing the desired user power by the sum of interfering signal powers. Based on SIR we compute ASE from

$$A = \frac{4}{\pi R^2 R_u^2} \log_2 (1 + \gamma) \qquad (2.47)$$

where R is the cell radius, R_u is the normalised reuse distance, i.e., the ratio of the reuse distance to the cell radius, and γ is SIR. This procedure is repeated a number of times to calculate the average ASE. The average ASE is computed for different values of the normalised reuse distance.

Theoretical upper and lower bounds for ASE are obtained in best case and worst case interference conditions. This implies that the interfering mobiles are located in the far or near boundary of their cells, respectively. We have evaluated the average ASE as a function of the normalised reuse distance in differenet channel conditions. Upper and lower bounds for ASE are given for the two-slope path loss model with both path loss exponents equal to 2. The average ASE for random user positions is simulated in two-slope path loss, log-normal shadowing, and Nakagami-*m* multipath fading channels. The results are shown in Figure 2.8 from which the reduction in ASE due to fading is clear. We used the standard deviation of 4 dB for shadowing and *m* parameter equal to 1 for Nakagami-fading. For more information on the parameters and models are given in Alasalmi (2002).

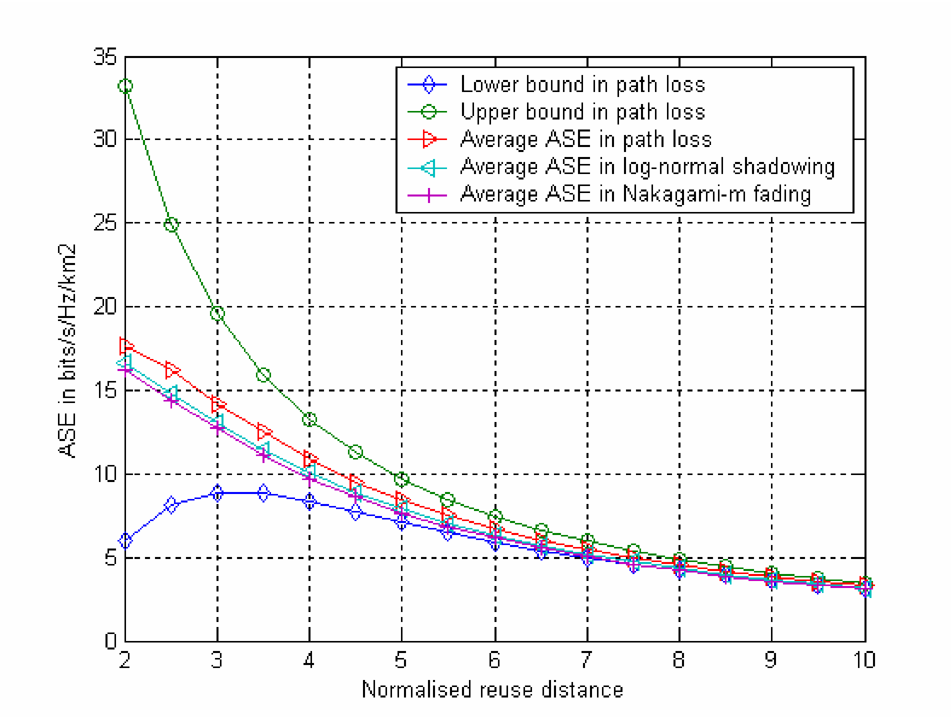

Figure 2.8. ASE as a function of the normalised reuse distance.

3. Quality of Service in wireless networks

Pekka Koskela

3.1 QoS overview

3.1.1 Quality of Service on different OSI -layers

In principal Quality of Service (QoS) can be enhanced at all Open System Integration (OSI) layers. New QoS solutions at the link layer (L2) are difficult in the current network because these kinds of solutions need new hardware standards. In respect, solutions at the network and transport layer (L3 and L4) will need development of standards and the added capability of a routing device like computing, buffer and memory capacity. The improvement of QoS at layers above the transport layer also needs the support of a network device, which maintains QoS in the network. The advantage of upper layer solutions is that the solutions can be made with software modules, which can be installed together with a software update of the operating system. The disadvantage of upper layer solutions is that they operate slower than the solutions of lower layers.

The medium access control (MAC) protocol of the link layer can restrict and control the amount of users as well as the amount of data traffic on the radio interface, which creates a framework for available customer QoS. If the MAC protocol does not restrict network users, but a large number of users are competing for the same resource, it may not be able to provide any QoS, which is then best effort. To restrict users or to reserve a part of the bandwidth from competition will enable QoS in network. When using a competition free resource reservation system, the utilisation of resource can be improved with adaptive reservation such as in HomeRF and 802.11e (Grilo et al. 2003, www.palowireless.com).

Control can be established with a distributed or with a centralized system, where the desired QoS is achieved by controlling the operation of border and intermediate routers. An example of a centralized system is resource reservation protocol (RSVP) and that of a distributed system is differentiated services (Diffserv) (Blake et al. 1998, Braden et al. 1997). The operation of RSVP is based on a pre-reservation of resource before a data transfer, which assures resources during a connection. The disadvantage of this kind system is that it needs an algorithm and routers, which can make and keep reservations. When the amount of traffic flows increases, at some stage the signalling mechanism of RSVP overflows and so the system does not scale. A possible method that overcomes the scaling problem is to aggregate flows, but then also the flow service will be aggregated. Diffserv does not need any pre-reservation and is scalable. The price of scalability is that it can not offer any hard QoS guarantees, but the QoS is proportional depending on the amount of users and traffic, i.e., soft QoS. Furthermore, QoS can be improved by tunnelling packets trough a network. The tunnelling can be established by labelling and encapsulating packets and by using routers, which can control and route packets according to marked labels, as for example in Multi-Protocol Label Switching (MPLS) (Rosen et al. 2001).

The advantage of solutions at the application layer is that they can be adopted with software and operating system (OS) updating. A restriction for this kind of implementations is that the created QoS depends fully on the support of network devices. In principal the solutions of the application layer resemble solutions of L3 and L4, where traffic through the network is controlled by the operation of routers, packet classification and the QoS control information of network. The main difference compared to solutions of L3-L4 is that the information is gathered and QoS control is made at the application layer, which makes it possible to use more versatile QoS mechanisms.

3.1.2 The requirements for QoS formation

Before a connection the user and a network provider have to agree on services and prices, called on service level agreement (SLA) (Blake et al. 1998). The service, which the user needs, depends on applications, which the user is running. In order to create the needed service at first we have to classify data. The classification can be made based on the application or in a more accurate manner, based on the importance of data packets. When "the importance of packets" classification is used it can allocate the resource more specifically, but then applications must have the capability to classify packets. When the data is classified and sent out to the network it can be treated in routers with a suitable mechanism and that way create the desired QoS. The mechanisms which are used in routers are packet separation according to a packet class and a policy of class treatment, which can be made with buffering, access control and different queuing algorithms. In order to maintain QoS it should monitor traffic, which makes it possible to detect congestion beforehand. The monitoring could be performed by intermediate routers, which would forward information to border routers. When the border routers receive the information about congestion they can make suitable acts like restrict traffic to the network. When the user changes network operator, the user has to make a new SLA. In addition, border routers and other QoS equipment have to be able to interconnect the QoS of different kinds of networks. The needed functionality is depicted generally for end-to-end QoS in Figure 3.1.

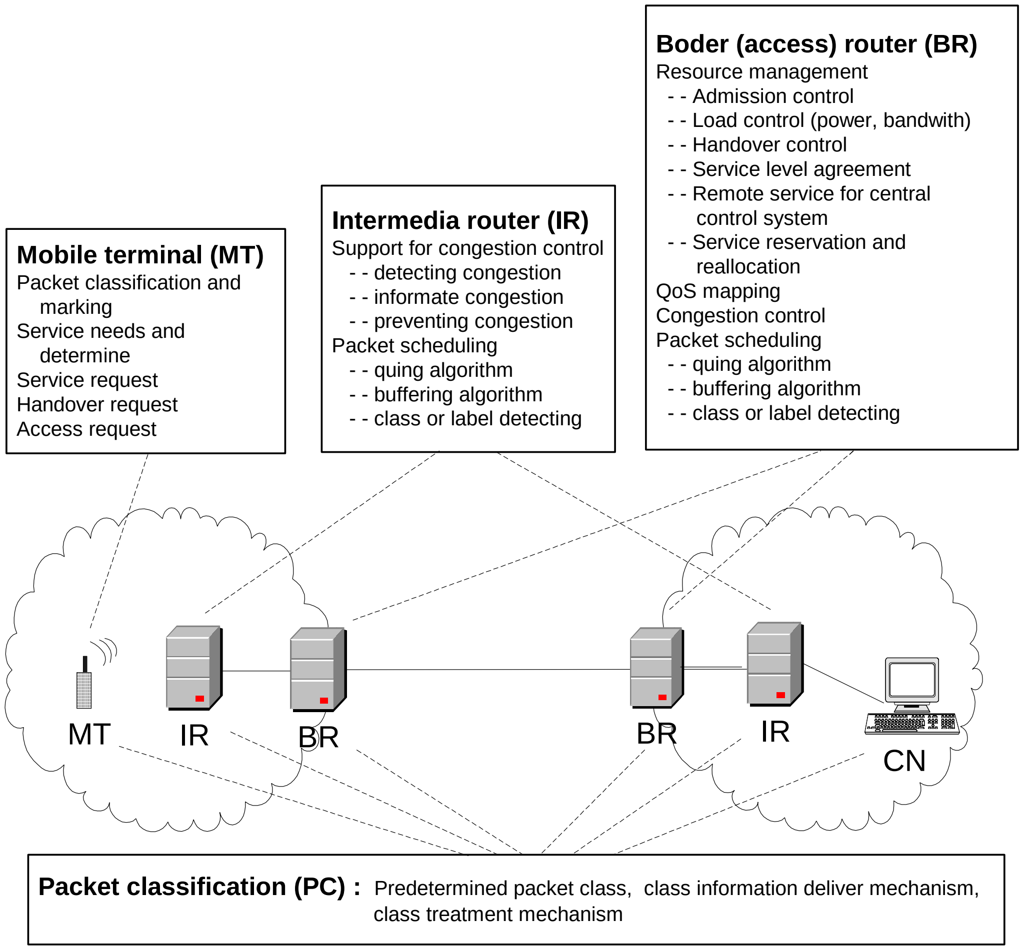

Figure 3.1. The needed functionality of end-to-end QoS between a mobile and fixed network.

3.2 QoS parameters and QoS classification

QoS can be roughly divided into two classes depending on which a point of view is taken. From the customer's point of view they are only interested in and ready to pay for the performance of their applications, i.e., quality of experience (QoE). On the other hand a provider of network service is interested to know how the network and QoS should be managed, so that user's desired service experience can be created in the most economical way and the services will be available for as many users as possible.

3.2.1 Classification of data

The data traffic, which different applications generate and receive, can be divided according to bursty and continuous traffic. A characteristic of bursty traffic is unpredicted temporary transmission peaks, whereas the amount of traffic is predictable for continuous data streams. Reserved data channels are preferred for continuous traffic, whereas using reserved channels with bursty data results in a waste of channel resource.

The data traffic can also be divided according to time sensitivity: real-time and non real-time traffic, where real-time traffic can include interaction. When the data traffic is real-time it has to be guaranteed a specific data rate with restricted delays. Indeed, when it is needed for safety, more requirements will be needed like the faultlessness of traffic and a stability of connections. The strictest demands for QoS bring up the need for real-time and safety, which depend on the running application. Typically interactive applications like audio and video can tolerate some packet lost, but are very sensitive to delays between packets, see Table 1.

Table 1. Video and audio application demand of delay, data loss and bit rate.

Interactive real-time	**Delay**	**Frame erasure rate**	**Bit rate**
Audio (conversational)	< 150 ms (good) < 400 ms (tolerable)	<3 %	4–13 kbit/s
Videophone	same as above audio	<1 %	32–384 kbit/s

3.2.2 Applications specific QoS parameters

The requirements of different kinds of applications create the basic criteria of QoS parameters. When building up a data channel for applications, among others the following QoS parameters are used for channel determination.

- Throughput
- Delay
- Jitter
- Packet fault between sending and receiving
- Stability of connection
- Packet loss
- Packet order

The throughput of the network can be depicted with by maximum bit rate and guaranteed bit rate, where the maximum bit rate is a physical upper limit that the network can not exceed. However, from the application point of view it is more important to establish a connection, where a specific, minimum data rate is fulfilled and so the bit rate is guaranteed. The most guaranteed data connection will be achieved by a channel reservation, but the risk is that the whole reservation is not used in an efficient way but a part of the capacity is under utilised and so the network capacity will be wasted. The shortage of the reservation can be minimised by using reserved channels, which are dedicated for many users, e.g., shared channels or those whose reservation changes dynamically according to traffic. The cost of this adaptivity is that it is more complicated and difficult to implement.

The guaranteed bit rate can be achieved by prioritising pre-determined data classes and by handling traffic according to priority. The prioritisation is a straightforward method and is easy to implement, but the shortage is that when the traffic increases at some stage all the traffic becomes best effort. Another simply way to give a service, which the customer requires is to use a large enough, oversized bandwidth. Although this kind of arrangement will operate well in the user's point of view it is not economical from a network provider's point of view. When the data traffic is bursty the capacity must be reserved according to the maximum data peak and between peaks resources will stay underutilised. Furthermore, when oversized bandwidth used the network provider has difficulty to create guaranteed QoS, hard QoS, which they would be able to charge to users.

Faulty and lost packets diminish the performance of applications. How faulty and lost packets affect performance, depends on the application and included fault correction mechanisms in the system. In wired networks almost all packet loss is caused by congestion, which can be solved by proper traffic management and packet priorization. In wireless networks the packet loss is caused by corruption of wireless channel, which may result from increased interference. The interference level can be controlled with power and transmission management. If packet loss has already occurred, one correction mechanism is to send the lost packets again. This however causes delay. The caused delay can be reduced by classifying packet/bit importance and sending the important packet/bit as double. Because the character of packet loss is temporary and occurs often randomly, the regular double sending wastes resource and reduces available bandwidth.

The jitter causes choppiness of real-time application like audio and video. Jitter can be corrected by buffering, which anyway causes delays and which may cause new problems.

The weak stability of connections causes packets loss and increases delays between packets. When a connection goes down it is essential how fast the connection can be re-established. If it is due to a weak signal a correction is an increasing in transmission power. This is not always possible because an increasing in transmission power increases the interference on the channel.

The packet loss with retransmission and other delays may change the received order of packets, which changes reserved information. In order to preserve information the packet order should be the same both at the sending end and at the receiving end. The possible ways to correct the received packet order is buffering and/or resending of packets.

In the wired traffic the capacity of the network can be increased by increased the size of packets, where the relative part of payload data increases and that way improves QoS. In the wireless network this is not true. An increase in packet size after a threshold packet size increases the probability of packet loss and corruption, therefore reducing the capacity of the network and QoS.

3.2.3 Operator specific QoS parameters

In addition to the QoS requirements of applications, we can determine parameters that relate purely to the maintenance and operation of networks, like handovers, regardless of the application. These kinds of parameters include: the signal interference ration (SIR), bit error ration (BER), transmission strength, connection dropping rate (CDR) and connection blocking rate (CBR). If a handover access is prioritised highest then the blocking rate of new connections will increase as the amount of handovers increases. So the blocking rate can be affected by reducing the amount of handovers and the handover time. The amount of handovers can be restricted to using a clever handover decision algorithm, to avoid useless handovers and to increase/preserve the size of network cells. The handover time can be reduced by decreasing control traffic and speeding up the handover process.

When the amount of users and the transmission strength increase, this increases the interference of the network, which decreases the quality of the signal with an increasing bit error rate. By decreasing the cell size users can be divided into several cells and the transmission strength can be decreased. That way the interference of the network can be reduced. On the other hand the decreasing of cell size increases the need for handover, which increases control traffic and blocks new users.

3.3 QoS architecture of some existing systems

3.3.1 UMTS and UTRAN

In a UMTS network data traffic is separated, mainly according to delay sensitivity, into different classes (see table 2), where the Conversational class is very sensitive to delays and Background traffic tolerates delays the best.

Table 2. QoS class of UMTS network (3GPP TS 23.107).

Traffic class	**Conversational class** **Real Time**	**Streaming class** **Real Time**	**Interactive class** **Best Effort**	**Background class** **Best Effort**
Fundamental characteristics	Preserve time relation (variation) between information entities of the stream Conversational pattern (stringent and low delay)	Preserve time relation (variation) between information entities of the stream	Request response pattern Preserve payload content	Destination is not expecting the data within a certain time Preserve payload content
Example of the application	Voice	Streaming video	Web browsing	Telemetry, emails

A network provider can offer specific traffic class connections, specific dominating QoS parameters and limit values in each class have to be determined (3GPP TS 23.107).

- Traffic class ('conversational', 'streaming', 'interactive', 'background')
- Maximum bit rate (kbps)
- Guaranteed bit rate (kbps)
- Delivery order (y/n)
- Maximum SDU size (octets)
- SDU format information (bits)
- SDU error ratio
- Residual bit error ratio
- Delivery of erroneous SDUs (y/n/-)
- Transfer delay (ms)
- Traffic handling priority
- Allocation/Retention Priority
- Source statistics descriptor ('speech'/'unknown')

The QoS classes of UMTS were determined by ETSI standards. The standards do not tell how classes will be implemented in different parts of the network, but work is left for network operators. In a wireless radio network the QoS can be created by channel sharing of UTRAN, which has been supported with MAC level prioritisation and the RRC function.

Figure 3.2 depicts the UMTS user plane protocol stack (3GPP TS 23.060) and two possible ways to build up end-to-end QoS in an UMTS network. In the UTRAN and UMTS core network (Node B-RNC-SGSN-GGSN) the traffic is ATM based, with related QoS features. When ATM based QoS is used before routing data to UMTS, the data is mapped in the GGSN with an external QoS architecture. The mapping should be able to match, in a logical way, different kinds of QoS classes to each other and transfer QoS management information, like congestion control information, between different QoS architectures. Another possibility to create end-to-end QoS is to use MPLS, which is an IP based protocol. The advantage of using MPLS is that the mapping with the external network QoS will be easier than in a purely ATM solution and if the external QoS protocol is also MPLS no mapping is needed. Only an ATM/IP address resolution protocol will be needed. The other advantage of using MPLS is that the formation of QoS is more flexible. The disadvantage of using an "all IP" solution compared to ATM QoS is that it will need an MPLS implementation in the UTRAN and UMTS core whereas ATM QoS exists already. A totally different way to build end-to-end QoS is to use IP routers with no ATM base, but from an economical point of view it will not appear as a potential alternative. IP routers with no ATM will be more expensive and slower than current ATM routers.

The UMTS core network uses a tunnelling protocol called GTP. MPLS also tunnels packets trough a network so it might be unnecessary to use two tunnelling protocol. If GTP is used together with MPLS then GTP should develop so that it does not affect the performance of MPLS.

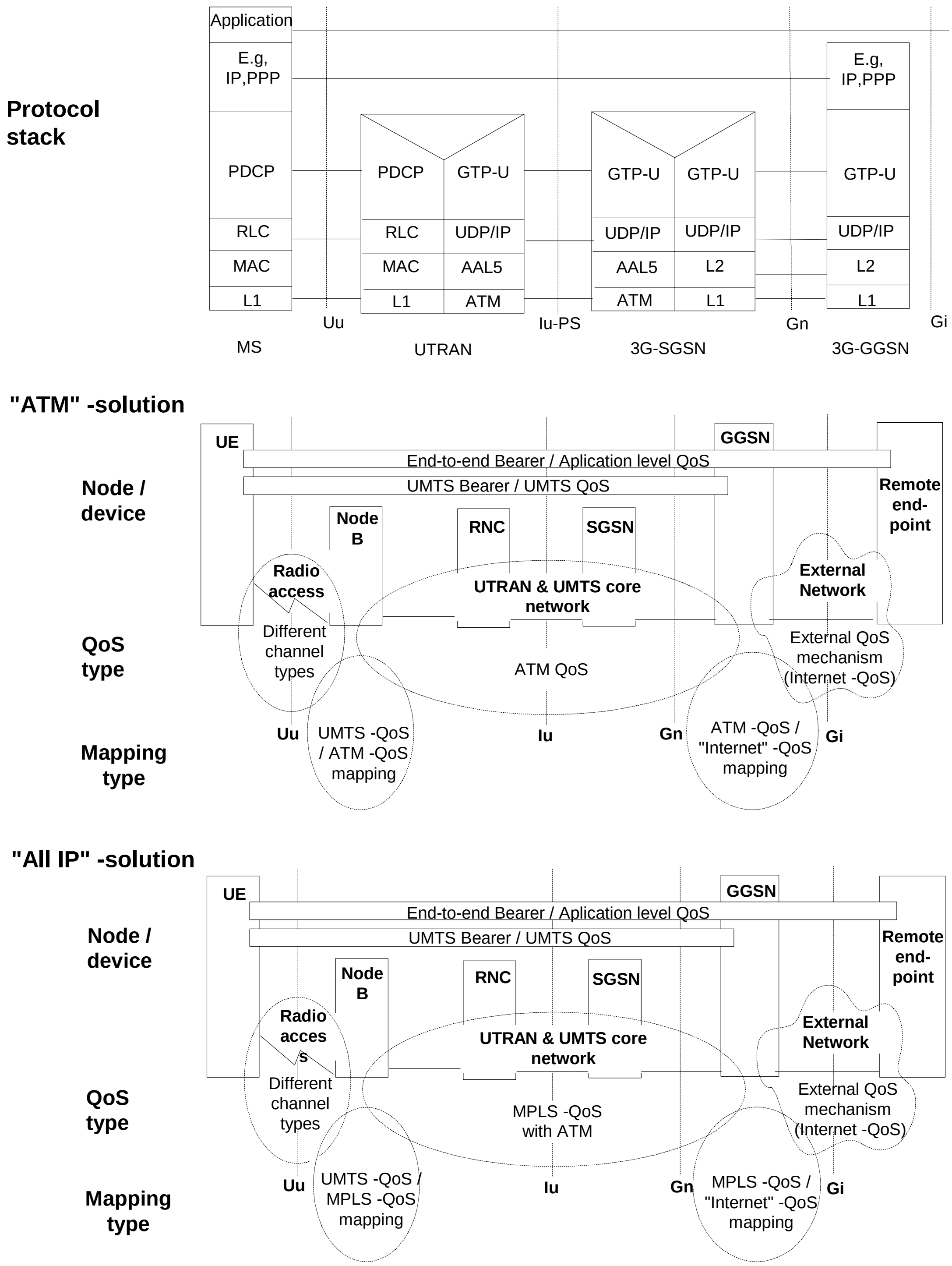

Figure 3.2. UMTS user plane protocol stack and possible end-to-end QoS solutions in UMTS.

In the ETSI standard it is not defined how the QoS classes, determined by the UMTS standard, will be implemented, but QoS creation is left to operators. In generally the operators can use QoS 1) to restrict access of users in a cell and 2) to direct traffic both to the dedicated and undedicated channel so that the QoS that users request can be created.

In order for operators to create QoS in efficient they should determine an algorithm, which classifies packets in such a way, where only packets that need specify QoS will be directed to the dedicated channel. Furthermore, the dedicate channel should be able to use dynamic reservation in order to avoid wasting resources in consequence of bursty traffic. One possible way to create a dynamic reservation system is to develop and standardised a common dedicated shared channel for several users, whose size would be control feedback information from the application.

3.3.2 ATM -network

Table 3. QoS service class of ATM network.

Attribute	ATM Layer Service Category				
	CBR	rt-VBR	nrt-VBR	UBR	ABR
Traffic Parameters					
PCR and CDVT(4.5)	specified			specified[2]	specified[3]
SCR, MBS, CDVT(4.5)	n/a	specified		n/a	
MCR[4]	n/a				specified
QoS Parameters					
peak-to-peak CDV	specified		unspecified		
maxCTD	specified		unspecified		
CLR[4]	specified			unspecified	See Note 1
Other Attributes					
Feedback	unspecified				specified[6]
n/a = not applicable					

In ATM networks there are five different QoS service class, shown in Table 3. The constant bit rate (CBR) service is used by traffic, which demands a constant bit rate during a connection. This kind of traffic includes audio and video. The rt-VBR (real-time variable bit rate) class is also used for real-time traffic but the bit rate is not assumed constant, and the traffic can be more or less bursty. The non-real-time variable bit rate (nrt-VBR), unspecified bit rate (UBR) and available bit rate (ABR) classes are for non real-time traffic, where UBR is the lowest service class. In addition to the above classes the forthcoming standard adds one more class, guaranteed frame rate (GFR). The new class will be used to get more visibility at the frame level. The QoS of ATM network based resource reservation done with connection admission control (CAC). Because an ATM network is a circuit-switching network, desired QoS can be achieved with a reservation based system.

3.3.3 WLAN (802.11xx)

At the present time basic WLAN 802.11a and 802.11b do not include QoS; all traffic is best effort traffic. The used MAC is contention-based, which tries to give similar throughput to all users. The MAC has no mechanism, which can favour some user at another's expense. One possibility to create QoS in 802.11a and b is use internet QoS protocols like DiffServ, if at the same time we can restrict the amount of users (using access control) so that the access point does not saturate with traffic, i.e., buffers do not overflow. The other way is using the optional polling MAC scheme, which can promote some users by giving more transmission time. The standardising IEEE work group has tried to develop a suitable QoS mechanism for WLAN, but there is not yet a published ratified standard. The proposed QoS standard draft is based on the prioritisation of traffic and divides the MAC into two parts, where in one part traffic can be sent without contention "contention free MAC" and one part is under contention "contention based MAC".

3.3.4 Internet

In the Internet QoS can be executed in several ways: Diffserv, RSVP, MPLS and IntServ, where each method has advantages and disadvantages. RSVP, which also IntServ uses (Wroclawski J. 1997), is a quite heavy protocol, which can implement desired guaranteed QoS. The disadvantage of RSPV is that it is not scalable and the network has to have routers, which can make QoS reservations. Meanwhile, although Diffserv is scalable, its problem is that it can not alone guarantee QoS, but QoS is related to the data traffic and throughput of the network. With MPLS we can enhance QoS by using labels to guide routing. When data packets are encapsulated and labelled the data packets can be tunnelled through routers, which can handle labels. MPLS speeds up routing decisions both at individual routers and at the network level by setting up a complete route across the network (the labelled switched path, LSP). An LSP decision can be influenced by RSVP and so the two technologies can work together. When both technologies are used routers have to be upgraded so that they are MPLS enabled with RSVP capability. Diffserv classes are reserved six bits in IP header field and theoretically it can use 64 different classes. At the moment the following classification has been proposed, see Table 4.

Table 4. DiffServ classes.

Class type	Objective	Example	Delay	Jitter	Packet Los Ratio	Bandwidth definition	DSCP
NCT1/NCT0	Minimised error, high priority	RIP, OSPF, BGP-4	100 ms	U	10^{-3}	Committed rate	111 000/ 110 000
EF	Jitter sensitive, real-time high interaction	VoIP	100 ms	50 ms	10^{-3}	Committed rate	101 110
AF4	Jitter sensitive, real-time high interaction	Video conference	400 ms	50 ms	10^{-3}	Committed rate & peak rate	100 000
AF3	Transaction data, interactive	Terminal session, Custom app	400 ms	U	10^{-3}	Committed rate & peak rate	011 000
AF2	Transaction data	Data base Web	400 ms	U	10^{-3}	Committed rate & peak rate	010 000
AF1	Low loss bulk data	FTP, E-mail	1 s	U	10^{-3}	Committed rate & peak rate	001 000
BE	Best effort	Best effort service		U	10^{-3}	U	000 000
U = undefined							

3.4 Frames for the QoS formation

In order for a terminal to be able to agree on QoS with the network, the terminal device has to be capable of access and to connect to the network confidentially. For this it will need authentication, authorisation and accounting (AAA) support of the network. After a successful connection the device needs a procedure, service level agreement (SLA), to negotiate its desired service (QoS). When both the customer and the network have accepted the service and its price, data traffic can be routed to a new network. Before the traffic is transferred to the new network, the services will be mapped together according to the agreed upon SLA. During the service mapping different service class have to be mapped in a logical way to each other and the information, which is needed in QoS management, has to be transferred to the new network. The QoS performance will be monitoring different programs and methods.

3.4.1 Access

When a device accesses a new network, the network and the device have to identify each other, in order for example to allocate a service and a corresponding charge. Because during identification private information is exchanged, a secure connection is needed. The secure connection is also needed when a user negotiates their service expectation. If it uses procedures, where user-specific information is transferred between the new and the old network, the networks should be able to authenticate each other and when it is needed to use a secure connection. One possible case, where this transfer would be take place is a vertical handover, when the handover process is sped up by enhanced QoS transferring.

3.4.2 Negotiate of QoS

When an AAA supported connection is created, it can be started on an agreed service, which customer needs. The service class and its definition and corresponding parameters must be agreed upon. Furthermore, there must be an agreement procedure, on which the negotiation can be made. The QoS parameters should be unambiguously defined, easy to understand, independent of technology and accurately measure and depict the user's performance of service. These kinds of parameters include: the establishment time of a new connection, packet loss, packet delay, build up time of a collapsed connection and collapsing frequency.

3.4.3 Mapping and flow control

Data can be either classified or unclassified. If data is already classified it can be transferred to a new class by mapping a current class to the new class. If the mapping does not succeed or the data is unclassified, data must be classified at first before transforming. When suitable classes exist, the transform will be done case by case, which depends on the architecture of the service and the classes. The other possibility would be to agree on common QoS classes, which all QoS methods use or transform their class to. This kind system requires a commonly accepted QoS class standard, where also QoS parameters would have common acceptance.

If QoS classification is missing in the current network, the classification would be made by the means that an application marks its QoS needs in a separate user profile. These kinds of files would be located in the user's home network, their terminal or the Internet. In the home network the creation of a user's QoS profile file is quite simple and it could be a part of the home network service. Problems might come when the user is travelling outside their home network. How would the profile be updated, and how would the information be retrieved there? In order that different networks can exchange information between each other, the networks have to create an AAA supported connection between each other. All this needs contracts between operators, which complicates information exchange. The advantage of using an Internet QoS information service would be the ubiquitous nature of the service, but problems might come from how the service would be arranged on the Internet. The advantage of saving the profile in the user's terminal is that the QoS profile does not need to be saved in any separate server, but the user itself takes care collecting, saving and transmitting QoS information. The disadvantage of this arrangement is that it consumes scarce resources of the user's device and demands specific properties of the device and user applications such as the mechanism for the application to identify its needs and to create, save and update the user's QoS profile.

If the application is not capable of announcing its QoS needs, the needs might be analysed from connection information by using a firewall type program. However, the simplest is to use the user's QoS profile, which is based on the announcement of the application. The information, which the application announces would include the following:

- Bandwidth need
- Maximum allowed delay
- Maximum allowed jitter
- Maximum allowed bit error rate
- Maximum allowed a connection of missing time
- Identification information of the application, in order to allocate service to it.

Because the data transferring capability of different networks can diverge remarkably from each other, it has to be able to follow and control a data flow so that a new network does not get congested as a consequence of transferring. To prevent congestion the traffic can be re-routing from overflowing routers in the new network or the traffic can be preventing from accessing the new network. In order that networks are able to react to and predict congestion TCP is not enough in a multi network system, but it has to use protocols, where intermediate routers can inform possible congestion to the border router. The control data transmission can use for instance IMPC messages. When the congestion information has arrived to border routers, the data traffic can be restricted, slowed down or re-routed according to the QoS class by using algorithms such as different version of random early detection queue management (RED) or BLUE. The basic idea of RED is to predict congestion and remove packets before congestion. The prediction is based on an average queue length dropping possibility.

The processing of routers can be affected by controlling when and what packets are forwarded, when and what packets are dropped from a buffer and which way buffer are filling, e.g., FIFO, Tail-drop etc. In solutions, which do not take care of the permission of a data flow, badly behaving data flows will fill buffers and block out flows which operate correctly. One way to solve that problem is to weigh the data flow before buffering, for instance, by using round-robin and so controlling access to the buffers.

3.4.4 Connecting IP base QoS to ATM QoS

Figure 3.3 generally depicts the QoS mapping between ATM and IP networks. The simplest QoS mapping between an ATM and IP network takes place when IP QoS is used with MPLS. The reason is that both ATM and MPLS use the same kind of mechanism to create QoS, called "labelled tunnels", and this way enhances the mapping and exchanging of QoS control information. When Diffserv is used, the DS class and the ATM virtual channel have to be mapped to each other and at the same time it should take care of resource management and control, which is more complicate than in the MPLS case.

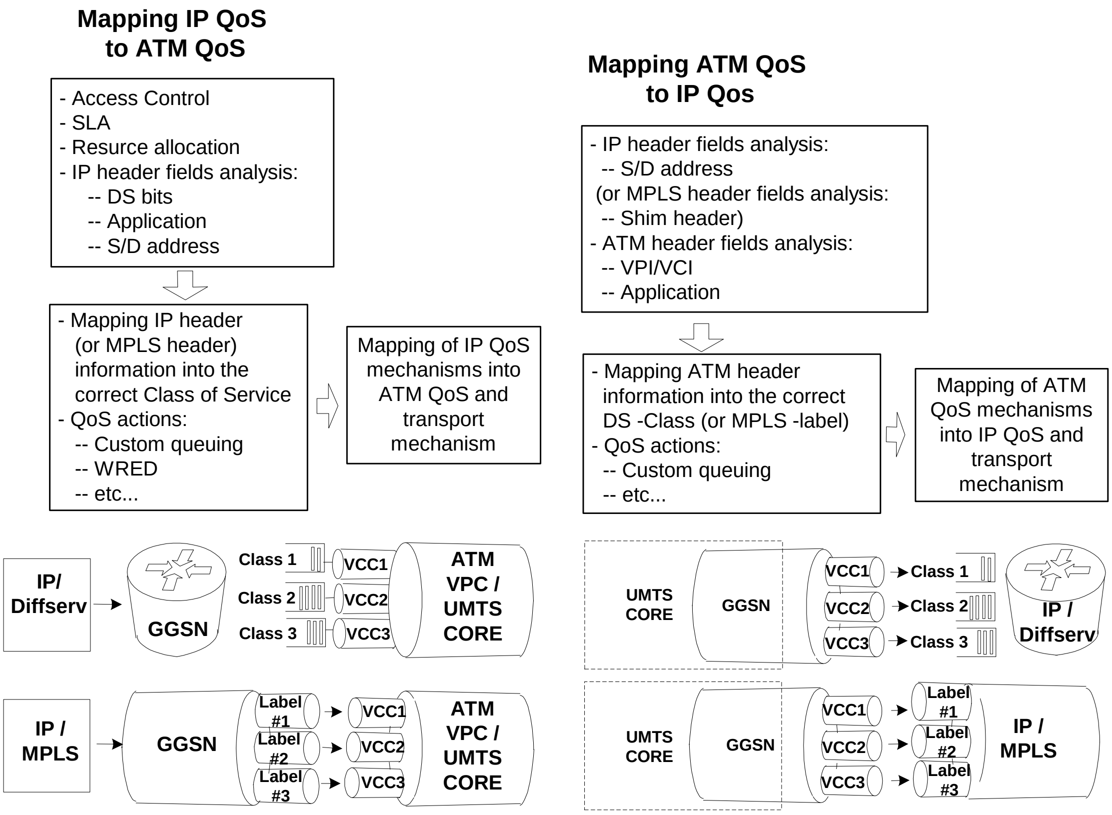

Figure 3.3. The QoS mapping between Diffserv and ATM and between MPLS and ATM.

3.4.5 Connecting WLAN -QoS to UMTS -QoS

Because current WLANs (802.11a and b) do not support QoS, data traffic has to be classified in the WLAN before accessing the UMTS border router (GGSN) in order to agree on the needed QoS. The QoS classes can use UMTS QoS classes, which enhance the mapping process. The classifying can be done either in WLAN routers or in the user terminal. WLAN routers do not currently have that kind of feature and a classifying function is probably easiest to add in the terminal. In the terminal the classifying can be done either in a separate application or in the application, which use QoS. When the classifying has been done, the terminal sends the SLA (data rate, delays, error tolerating etc.) request to the UMTS GGSN. The UMTS SGSN answers the request and negotiates possible changes and after that the UMTS SGSN and the terminal agree about the SLA. The UTRAN SGSN has a procedure which reserves needed channels, directs, slips and classifies packet traffic to suitable channels according to the SLA. The terminal cuts off an old connection or the connection is cut off automatically after a specified time.

Because current WLANs do not have QoS, UMTS traffic is changed to best effort traffic when it is transferred from UMTS to WLAN. So there is no QoS parameter mapping. If a WLAN uses some Internet QoS protocol then the UMTS QoS class can be mapped to an Internet QoS class. In that case it can not guarantee any UMTS like QoS, but instead QoS depend on the amount of traffic. The more traffic the poorer a WLAN can maintain QoS.

The interesting question when transferring data from a WLAN to UMTS is how to map the different throughput of the networks to each other. One solution is to limit the WLAN data traffic so that it can be transferred through current UMTS, in particularly through the UTRAN radio. This kind of arrangement means that the user allows some decrease in QoS. The amount of data can be reduced by using different codec techniques and by removing less important packets. The other possibility is to add a new access method in UTRAN, which allows throughputs of WLANs. This kind access could be for instance a UMTS supported WLAN access point, which is connected directly to the UMTS core network.

3.4.6 Monitoring

Monitoring information is used for:

- The tracking of network flow and network topology to find bottlenecks in a network.
- The shaping of network traffic in order to control and manage traffic reasonably.
- The composition of network traffic as a function of time in order to find out its affect on networking.

When the user and network provider have made an SLA, the agreement should be monitored some how. In order that the monitoring would be possible, the parameters to measure and their limit values in different service levels have to be agreed upon first. To enhance monitoring and controlling, the parameters should be unambiguously defined, independent of technology and easy to measure. These kinds of parameters include: the connection time of a new connection, reconnection time after breakdown, packet delay, jitter, bit error rate and packet loss.

The chosen QoS parameters can be monitored actively or passively. Active monitoring is based on the generating and monitoring of test traffic, which is sent over the network. This gives information about the performance of the network such as availability of the application, delays and packet losses. One widespread method is to use the Ping program, which uses Internet control message protocol (ICMP) echo request and reply packets as test traffic. With the Ping program the connectivity of a host, round-trip time and packet loss can be monitored. One disadvantage of the Ping program is that it uses ICMP packets as test traffic, which the network handles in a different way than data traffic.

Other active monitoring methods include the Traceroute program, end-to-end FTP transmission and the capability of the WEP service. The Traceroute program gathers a path list, which includes routers and transfer times. The disadvantage of the program is collected only in one direction, although in the opposite direction the network may operate in a different way. By sending FTP traffic across the network, the information collected will be either the packet throughput of network or the receiving capacity of the receiver, depending on which one is more restricted. The capacity of a WEP server can be measured by generating WEP service request and by measuring response times, thus finding out how fast the network operates. The down side of these methods is that response time is a summary of several delays such as DNS, connection, server and network delay and we do not know the actual components of delay, which would be more interesting when traffic is managed.

The other network monitoring method is to do it passively, where no test traffic is generated, but the monitoring information is gathered directly from original network traffic. One problem in passive monitoring is that the monitoring devices should disturb the operation of the network as little as possible.

Table 5. Corresponding processing time when a bit rate of port increases.

Throughput of port	100 Mbps	1 Gbps	1 Tbps
Processing time per packet	*) 0.12 ms	0.012 ms	0.012 μs
*) Total throughput / average packet size = port speed * in&out * utilisation rate of link / average packet size = 100 Mbps * 2 * 0.5 / 1500 * 8 = 8333 packet/s = 0.12 ms / packet			

In Table 5 we see that the faster the network is the less time there is left to process packets. So when the data rate of the network increases, we can not handle all packets, but we have to use sampling, where only a part of passing packets are studied. This sampling can be done either at a specific packet frequency or at a specific time frequency.

When passive and active methods are compared to each other, the advantage of the passive method is that it does not need to generate extra test traffic. The disadvantage of passive methods is that extra devices, which do monitoring, have to be implemented in the network, which includes additional cost. Furthermore, if with the passive method delays are measured, the measuring devices should be able to synchronise with each other. One possibility to synchronise is using GPS. In active methods the synchronisation is not needed, but delay times can be calculated from round-trip times.

3.5 Vertical handover between WLAN and UMTS network

Planning vertical handover does not only involve the handover algorithm, but is a much wider study area, where the following issues should also be taken into account:

- Seeking available service from surrounding networks
    - -- Mechanism to find out service
    - -- Preserving of service information
    - -- Candidate router list
- Establish new connection
    - -- Authentication
    - -- Access method
    - -- Possibly authorisation for information exchange between networks during vertical handover
- Routing to new network
    - -- Mechanism to get new IP address
    - -- Mechanism to re-route traffic from the old network to the new network
    - -- Timing of removing of old connection and establish a new connection

- Service transferring to other network
    - -- Service and price agreement
    - -- Mapping different QoS
        - -- Parameter mapping
        - -- Possibly re-mapping
        - -- Access and flow control
        - -- Management information exchange between different QoS
- Handover triggering
    - -- Trigger parameters and algorithms
    - -- Handover mechanism & decision process
    - -- Signalling between networks.

3.5.1 Seeking of available service from surrounding networks

In principal there are two possible methods for finding out the available service of surrounding networks. The network can send an advertisement broadcast (destination address 255.255.255.255) of offered services to potential customers' terminals, which decide if they need the available services. The other possibility is that the terminal requests from the surrounding network if it can offer any services. Based on gathered service information the user can make a candidate network list, which can contain the information of services, prices and access methods to the candidate networks.

If the broadcast method is used the method disturb the radio media as little as possible so that it does not consume the transmission capacity of media uselessly. It is important determine the tolerable frequency and power of broadcasting. The broadcast procedure should also be dynamic in respect to available service. When the network capability of offering services reduces, broadcasting should also be reduced. This can be done by decreasing the exponential broadcast frequency, when the offered service decreases.

Power issues are critical to all mobile devices. To save power the service broadcast message should be such that it can also reach devices that are in a sleeping mode. If a network has broadcast messages, like a paging message, which can reach a sleeping mode device, the service information could be piggybacked in the broadcast message.

In the request method a device listens to the surrounding network traffic and concludes, for instance based on the prefix of IP address, the existence of new networks. When a new candidate network is found the device tries to connect to it and if the connection succeeds the device start the request procedure. The request procedure can be similar to SLA negotiation and can be part of the SLA procedure.

If potential network exchangers are few, the procedure based on service requesting, loads the network less than the service broadcasting method. When the amount of potential exchangers increases broadcasting becomes the least loading alternative at some stage. The stage when broadcasting is the least loading is affected by the existence of overlapping networks. If there are several overlapping networks broadcasting will stay more loading than the request method.

When observing of new network is based on broadcasting, the procedure is not as heavy as the request procedure. In the broadcasting method a client already knows what services are available and can make a service request. Based on the service request the new network makes needed QoS reservations.

Broadcasting QoS information would include the following items:

1) Right of access (for instance: no, private, public, security code, occupied)
2) Offered service class (for instance DS -classes). Chosen class system should support development of service and respect needs of applications.
3) Stability of connection (95 %, 99 %, 99.98 %)
4) Recovery time of connection (<0.4 s, 0.4–2 s, >2 s)
5) Offered security class (for instance: no, private, office, bank/military).

The broadcast packet format could be as in Table 6, where M = Mandatory, O = Option and EOF = End Of File. The first bit length can be used if it is pre-agreed, what attributes different bit sequence respect.

Table 6. Packet format of QoS broadcast.

(M) AP / BS identification 32 bit	(M) Access condition 3bit/(16bit)	(M) Length 1 (data class) 8bit	(M) Length 2 (Add. service) 8bit	(M) Stability of connection 3bit/(8bit)
(M) Recovery time of connection 3bit/(8bit)	(M) Security class 3bit/(8bit)	(M) Data class 3bit/(8 bit)	(M) Price 16bit	(O) Data class 3bit/(8 bit) etc…
(O) Price 16bit	(O) Data class 3bit/(8 bit) etc…	(O) Addition service (16bit)	(O) Price 16bit etc…	(O) EOF (8bit)

When a mobile user's roaming coverage area includes different networks, where service information updates are performed by either broadcast or request it is necessary to list the offered service. The user can choose according to the active application, the most suitable network connections at the moment. The format and contents of the offered service table could be following (see Table 7):

Table 7. Candidate access point table format.

Access condition	ID of AP/BS	Stability of connection	Recovery time of connection	Security class	Data class	Price	Additional service	Price
(16 bit)	ip-addr (32 bit)	int (8bit)	float (8bit)	char (8bit)	int (8bit)	double (16bit)	char+int (8+8bit)	double (16bit)
(16 bit)	ip-addr (32 bit)	int (8bit)	float (8bit)	char (8bit)	int (8bit)	double (16bit)	char+int (8+8bit)	double (16bit)
etc...								

3.5.2 Connection establishment

How to establish a connection depends on the network. In WLANs a connection can be opened through an access point by using the MAC protocol. In order to have security the accessing device and the network should be able to authenticate each other. In WLAN the old authentication method is WEP, which has several security problems. As a substitute for WEP the WPA protocol has be developed. It corrects the known problems of WEP and forthcoming Wi-Fi devices will use it. In WPA new clients are authenticated by the extensible authentication protocol (EAP), where authentication can be done by the network interface card (NIC).

User can access UTRAN from the radio interface through a base station or from a fixed network trough the GGNS's interface. When accessing from the radio interface the user needs a SIM card, where authenticate information is saved. When access takes place the GGSN interface will use the Packet Data Protocol (PDP), which exchanges authentication information and service information.

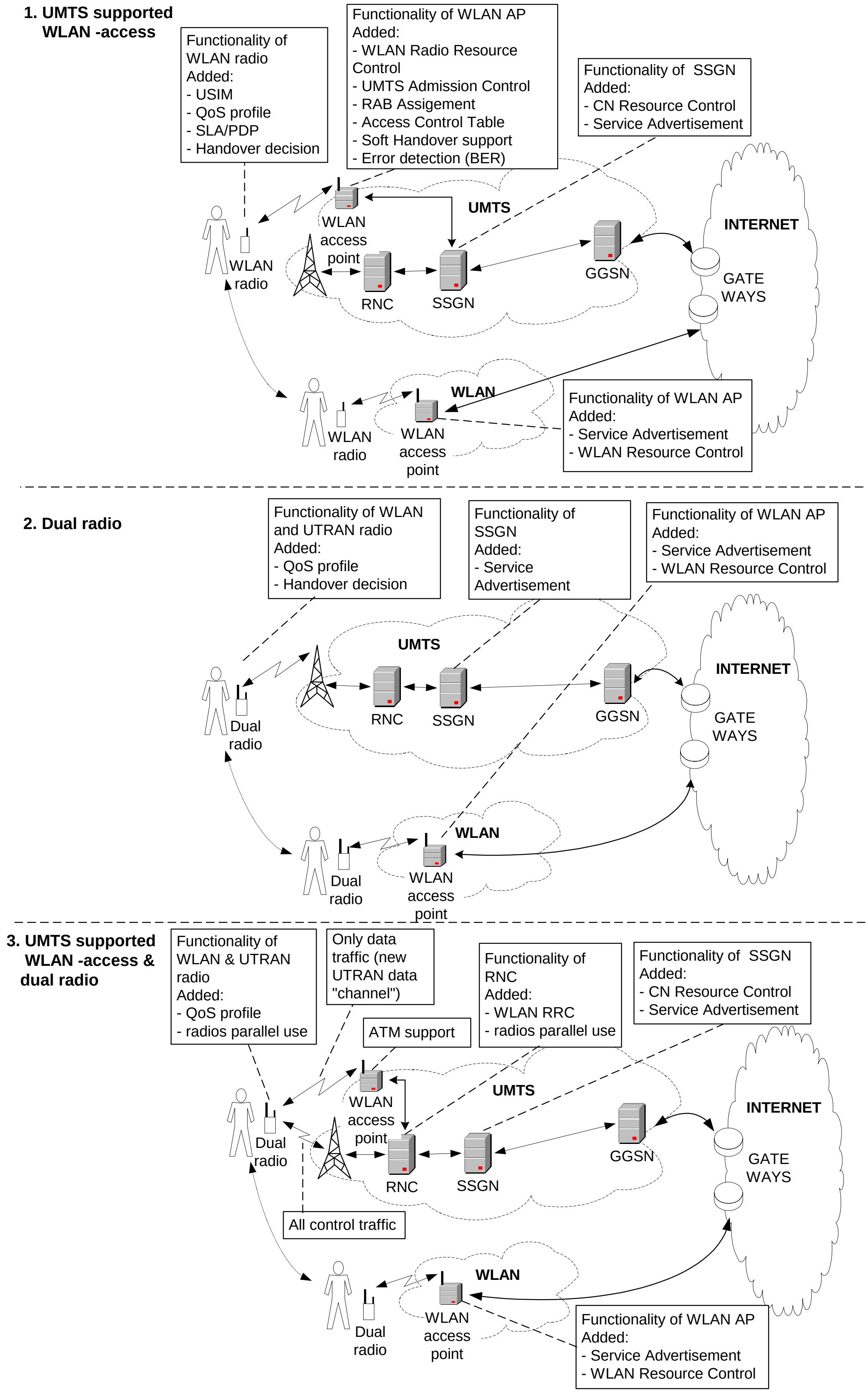

Figure 3.4. WLAN access to UMTS.

In principal a WLAN device can access UMTS if it can connect physically to UMTS and perform the needed access and authentication protocols. Because WLAN and UTRAN physical layers are totally different, the WLAN device should be a dual radio device, which supports both radio techniques, or the UTRAN interface should have a WLAN supported access point, see Figure 3.4. The advantages to using the dual radio technique are that a vertical soft handover implementation is possible, structures of networks do not need any change, like smaller cells. The disadvantage is that the dual radio consumes more power even though another radio will be mostly in sleep mode and the mobile terminal is more expensive because of the two radios. When the UMTS supported WLAN access point will be used it needs standardisation and implementation of new UTRAN access and devices. The advantage of this solution is that operator can use the WLAN access point for directing faster data streams, the operator has more radio capacity to delivery and new access point simplifies the mapping between WLAN and UMTS. The last alternative is to use both the WLAN UMTS access point and dual radio so the system has the advantages and disadvantages of both methods. One more advantage of using both methods is that it is possible to make a more sophisticated arrangement for data transfer, like using WLAN only for data where UMTS channels give control traffic support.

All the presented access methods need some mechanism, which can transfer a user's service and identification information between WLAN and UTRAN. The simplest case is when the WLAN and UMTS operator is the same. Then the same NIC/SIM card can be used for both networks. This kind of solution is not ubiquitous, where the user can roam freely between different networks and choose the best service. Thus the solution where the same operator offers both WLAN and UTRAN network services is not necessarily to the benefit of the user. Another possibility is to make and use a multi-SIM/NIC card, which contains the access information of different networks or to determine a common SIM card, which different kinds of networks can accept.

3.5.3 Routing and transferring services

When the network is changed some mechanism to re-route packets and to transfer QoS is needed. One possibility is to use Mobile IP (Perkins 2002) for re-routing. The IP address of the new network, which Mobile IP needs, can be detected with the address auto-configure algorithm (Thomson S. and Narten T. 1998, 2001). A problem when using Mobile IP is that re-routing takes considerable time and if the connection is cut several packets will be lost. Thus, to preserve QoS a soft handover mechanism should be used instead of hard handover, where the mobile device can hear both networks simultaneously. A significant disadvantage of this arrangement might be increased power consumption in the case of a dual radio device, when both radios are active.

Before vertical handover the mobile user has agreed on services with the new network. Based on the agreement the network operator makes necessary arrangements like classification, mapping and reservations. How this will be implemented in practice depends on the QoS and network technologies that which are used.

3.5.4 Criterions and analysis of handover

The following handover criteria could be used for transferring from WLAN to UTRAN. The most important criteria for the user are quality of the radio link, stability of connection and price. Correspondingly for an operator the most important criteria are the possibility to produce, as economically as possible while saving resources, the services that users request. The quality of the connection could be estimated by the following parameters:

- Delay between packets
- Jitter
- Packet corruption and loss
- Bit error rate (BER).

The stability of a connection could be predicted according to the state and change of state at the moment:

- Strength of signal
- Gradient of packet loss and bit error rate
- Gradient of signal strength.

From an economic point of view of the operator, important criteria for handover are: To prevent useless repeating of handover and to reallocate quickly reserved, but unused resources and to minimise control and duplicated traffic. The following parameters could be used for the handover decision:

- Elapsed time of previous handover (prevent ping-pong phenomena)
- Reallocation time of unused resources & handover time (to anticipate release of resources).

There are many parameters to use in the handover process and the decision for handover is the combined effect of all parameters. That kind of multivariable process control is maybe the simplest to implement with fuzzy logic, although the testing would be more labour than with purely algorithm based control. Other advantages that might be achieved by using fuzzy logic are: reduction in power consumption and the enabling of a stepwise decision process, i.e., several limit values are possible. With the stepwise handover process it is possible to give notice and prepare a new network before the actual handover. This way, for instance a powered down connection can be activated in advance, speeding up the original handover process.

The basic principal for fuzzy logic control is that member functions and their interdependency can be formulated within limit of the human power of deduction, which restricts the amount of usable parameters in one fuzzy calculation. The other principle is that when making a control system it should be as simple as possible, in which case its implementation and testing will be easier and functionality and reliability become better. Based on the above a fuzzy solution system is proposed, where parameters are classified classes, where the first results of fuzzy calculation are collected for the next fuzzy calculation, from which we get the expectation value for handover. If the expectation value is bigger than a specified limit value then preparation for handover or a finally handover will occur. Table 8 presents one possible variable grouping for fuzzy analysis.

Table 8. A possible variable grouping for fuzzy -analysis.

1. Compound price (*1. fuzzy -analysis*)

Price	cheap (<0.7)	tolerably (0.7–1.2 %)	expensive (>1.2)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
price difference between networks	big (> 50 %)		small (< 20 %)
member function	*"down diagonal step"*		*"up diagonal step"*

2. Compound quality of connection (*1. fuzzy -analysis*)

corruption-loosing amount	small (<0.1 %)	tolerably (0.1–15 %)	big (>10 %)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
packet delays	small (<150 ms)	tolerably (100–400ms)	big (>200 ms)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
jitter	small (< 2 %)	tolerably (5–20 %)	big (>20 %)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*

3. Compound stability of connection (*1. fuzzy -analysis*)

gradient of bit error rate	slow (1 % / s)	tolerably (3–10 % / s)	fast (> 20 % / s)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
gradient of packet loss-corruption rate	slow (1 % / s)	tolerably (3–10 % / s)	fast (> 20 % / s)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*

Collection of compound variable (*2. fuzzy -analysis*)

Compound price	cheap (<0.4)	tolerably (0.3–0.7)	expensive (>0.6)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal steps"*
Compound quality of connection	good (<0.4)	tolerably (0.3–0.7)	bad (>0.6)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
Compound stability of connection	stable (<0.4)	tolerably stable (0.3–0.7)	unstable (>0.6)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
Elapsed time of previous hand over	short time (< 1 min)		long time (> 10 min)
member function	*"down diagonal step"*		*"up diagonal step"*

Handover result

expectation value	no hand over (≤0.6)	prepare hand over (0.6<ev<0.7)	do final hand over (>0.7)

Rules

If price not expensive and quality good and stability stable no hand over

If price not expensive and quality tolerably and stability stable and elapsed time of previous hand over is long prepare hand over

If price is tolerably and quality not bad and stability tolerably and elapsed time of previous hand over is long prepare hand over

If price is expensive and previous hand over is long do hand over

If quality is bad and previous hand over is long do hand over

If stability bad do hand over

Handover from UTRAN to WLAN would be take place when WLAN can offer a clearly cheaper connection for active applications. Because the original WLAN can not offer guaranteed QoS, the only thing that can guarantee some probability of QoS is a large enough data throughput. The handover decision could be based on a relative price (compared to UTRAN) and an available relative throughput (compared to UTRAN), see Table 9. The decision itself could be done with fuzzy analysis.

Table 9. Handover parameters for WLAN to UMTS handover.

Relative price	cheap (<50 %)	slightly cheaper (20–50 %)	equaly (>20 %)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
throughput	bigger (> 2-times)	slightly bigger (1.5–2-times)	equaly (<1.2-times)
member function	*"down diagonal step"*	*"triangle"*	*"up diagonal step"*
elapsed time from previous handover	short time (< 1 min)		long time (> 10 min)
member function	*"down diagonal step"*		*"up diagonal step"*

Handover result

expectation value (ev)	no hand over (≤0.6)	prepare hand over (0.6<ev<0.7)	do final hand over (>0.7)

Rules

If price is cheap and throughput is bigger and elapsed time is long from previous hand over do handover

If price is cheap and throughput is slightly bigger prepare handover

If price is slightly cheaper and throughput is bigger prepare handover

If price is equally no handover

3.5.5 A scenario of vertical handover mechanism between WLAN and UTRAN

When a terminal moves in UTRAN, it listens to candidate WLANs and offered services in surrounding networks, see Figure 3.5. The terminal lists all available services and creates a "power ready" connection to suitable WLANs. In this context a "power ready" connection means a silent connection, where the WLAN radio is powered off and services are not agreed upon, but the device has an IP address and permission to access the network. If some connection is available with a better price/service relation than the current network, the terminal changes connection to a new network and leaves the UTRAN connection in sleep mode. When the terminal goes on roaming the connection may deteriorate or some other connection will become preferable. In that case the terminal changes back to UTRAN or a new WLAN depending on which one is preferable. Before a new (WLAN) or updated (UTRAN) connection the terminal and network agree on services and after that data traffic can be directed to the new network. Depending on the QoS architecture used it will be done mapping between different QoS. During mapping the requested service level and the loading state of the new network will be taken into account.

1. QoS BROADCAST. Surrounding networks send broadcast advertisements of QoS, i.e., what they can offer. The advantage of broadcast for the user is that they can straight away by listening find out what services are available without separately requesting from each network. The user can instantly chose the most preferable network. When a device is in power off state (WLAN radio), it has to have a mechanism, where the device wakes up from time to time for a short broadcast listening period. If broadcasts are done at a defined frequency the listening should be done randomly in order to avoid that all broadcasts take place at power off time. When the device is in sleep mode (UTRAN) the broadcast message can piggyback a suitable control message, which the UTRAN device listens to in sleep mode.

In WLAN access points send broadcast messages. Before QoS and price information can be sent, the AP should have a program, which is able to create and serve a state and price information packet. Price information may be retrieved from a separate server and service state could be estimated from the loading state of the access point. Some estimation of the loading state could be gotten for instance with a firewall program. If the user is charged, a network has to authenticate the user under a secure connection. One possibility is to use the WPA protocol together with a SIM/NIC card, where security information can be checked from a separate server (RADIUS), which has been installed on the network.

In UTRAN base stations send broadcast messages, which come from the RNC. At the moment price and loading information can be found from UTRAN, but it will need a procedure that creates a broadcast packet and send it to the base station, which broadcasts it. One possible way to broadcast a QoS advertisement is to piggyback it to a suitable control message, which the terminal can listen to also in sleep mode. Another possibility is to make a custom advertisement information packet, which is sent through the broadcast channel.

2. Mobile user listens to the QoS broadcast of different networks and chooses the preferred network, where it establishes a data traffic connection. If the new connection is a WLAN connection, the UTRAN connection is set to sleep mode.

3. Mobile user starts data traffic to the chosen network and occasionally listens to service advertisements to find a preferred network

4. According to the handover analysis the mobile device triggers the need for a vertical handover. The handover process could be done in two stages, where the first stage prepares the new connection for the second stage, which finalises the handover and completes the connection.

5a. If the new network is UMTS, the mobile sends a request to the RNC to weak up the sleep mode connection with specific QoS requirements.

6a. The RNC gives an ACK about the connection establishment. A separate negotiation of service is not needed, because the client already knows based on the QoS broadcast, what kinds services are available.

7a. When a connection is ready, data is directed to "Node B", where its QoS maps to UTRAN's QoS.

8a. The WLAN connection is cut, when mobile user stars to receive packets from UTRAN.

3b. If the new network is WLAN, the mobile powers up the connection and requests services from an access point in the case where the WLAN offers services.

4b. The access point replies to the service request.

5b. The mobile informs the home network about the network change with Mobile IP.

6b. When the connection is ready, data is directed to the WLAN access point where UMTS QoS is mapped to WLAN QoS.

7. When the mobile user start to receive packets trough WLAN, the UTRAN connection is changed to sleep mode.

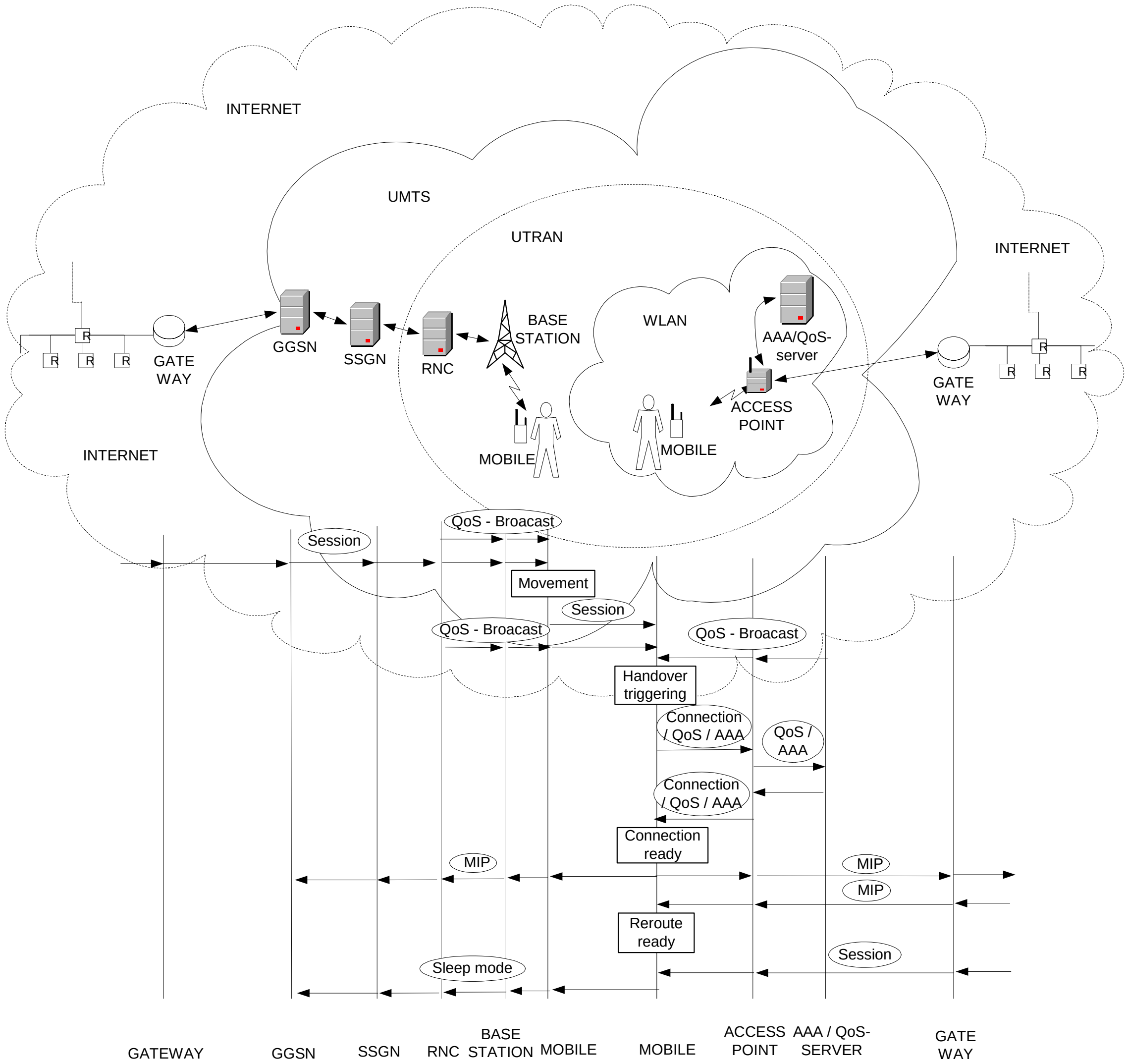

Figure 3.5. Vertical handover between WLAN and UMTS.

3.6 Mobile QoS measurement tool

Marko Palola

This chapter talks about the developed mobile service QoS measurement tool for measuring the performance of GPRS connection from a mobile phone to a WWW service. QoS measurement tool has been improved by adding new features and tested in Nokia 3650 and 7650 mobile phones during the Aura project.

The QoS measurements are focused on the network side, especially, in tele-communication networks QoS measurement hardware and software have been used several years. On the other hand, measuring the user experience in networks have gained popularity. The only way to measure the user experience is to use real terminals, which are connected to real networks and services.

The measurement tools with the end user viewpoint were not available. Our method was to develop a generic tool to mobile phones, which allows making QoS measurements over mobile network to services. Mobile phone as a tool is easy to carry along and measurement software can be downloaded to the terminal over the network at the measurement location.

The system architecture is shown in Figure 3.6. The current system consists of a data server for collecting benchmark results and a J2ME (Java 2 MicroEdition) application in a GPRS mobile phone for making the QoS measurements.

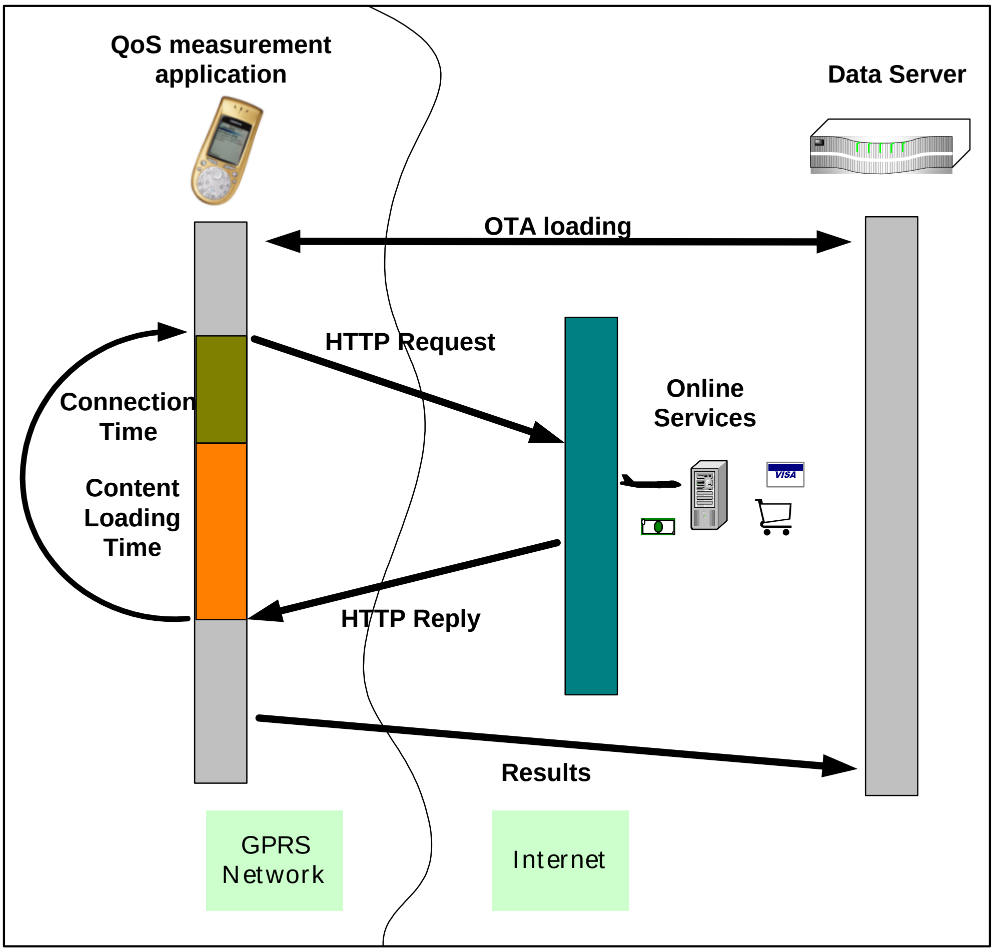

Figure 3.6. Mobile QoS measurement system.

The system measures performance information of HTTP (HyperText Transport Protocol) requests over GPRS (General Packet Radio Service) network as perceived by the user.

QoS measurement application in mobile phone can be used to measure different kind of services. In a typical measurement, the measurement application collects and calculates the measurement results for each individual HTTP request which includes the time needed to load the service content into the phone and the total time of the HTTP access. Mainly two QoS parameters are measured from the service access; throughput and delay as observed by the mobile phone user. The measured parameters include the properties of the user terminal, access network and service. The tests can be repeated, so the application is capable of calculating the total amount of data transfer in bytes, total count of succeeded requests, total amount of bytes transferred in each request and the average, maximum and minimum times.

Data server provides a few fixed tests for the measurement application such as variable length WWW pages and also more general WWW server capabilities including OTA (over-the-air provisioning) for loading measurement application into the phone. New benchmarks are can be developed into the mobile phone application without changing the data server.

The actual measurement steps involve the following:

1) Mobile phone user loads a measurement application from the data server into the phone using OTA over GPRS, or using infrared or cable.

2) After the application is installed, user can specify user's name, user's address, network operator name, and phone model into the application. User information can be sent to the data server over GPRS.

3) User can browse the available benchmarks on the phone and start execution of one of the available benchmarks.

4) Phone executes the benchmark, which typically sends HTTP requests over GPRS to the specified WWW address. Typically, the measurement is repeated several times over a long period of time. The benchmark results are stored into the phone.

5) The user can browse the results on the phone and submit selected results to the data server for further analysis. The data server receives test results from the phones. Test results can be viewed, saved into a file, removed, and transferred to other programs.

Figure 3.7 and Figure 3.8 shows measurement results taken by the QoS measurement tool. The graphs show measured delay of downloading content (HTTP GET) and uploading content (POST) into a HTTP server over GPRS. These results have been taken using different operators (Sonera and DNA), phone models (Nokia 3650 and Nokia 7650) using different access methods (uplink at right and downlink at left) data transfer, and different days, so the results are not directly comparable.

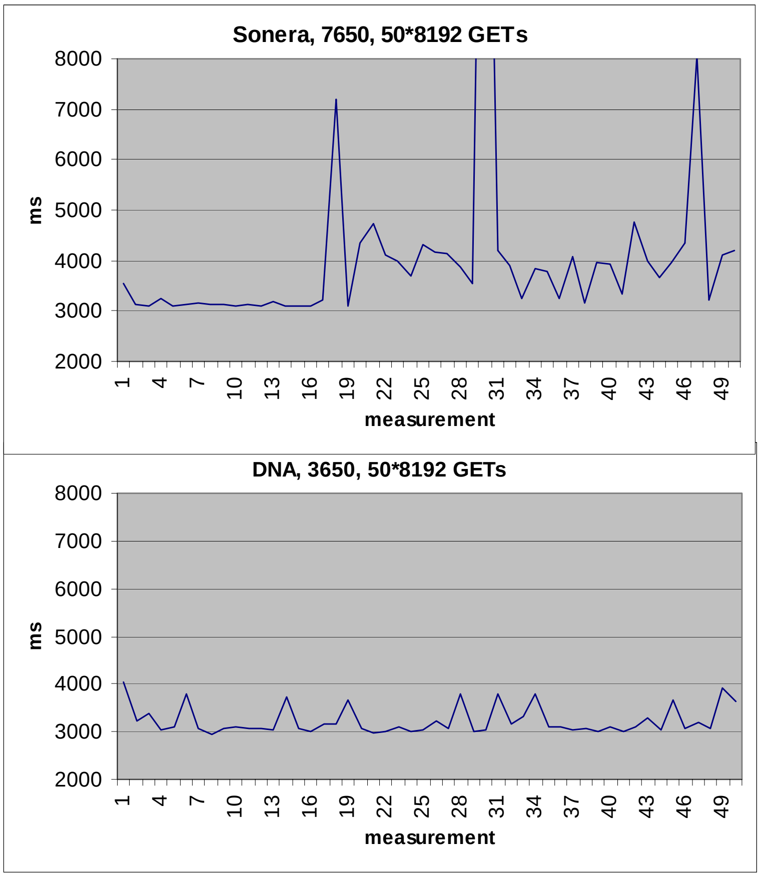

Figure 3.7. Mobile download measurements.

A large fluctuation in the access times is due to the GPRS network configuration and the amount of simultaneous users in the cell. Delay is also affected by lost packets, Packet Data Channel configuration, error coding scheme, and the number of transceivers in the mobile terminal (Porcarelli, et al. 2003, Andersson, et al. 2004).

The lost transmission in GPRS is detected by using TCP protocol and the timeout of the packet acknowledgement, which can be set by the GPRS operator. This timeout is probably the reason for the Sonera GET graph's 40 seconds request delay. Other possible reasons for delay can be too many simultaneous voice users in the cell.

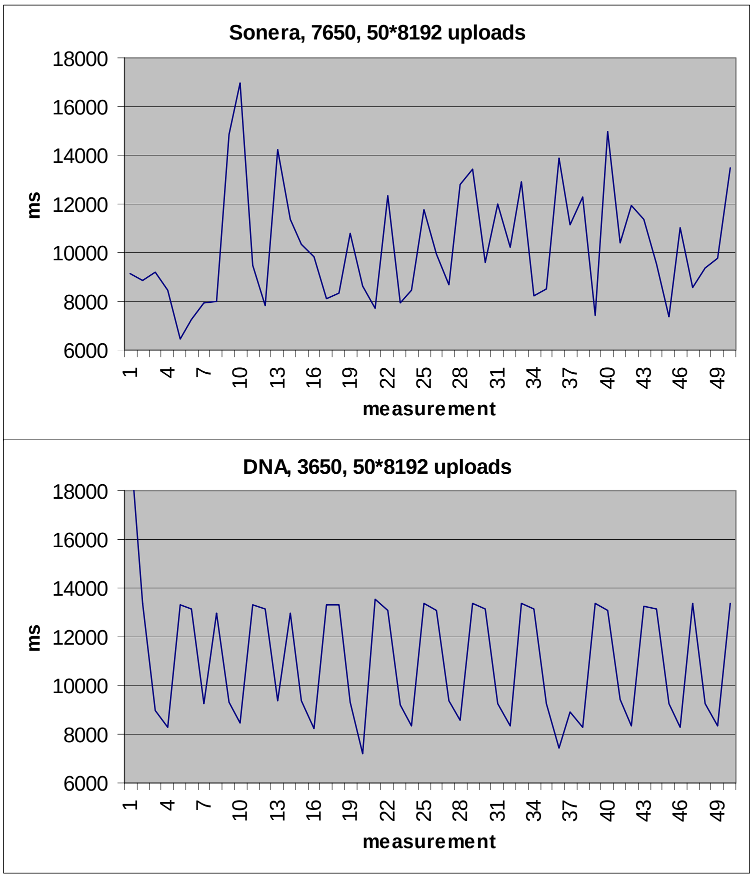

Figure 3.8. Mobile upload measurements.

In the GPRS network, PDCH (Packet Data Channel) can be in "dedicated" mode or "on-demand" mode based on GPRS network configuration. Both Finnish operators are providing "on demand"-mode, in which the voice and data communications share the slots of the TDMA (time division multiple access) frame and GPRS data is sent only when free slots are left over by the voice traffic (Porcarelli, et al.). Packet Data Channel is mapped to a single time slot of TDMA frame. One TDMA frame consists of 8 slots and requires one carrier channel from total of 124 carrier channels over 25 Mhz band. The BTS (base station) determines the number of available GPRS data channels. Each PDCHs is capable of transmitting 9, 12, 14 and 20 Kbps depending on the error coding scheme (CS). In GSM/GPRS, the coding scheme vary from CS-1 to CS-4 depending on the link quality, CS-4 being for the best connection (20 Kbps) with minimal extra coding and CS-1 low quality connection with good error coding (8 Kbsp).

The data transfer is also limited by the amount of transceivers in the terminal. If terminal has 3 transceivers it can support GSM call and GPRS at the same time, with 2 transceivers, it can transmit and receive GPRS data during the same timeslot, and with one transceiver it can only send or receive during a timeslot. For example a 4+1 terminal with 1 transceivers have 4 changes per frame to receive data and 1 change per frame to transmit data. The "free" 3 slots per frame can be used for signalling with basestation, maintaining the connection, and monitoring other cells for possible handover (Andersson et al. 2001).

Since one TDMA frame consists of 8 slots and requires one carrier channel from total of 124 carrier channels over 25 Mhz band. The BTS determines the number of available GPRS data channels. A theoretical maximum is reached with CS-4 coding with a theoretical mobile supporting 8 time slots per frame, which would result throughput of 160Kbps. Typically, a mobile may be using CS-2 and it supports four time slots (3+1), so this will result a maximum transmission speed of 48Kbps (Andersson et al. 2001).

In the measured results, the type of GPRS terminals was 3+1, this is clearly shown in measurements between uplink and downlink. Uplink data transmission is about 3 times slower than downlink. The average times of downlink access was 4.1 seconds in Sonera and 7650, and 3.2 seconds in DNA and 3650. Corresponding uplink times were 10,2 and 11,0 seconds in average.

4. Positioning technologies in wireless networks

4.1 Outdoor

Suvi Juurakko

Location techniques in today's and in the next generation cellular networks have gained extensive interest among cellular operators and manufacturers. Behind the interest towards technology deveplopment is the need to be able to offer new kind of commercial and non-profit services for mobile terminals. The non-profit services may include e.g. emergency call location, which has been mandated with high accuracy requirements in the US. Also in the EU a directive for emergency location was given in July 2003, although without accuracy or technical restictions. In commercial services the knowledge of the mobile terminal user location is applicable e.g. to transport logistics, yellow pages like applications, friend, personnel or children tracking, etc. Currently, the scope of commercial services is, however, quite narrow and only the simplest location technique, namely cell identification (CID) is typically used by the operators.

4.1.1 Location techniques

As mentioned the simplest method for location is the CID in which some pre-determined coordinates, e.g. coordinates of the serving base station or the centre of mass of the cell coverage, are given as the location estimate. The accuracy of CID depends directly on the cell size and therefore it may be very poor in rural or suburban areas (200 m–30 km). On the other hand, the CID is directly available in all networks and it can locate any mobile in the network including the legacy phones, which are the main reasons behind the popularity of this technique.

A major part of the other conventional outdoor location techniques rely on time of arrival (TOA) measurements between the mobile terminal and several base stations. In these cases, the location is typically computed by triangulation i.e. location computation requires measurements from at least three base stations. The more advanced timing based techniques, such as Observed Time Difference of Arrival (OTDOA) in UMTS and Enhanced OTD (EOTD) in GSM, are based on measuring time differences between base station pairs, but even these do not perform optimally in dense urban areas due to rich multipath propagation distorting the measurements. In CDMA networks (e.g. UMTS) the so called “hearability problem” may also prevent location estimation by triangulation based methods. This happens, if the mobile terminal can not receive the required amount of base stations due to e.g. high interference or minimized pilot pollution in the network planning phase.

The most accurate (2–20 m) location can be obtained by satellite based Global Positioning System (GPS), which requires a GPS-receiver in the mobile terminal. Currently, the density of GPS-capable terminals is, however, very low due to increased terminal prices, sizes and power consumption. Location reliability and availability of GPS in urban environment is also often compromized due to blocked line of sight paths to satellites. Nowadays, the trend seems to be towards network Assisted GPS (AGPS). It is gaining ground in the location industry due to smaller and cheaper GPS chipsets and wider availability compared to the traditional GPS. This accords permission to expect much higher penetration of AGPS-capable mobile terminals in the future. When considering the current situation and near future, location techniques based on standard network signals remain the main target due to their capability to support legacy phones, thus, offering much wider user penetration than AGPS.

4.1.1.1 Accuracy

Overall, location accuracy depends mainly on measurement accuracy and service environment but also the sensitivity of the used location algorithm to errors in the measurements or variations on the propagation path has an effect. An overview of the location accuracy of different techniques in different environments is given in Figure 4.1.

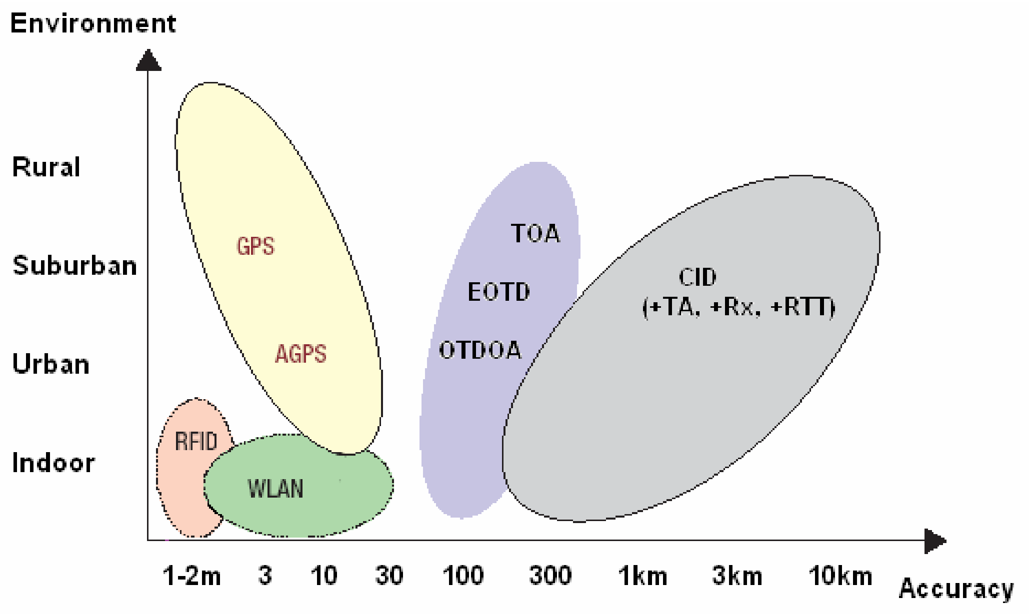

Figure 4.1. Accuracy of location techniques in cellular networks. Accuracy depends mainly on the environment. Typically, urban areas are most problematic due to multipath propagation.

4.1.1.2 Research goal in AuRa

The aim of the outdoor location research in AuRa was to develop a location method, which can overcome the problems encountered by conventional location techniques, especially in urban areas. The target accuracy range was the gap between the GPS based techniques and conventional techniques seen in Figure 4.1 and the development was targeted for UMTS and other wideband systems of the third generation.

The idea of the developed method is to take advantage of wideband channel delay profiles in the Database Correlation Method (DCM), developed in VTT Information Technology. This location method is based on *a priori* measured or computed signal database, which contains measured or computed signal fingerprint information from the service coverage area. Location estimation is then based on a correlation type algorithm between a measured fingerprint and the database.

4.1.2 Test environment

Urban areas can be expected to be the most favourable when considering mass market targeted location based services. Location estimation in these areas is, however, problematic independently of the used technique. Due to complex propagation environment and blocked visibility to satellites, both satellite and conventional network based techniques may fail.

In the project the test network was constructed in the centre of Helsinki in 1 km^2 area by a network planning tool that features full 3D-ray-launching method. The test environment includes the network and the UMTS system level environment. The network is illustrated in Figure 4.2 consisting of 24 microcell base station sites with 1–3 directional antennas in each. The signals were transmitted at the UMTS downlink frequency 2.15 GHz with 33 dBm transmission power. The range of antenna heights was 8…20 m depending on the surrounding building rooftop levels. The system level implementation and location algorithms we implemented in Matlab®.

In the urban microcells radio propagation is characterised by multipath propagation, shadowing and pathloss. The multipath channels of the test network were modelled by computing channel impulse responses between the transmitter (cell antenna) and the receiver (mobile terminal) with the ray-tracing tool. The used receivers were defined as a grid. The size of this receiver grid (inside the square in Figure 4.2) is 57^2 points with 12.5 m resolution and the receiver, i.e. mobile terminal, height was 1.5 m. The computed impulse responses were then used to simulate the delay profiles at the receiving UMTS terminal.

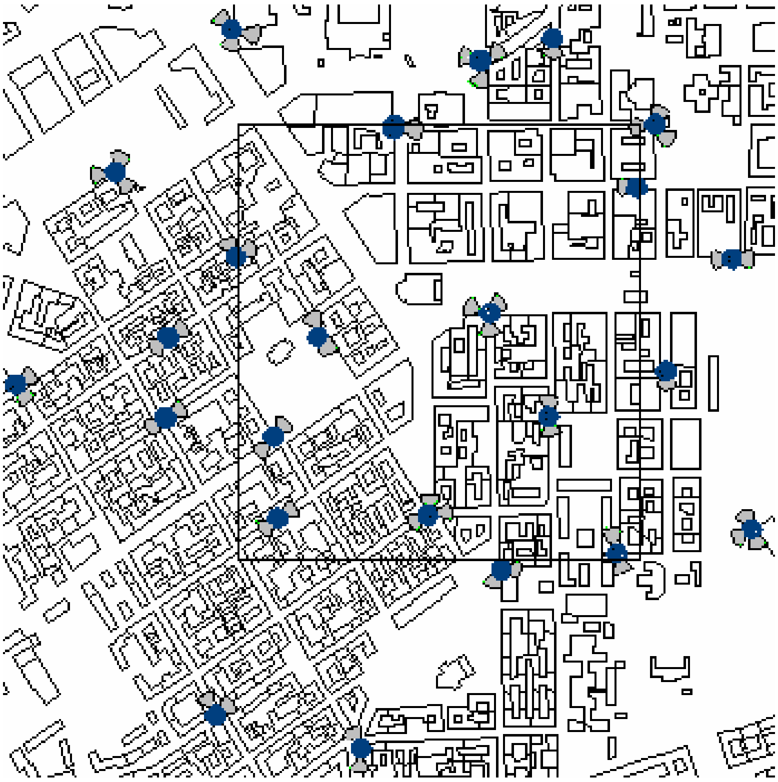

Figure 4.2. The microcell test network base stations in the centre of Helsinki. The receiving grid area is defined by the square.

In UMTS the time delay estimation is based on the common pilot channel (CPICH) signal, which is unique for each cell sector and was used also in this research task. The created simulation environment consists of the transmission code generation for CPICH (spreading, scrambling and pulse shaping), radio channel, delay estimation and location algorithms. The main system level parameters are given in Table 1.

Table 1. System level simulation parameters.

Parameter	Value
Carrier frequency	2.15 GHz
Chip rate	3.84 Mchip/s
Oversampling rate	4
Modulation scheme	QPSK
Pilot length used for	3×256 chips
Pulse shaping roll off	0.22

The influence of the modelled radio channels on the transmitted codes was achieved by convolving the codes with the calculated impulse responses. The impulse response data from the test network was saved in Matlab® format.

4.1.3 Location algorithm development

The aim was to develop a robust location algorithm that is able to locate a mobile terminal based on power delay profile (PDP) measurements. Thus, the solution can offer high tolerance for multipath propagation and therefore, better accuracy compared to the conventional techniques. The developed DCM algorithm uses a database of previously computed delay profiles (by ray tracing), but also measured delay profiles can be used. The location estimate is determined by a correlation algorithm between the database and a measured delay profile (by the mobile terminal).

Due to the use of delay profiles, the developed location algorithm is insensitive to small changes in the propagation environment, such as shadowing by vehicles. The rich multipath structure of urban areas can even improve the accuracy by contributing to more unique delay profiles. Therefore, it is possible to locate a terminal using measurement from only one base station and the common hearability problem in CDMA networks can be avoided.

The DCM algorithm can use several measured values (here PDPs) for location. In the research it was noticed that the most reliable results are, however, obtained by using only the measurement results from the strongest cell. This is, since the strongest cell had often in a line of sight path to the reception point, contributing to one unique and distinctive peak in the delay profile. The correlation between the received delay profile and the database was computed by the algorithm and the receiving point with the highest correlation coefficient was taken to represent the coordinates of the mobile terminal. The principle is illustrated in Figure 4.3.

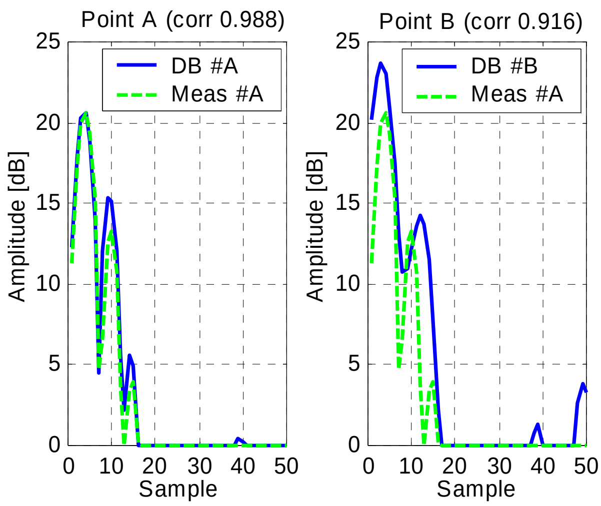

Figure 4.3. The cross-correlation between the measured delay profile (dashed line) and the delay profiles in adjacent database points, A and B (solid lines). The location estimate is given based on the correlation. Here the point A coordinates are chosen (true location).

Also another algorithm was implemented for comparison reasons. The chosen technique was OTDOA which is standardized by 3GPP for UMTS. The OTDOA location algorithm is based on hyberbolical triangulation. Location computation requires time differences of signals of base station pairs measured at the mobile terminal. Time-difference measurements of combinations of three base stations provide enough information for a two-dimensional coordinate solution. Therefore, as mentioned earlier, the hearability problem may prevent obtaining the required amount (≥3 hearable base stations) of measurements for OTDOA.

4.1.4 Location simulations

Location accuracy was evaluated by simulations made in the receiving grid area of the test network (inside the square in Figure 4.2) and the location estimates were computed only at the outdoor points.

The delay estimation for the OTDOA algorithm was made at the reception by correlating the received signal with the code used on transmission. Thus, the PDP of the channel was obtained. A threshold value was used to reduce contributions of noise power and interference from other codes. Under the determined threshold the signal was regarded as null and the part of the delay profile above the threshold was stored into the database. The maximum peak of the correlation output obtained from these cut PDPs was taken to indicate the TOA of the received signal. The DCM algorithm, on the other hand, took advantage of the entire measured PDP above the threshold.

The DCM simulations require two realisations of the channel at each receiving point. The first, which is directly given by the ray-tracing tool, is used to calculate the PDP database. The second realisation is needed to simulate the PDP measured by the terminal to be located. Fast fading and shadowing effects were taken into account in the second realisation by randomly changing the amplitudes and phases of the channel impulse responses. The modelled random attenuation used for each ray is a uniformly distributed random number between 0 and 1.

4.1.5 Location accuracy

The simulations of the DCM and OTDOA algorithms in the test network revealed a major difference in the accuracy. The used performance measure is the cumulative distribution function (CDF) of the absolute position error in metres. In Figure 4.4 are presented the CDFs of the location error for the OTDOA and DCM simulations.

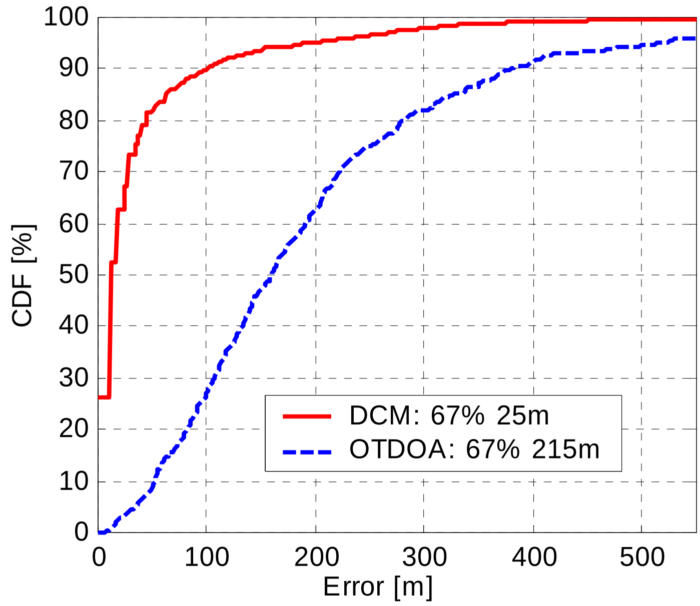

Figure 4.4. Cumulative distribution of location error in dense urban environment. Dashed line: OTDOA (67 % error, 215 m), solid line: DCM (67 % error, 25 m).

It can be seen from the results that multipath propagation and non-line-of-sight (NLOS) situations in the dense urban environment degrade the performance of the OTDOA algorithm. The time of arrival estimation process in multipath conditions is very difficult degrading OTDOA accuracy due to unrealistic delay values compared to the direct route delay value. DCM, on the other hand, seems to be insensitive to small changes in the propagation environment, such as shadowing by vehicles. This was tested by adding attenuation to certain directions or blocking the incoming signals from a certain sector completely. The DCM algorithm was proved to be able to locate a terminal using only one base station, therefore, avoiding the hearability problem common to OTDOA.

4.2 Indoor

Ville Haataja

In Aura project indoor location methods were studied and in this chapter we concentrate to a WLAN and Bluetooth based indoor positioning methods and solutions. WLAN positioning algorithm is presented and applicable positioning algorithm for Bluetooth environment is studied.

4.2.1 WLAN positioning

There are two main methods for acquiring location information based on measured WLAN access point signal strengths (RSSI). Both require measuring the signal strength of at least two, preferably three or more, WLAN access points. The first method, propagation model method, uses theoretical model of attenuation of the signal strength as the basis of the location calculation (Latvala et al. 1999). In its simplest form, signal strength is assumed to be a function of the distance to the access point. This method can be enhanced by including some assumptions about the effects of antenna, walls and other structures to radio propagation. This makes calculations laborious and some error sources remain, since the effects of e.g. walls are not easily modeled.

The second method, called location fingerprinting, is based on learning process (Bahl & Padmanabhan 2000). The signal strengths associated with a number of locations are measured in the teaching phase and recorded into a table containing the co-ordinates of each location, the signal strengths and possibly some related semantic information, such as the name of the room. The locating step uses k-nearest neighbors algorithm for determining an unknown location based on measured signal strengths and the table.

As the basic solution, the developed location method uses location fingerprinting to which some restrictive properties has been added from the propagation model method. Location fingerprint method uses one to three nearest neighbors to determine the location of the terminal. A Kalman-filter was added to the positioning system to estimate target co-ordinates and velocity and also to filter noisy signal strength measurements. Positioning method is presented in Figure 4.5. Larger circulars describe terminals maximum distances to the access points and these circulars are used to restrict the area where to find location fingerprints. The user's last known location is also used for the same purpose and it is described by the smaller circular. Black points are locations where the location fingerprints are measured.

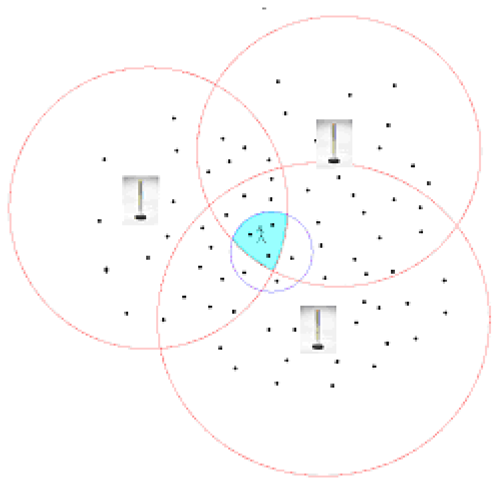

Figure 4.5. Positioning method for WLAN.

Requirements for using WLAN based positioning method are the WLAN infrastructure which contain at least two, preferably three, WLAN access points and of course the terminal with WLAN card. By increasing the amount of access points, larger location and WLAN area is covered and the positioning becomes more accurate.

In test cases average positioning accuracy have been about 2 meters, see Figure 4.6 where is described test case for a moving person. A test set-up consisting of three WLAN access points was built in a typical office building having long corridors and rooms along them. The area was L-shaped with dimensions of ca. 30 x 60 m and an area of 1000 m2, see Figure 4.6. The learning phase consisted of giving the co-ordinates of 240 points. The performance test included measurements in stationary situation and while moving at slow walking pace (0.4 m/s). For the stationary situation, the average error was below 1 m and worst case error was 1.4 m. The average error increased to over 1.5 m when the terminal was moving during measurement. Figure 4.6 represents positioning accuracy for the moving person. Straight line describes the true walking route and other line is calculated route from the positioning system. The increase in average error is due to the fact that the individual measurements are done at 0.3 s intervals and the averaging algorithm uses one to three nearest values for calculation. If three values are used, they actually represent three different locations and if only one value is used, the stochastic error of individual measurement is not smoothed by averaging. For this application, in which it is most important to locate the room correctly, the accuracy of both stationary and mobile cases is sufficient.

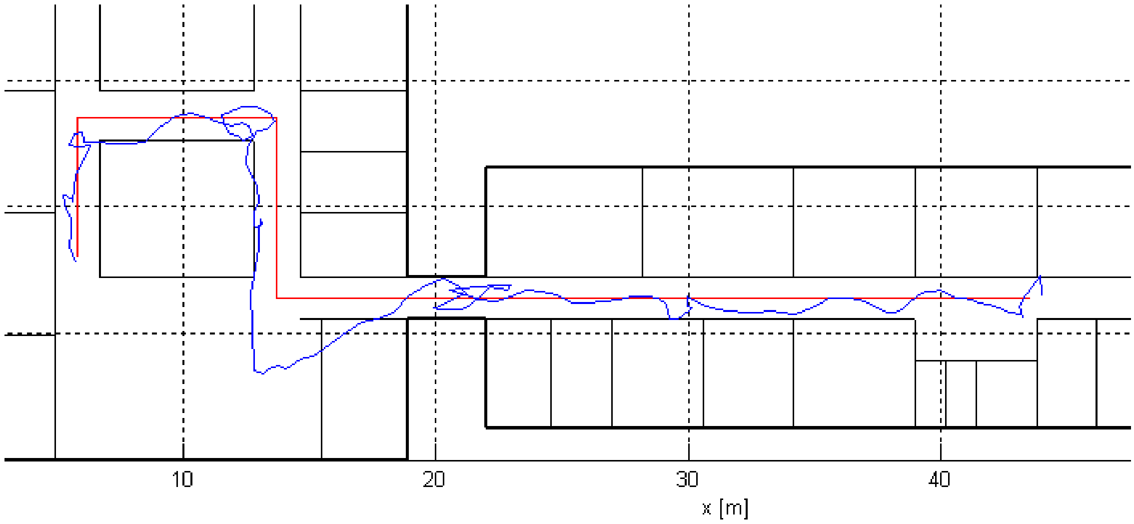

Figure 4.6. Positioning accuracy in WLAN based system, straight line is real route for a moving person and another line is positioning result get from WLAN positioning software.

WLAN-positioning software often consists of two modules: measurement module and positioning module. The measurement module resides in the terminal (a PDA) and it measures the access point signal strengths and sends them to the server for the calculation. Positioning module, residing in the server, performs the positioning computation and determines the estimated location of the terminal.

In Aura project there is made some improvements to this WLAN positioning software like positioning algorithms, usability and new Configuration Tool. Previous version of Configuration Tool was implemented by Matlab and the new one made in Aura is implemented with Visual C++. Configuration Tool is software which is needed to measure signal strengths used with location fingerprinting method.

Instead we concentrated on WLAN positioning and improvements in its algorithms and usability. Current WLAN positioning system can now use dynamically up to 25 base stations for determining the location with average accuracy of 2 meters.

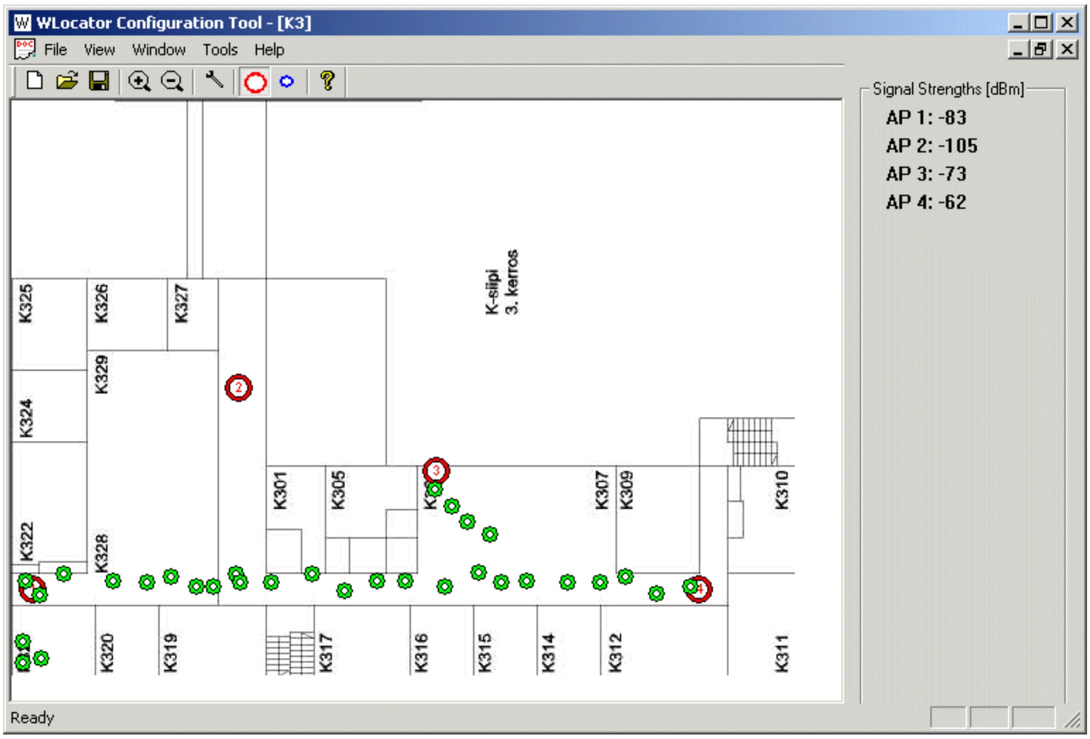

Figure 4.7. Configuration Tool Screenshot.

In Figure 4.7 is presented a screenshot about Configuration Tool. Bigger circles shown in picture describe locations of WLAN access points and smaller dots are locations where signal strengths are measured. Typical amount of measured points is 200 points / 1000 m2 but the more measured locations are in the area the more accurate results can be acquired from positioning engine.

4.2.2 Bluetooth positioning

WLAN-positioning method is quite progressive so what comes easily into mind is to study if it is possible to use same location fingerprinting technology with other wireless communication technologies like Bluetooth.

However the problem with the Bluetooth as a positioning system lies in its connection oriented nature in communication. This means that to measure the signal strength from base station to client the connection must be established and running. Any other signal strength measurement during that time is not possible. In principle one workaround that could pass problem is to make the system connect sequentially to each Bluetooth base station and measure the radio signal strength indication (RSSI) value during that short connection. However this causes slow functionality in positioning results. Measurement process could take several seconds but acceptable time for that is below 1 second.

Another problem is that the RSSI values are scaled between -10 and +10 where the normal connection quality stays at value 0. Practically this means that Bluetooth is not very usable for same sort of positioning purposes as WLAN. WLAN provides larger RSSI value scale making it easier to determine signal strength variation in different locations in the area.

However the Bluetooth base stations can be used to positioning purposes if base stations themselves contain information about their location. For instance the base station ID can be set to contain a name of a specific place, e.g. office. The client device can detect this base station ID and during connection.

The Bluetooth technology was studied in Aura project but it was discarded for the reasons listed above. But it is useful for purposes where we could use base station ID based location information for example we need to know room name where we are or some information about the room. WLAN positioning system is better for heavier systems where we need continuous positioning and tracking. Example case could be some moving person in factory or moving vehicle in factories or mines.

5. Wireless network routing

5.1 Ad hoc routing

Ville Typpo

The mobility of independent wireless hosts has gained more and more interest in telecommunication research. The most common approach is to provide network connectivity to mobile users moving in the coverage area of base station(s). However telecommunication research has also focused to study independent, self-organising networks between several mobile hosts. These kinds of networks are called ad hoc networks.

Ad hoc wireless mobile networks are self-organising and self-configuring networks. They can be established anytime and anywhere, without the static radio base stations and/or fixed backbone network connection(s). Mobile nodes themselves are maintaining the network, being able to both send/receive data and route the data between other hosts of the network. Naturally the ordinary routing methods are not applicable here, but there is wide spectrum of specific ad hoc routing protocols defined (Feeney 1999, Toh 2002, Perkins 2001). It is worth of noting that several mobile nodes are always needed to be present in order to create the ad-hoc network.

The concept of ad hoc network with wireless nodes requires also power supply independent of any electrical network. Mobile nodes today are powered with their own batteries. Efficient use of battery energy is therefore an essential issue. Battery life is not only effecting to mobile nodes own active operating time, but also to overall network communication performance of the whole ad hoc network.

In this section we take a look on different methods in order to provide power-efficient ad-hoc networking in a variety of usage scenarios based on the work presented in Sukuvaara (2002). The methods developed for advanced routing and MAC level operation will be covered.

5.1.1 Energy-efficiency within ad hoc networks

Many issues on energy efficient ad hoc networking are rather widely studied. Some overviews for the research have also been done, one of the latest presented in Goldsmith & Wicker (2002). Originally ad-hoc networking research had not emphasis on the energy efficiency, but later on it has been noticed to be an important element in the most ad-hoc networks. Usually the energy efficiency research in ad-hoc networking has focused to provide individual solutions to problem areas on routing, data processing, transceiver controlling etc. As stated in Goldsmith & Wicker (2002), truly energy efficient operation requires all the networking elements to work together. However, there are no this kind of "combined" solutions available yet.

There are different approaches for providing energy efficiency. For example, energy efficient processing techniques are needed to minimise the energy spent for different computational operations inside the networking device. The radio transceiver must be controlled by MAC in a way that transmission and receiving times are optimised. The ad hoc routing must be realised in an energy efficient (and fair) manner. Efficient coding, adaptive resource allocation, channelization, scheduling and the use of multiple antennas may also advance energy savings. The cross-layer design of these issues is also one important approach. A general approach might be derived, but such will hardly be an optimal solution for the many different use scenarios of ad hoc networks. Rather, an energy efficient solution could be biased differently for each use, depending on the desired amount of traffic, the level of mobility, and the number of hops in the network, a couple of criteria to mention.

Usually the energy efficiency is more or less provided by trading off the connectivity. Good connectivity requires constant listening of the network, quick response to connection request(s) etc., while the power efficiency basically means as few communication operations as possible. Here we aim to maintain the connectivity of every node in acceptable level, still keeping the energy efficiency as a primary goal.

In AuRa project we have studied several approaches for providing energy-efficient features in ad hoc networks. Following sections present the most promising methods proposed in the literature. We have tried to identify the most optimal solutions for the different use scenarios.

5.1.2 Traditional ad hoc networks

A traditional ad hoc network means here simply the general purpose autonomous ad-hoc network, not specifically optimised to any use scenario. In this kind of network, the effective routing and network operability has to be maintained on an acceptable level, regardless of the power optimisation. Therefore, the methods suitable for this purpose are power-efficient and fair routing methods and MAC level transceiver powering-off methods. The latter one must not save power at the cost of connectivity.

As a routing method, either energy aware routing with route probabilities (Shah & Rabaey 2002), Conditional Max-Min Battery Capacity Routing (CMMBCR) (Toh 2001, Toh et al. 2001), or minimum total transmission power routing (Scott & Bambos 1996) are considerable solutions. Based on comparisons in (Safwat et al. 2002) the last one performs well in most cases. However it lacks the means for ensuring fairness of power consumption between nodes. The other approaches deal with the fairness in routing task allocation too, being therefore more applicable to general purpose networks. Energy aware routing with route probabilities outperforms CMMBCR in simplicity, which could make that first one a primary choice.

MAC related transceiver power control is the other way to enhance the power efficiency of the traditional ad hoc network. As we are concerned about the undisturbed operation of the ad hoc network, several MAC mechanisms proposed in the literature are not valid here. The Dynamic Switching-off Network Interfaces (Cano & Manzoni 2001) seems to offer the best solution. Here the node hearing the RTS/CTS sequence (in IEEE 802.11 type of system) and receiving the header field of the first data packet, can turn its transceiver off until the end of the data packet transmission, whenever the packet is not destined to itself. Cano and Manzoni (2001) have reported up to 60 % energy savings achieved by utilising this mechanism. This is a rather simple enhancement to power efficiency, not considerably decreasing the overall performance of the ad hoc network.

5.1.3 Ad hoc networks with mixed-capability nodes

These ad hoc networks consist of two types of network nodes: traditional mobile ad hoc nodes with limited power resources and (fixed) nodes with unlimited power being connected to fixed power line or having a generator or a very large battery compared to the nodes with limited power resources. In this kind of system the power-unlimited nodes can take the most of the routing tasks in order to save the batteries of the power-limited nodes.

The approaches presented for the traditional ad hoc networks are basically valid here as well, but there are some other methods exploitable by the power-unlimited nodes. The Span (Chen et al. 2001) method is more related to sensor networks, but when matching the elected coordinators, which are responsible of routing tasks, with the power-unlimited nodes, we end up into routing mechanism where the routing load of the power-limited nodes is minimised to transmission of their own traffic only. If some of the elected coordinators are power-limited (a non-ideal case), a re-election procedure need be implemented in order to ensure fair power consumption for the power-limited nodes.

Turning the transceiver off occasionally, can be utilised as well, although not as straightforwardly as in the Span approach. Instead, we could dynamically switch-off the network interfaces of power-limited nodes, whenever it is known that no transmit/receive operations need be executed. As described in Cano & Manzoni (2001), these moments can be identified by overhearing the beginning of a transmission destined for another node. Then, it is safe to switch off for the rest of the transmission.

5.1.4 Single-hop communication networks

A single-hop communication network is formed among devices in the single-hop distance of each other. This network can be seen as master-slaves type of network, where single master control the network of several slaves. The distinction between the forthcoming single-hop sensor networks lies in communication capacity requirements. These networks require higher bandwidths for unforeseeable traffic patterns, while sensor networks usually provide periodical information with minimum communication capacity. Furthermore, the sensor networks are highly power optimised, while energy efficiency requirements might not be as crucial here.

A general solution, which has gained markets already, is the Bluetooth (Bluetooth SIG 2001). With gross bit rate of 1Mbps and support for both audio and data traffic, Bluetooth offers a reasonable solution for single-hop communication networking.

Other upcoming technologies include the work done in the several task groups of IEEE 802.15 (WPAN) working group. IEEE 802.15.1 follows the Bluetooth standard in compliance with IEEE 802 standard family. IEEE 802.15.3 together with WiMedia alliance aims to define high rate Wireless Personal Area Network technology able to reach 55 Mbps data rate. First products are expected in the second half of year 2004. The extensions planned in task group IEEE 802.15.3a will provide even 480 Mbps data rate with short distances. Another line of work is being carried out in IEEE 802.15.4 along with ZigBee alliance, aiming to low power and low data rate (below 250 kbps) WPAN. The draft standard has been completed and the first products are expected in the beginning of year 2004.

5.1.5 Single-hop sensor networks

As the sensors are typically really small, they have to be simple devices with simple communication facilities. The data throughput requirement is moderate, consisting of periodical minimal data burst (from several bits up to few kilobits). The energy efficiency is the main optimisation issue, communication must be realised with minimal power consumption.

For this kind of environment the Energy-Efficient Permutation Routing presented in Nakano et al. (2001) is probably the most efficient one. Here each node of the sensor network stores its data (to be transmitted) and keeps the transceiver off until certain predefined time window, in which one node at a time will transmit its data to the central unit. Energy efficiency can be increased by adjusting the sleep period at the cost of buffering requirements in the nodes and latency of the data delivery.

5.1.6 Multihop sensor networks

Multihop sensor network has basically the same requirement of extreme efficient power usage, but here the sensors must be able to route the sensor information. Therefore, power optimisation can not be as efficient as in the previous sub-chapter, but it is still clearly the main issue. With this scenario we make a distinction between static (or almost static) and mobile sensors, which have rather different optimal solutions.

In the case of static (or almost static) sensor network, the dominating set based routing presented in Wu et al. (2001) provides probably the most efficient solution. In this solution, nodes are dynamically divided into local subgroups, in which the routing to other subgroups is handled by the (selected) central nodes. The central nodes are changed from time to time, avoiding the fading of the routing nodes, and allowing low mobility (almost static) of nodes.

With relatively high mobility, this solution is not efficient anymore, since especially the movement of central nodes is clearly affecting to the performance. In the (highly) mobile sensor networks, the most appropriate solution is directed diffusion paradigm presented in Intanagonwiwat et al. (2000). In this method the route request is flooded to the network, and response will be delivered through the best route only, after which the transmission can be completed. The use of flooding in route acquisition may seem inefficient from the power consumption point of view, but in this case when the old route is most probably unusable due to mobility, the flooding provides the simplest (an that way efficient) way to achieve a route for data transmission.

5.1.7 Static or mostly static ad hoc networks

This is a special case of ad hoc network, in which it is known beforehand that the nodes are not moving, or move so slowly that the network structure is changing only occasionally. In this case we can accept more power-consuming route creation procedure, if it leads to more power-efficient routes.

For this purpose the methods presented for traditional ad hoc networks in section 5.1.2 are valid here as well. However, here the operation can be made more effective by employing minimal power level to reach the next-hop node. One solution providing this kind of property is presented in Wu et al. (2000), occupying separate signalling channel for RTS/CTS communication. Based on the signal strength on the separate control channel the actual transmission power can be adjusted. However, this solution requires rather complicated transceiver, and may therefore become unattractive.

5.2 Ameba – graphical user interface for network simulation

Mikko Majanen

In this section a graphical user interface called Ameba for network simulation is introduced. Network simulations are typically discrete event simulations. That’s why there is a short introduction to sequential and parallel discrete event simulation techniques at first.

5.2.1 Sequential discrete event simulation

Simulation is the technique of using computers to imitate the operations of actual or theoretical facilities and processes (Liu 2001). In a simulation model, these operations are abstracted using mathematical relationships in terms of state variables. The evolution of the simulation model is specified as state changes over time. The time advance function describes how time progress in the simulation and according to it simulation models can be divided into time-driven and event driven models.

In a time-driven simulation model state changes occur at small time steps, which gives an impression that the system evolves continuously over time. Time-driven simulation model is usually used when the system can be represented by differential equations that models the continuous time evolution of the system. (Liu 2001)

An event-driven or a discrete event simulation (DES) model assumes that the system changes state only at discrete points in simulated time (Fujimoto 1990). The simulation model jumps from one state to another upon the occurrency of an event. DES model is usually used when the system can be described by stochastic processes. Network simulations belong to this group, allthough certain aspects in these simulations, such as interference calculations in the simulation of wireless systems, can be better described by continuous time functions (Liu 2001). That’s why we concentrate on discrete event simulations in this section.

Figure 5.1 shows how a sequential DES usually works. It has three data structures (Fujimoto 1990). Firstly, the state variables that describe the state of the system, secondly, an event list that contains all pending events that have been scheduled, but not yet processed, and thirdly, a global clock variable that denotes how far the simulation has progressed.

Each event contains a timestamp and usually denotes some change in the state of the system. The timestamp tells when this change occurs in the actual system.

The “main loop” of the simulator repeatedly removes the smallest timestamped event from the event list and processes it. Then it updates the state variables and the global clock before processing the next event.

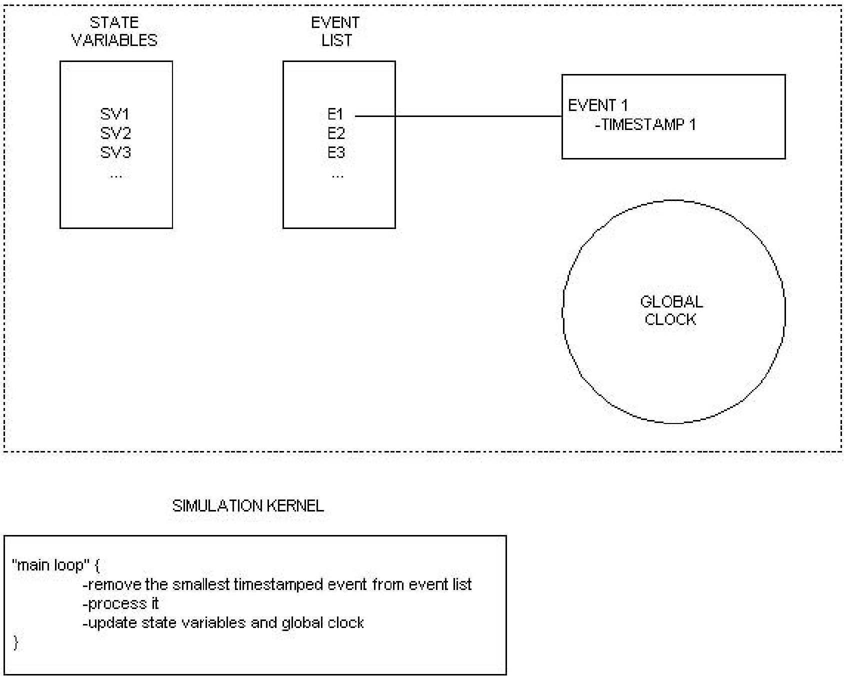

Figure 5.1. The sequential DES operating principle.

It is very important to always select the smallest timestamped event, because selecting larger timestamped event before smaller one might lead to a situation where future events could have affected the past! This kind of errors are called causality errors. (Fujimoto 1990)

5.2.2 Parallel discrete event simulation

Modeling large and multi-protocol communication networks such as the Internet or large-scale wireless ad hoc networks is very challenging due to their size, complexity and rapid change. Sequential simulation of such networks requires much of memory and CPU time, which can even limit the scale of the simulation to just hundreds of nodes. However, it is important to be able to simulate very large-scale networks consisting of hundreds of thousands of nodes, because some phenomena occur only at large-scale, for example congestion storms of the Internet.

There are two ways to improve the scale of the simulations: abstraction and parallel simulation. Abstraction means that unnecessary details of the simulation are eliminated.

This saves time and memory, but it may change the results of the simulation. Several abstraction techniques can be studied from (Huang 1999). Figure 5.2 shows how a parallel discrete event simulation (PDES) usually works.

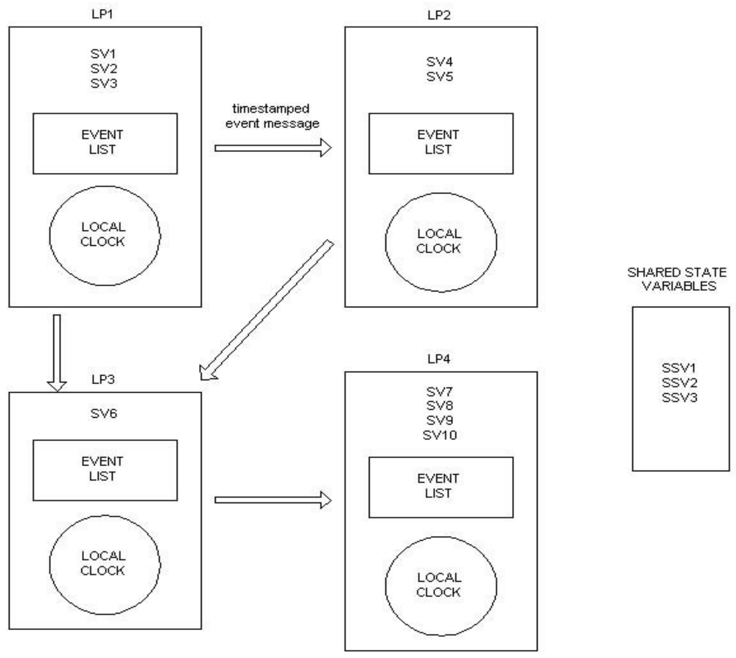

Figure 5.2. The PDES operation principle.

In PDESs the system being modelled, the physical system, is viewed as being composed of some number of physical processes that interact at various points in simulated time. The simulator is constructed as a set of logical processes LP_0, LP_1, …, one per physical process. Each LP can be distributed to be processed at different processor. All interactions between physical processes are modelled by timestamped event messages sent between the corresponding logical processes. Each logical process contains only the state variables that correspond to the physical process it models and a local clock that denotes how far the process has progressed. There is no global clock in parallel simulations. (Fujimoto 1990)

The so called local causality constraint is sufficient (but not always necessary) to guarantee that no causality errors occur (Fujimoto 1990): A (P)DES consisting of logical processes that interact exclusively by exchanging timestamped messages, obeys the local causality constraint if and only if each LP processes events in nondecreasing timestamp order.

The difficulty in PDES is how to ensure this nondecreasing timestamp order at each LP. An event E_1 at LP_1 may schedule an event E_x to LP_2 that has a smaller timestamp than the events already processed at LP_2. How LP_2 can know which events are safe to process? This is the fundamental issue that PDES strategies must address. According to this the PDES mechanisms are divided into the conservative and optimistic ones. (Fujimoto 1990)

In conservative mechanisms processes that do not have any safe processes to process must block. This prevents any causality errors ever occurring, but on the other hand, this can lead to deadlock situations if appropriate precautions are not taken. Lookahead plays an important role in conservative mechanisms. It refers to the ability to predict what will happen, or more importantly, what will not happen in the simulated future. Non-zero minimum timestamp increment, let's say *M*, is a good example of lookahead: the process can then guarantee that no new event messages will be created with timestamp smaller than *Clock+M*, so the process has a lookahead of at least *M*. Thus, lookahead can be used in determining which events are safe to process. (Fujimoto 1990)

Many conservative mechanisms require static configurations: one cannot dynamically create new processes, and the interconnection among logical processes must also be statically defined. The degree to which processes can look ahead and predict future events plays a critical role in the performance of conservative mechanisms. Simulation programmer must also be concerned with the details of the synchronization mechanism in order to achieve good performance. (Fujimoto 1990)

Optimistic mechanisms do not strictly avoid causality errors. Instead of determining when an event is safe to process, they just process events. If an event message is received that contains a timestamp smaller than the process's clock (this kind of event message is called a straggler) the process just undo all the effects of all events that have timestamps larger than the straggler. In order to be able to do this, the process must periodically save its state. (Fujimoto 1990)

Optimistic mechanisms can exploit the parallelism in situations where causality errors might occur, but in fact do not. The critical question in case of optimistic mechanisms is whether the system will exhibit thrashing behaviour where most of its time is spent executing incorrect computations and rolling them back. Optimistic algorithms tend to use more memory than their conservative counterparts because of the need to save the state of each logical process periodically. (Fujimoto 1990)

5.2.3 PDES languages and libraries

Developing a PDES from scratch requires enormous effort. Luckily, several PDES languages and libraries exist that help the developer by providing a pre-built parallel simulation kernel and application development tools. A simulation language usually provides a full set of well-defined language constructs for the user to design simulation models, whereas a library only provides a group of routines to be used with a base programming language such as C, C++ or Java. Libraries usually give the user more flexibility in controlling the simulation application in terms of the underlying synchronization protocols. Most PDES languages hide these synchronization protocol issues from the user. Examples of PDES languages are among others APOSTLE (Bruce 1997), Maisie (Bagrodia & Liao 1994), its object-oriented version MOOSE (Waldorf & Bagrodia 1994) and Parsec (Bagrodia et al. 1998), which is a next generation version from Maisie. PDES libraries include for example GloMoSim (Bajaj et al. 1999), SSF (Cowie et al. 1999) and PDNS (Riley et al. 1999), which is a parallelized version of the popular sequential ns-2 simulator (Breslau et al. 2000). (Low et al. 1999)

Despite the above, and many other, efforts of building parallel network simulators, most network simulations today are still performed using sequential simulators such as OPNET and ns-2. One important reason for this is that transitioning to a new simulator, especially to a parallel simulator, is often difficult. One must abandon previous investments in sequential simulators and master a new, unfamiliar simulation language. (Wu et al. 2001)

OPNET (Chang 1999) is a commercial network simulator and it consists of four major components: an event-driven simulation engine that manages an event queue and processes events in timestamp order, a set of application interfaces implemented as C libraries, graphical tools and commands, and a large library of network protocol models. (Wu et al. 2001)

Ns-2 is a sequential discrete event simulator targeted at networking research. It is publicly available and provides substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless (local and satellite) networks. It also supports abstraction techniques. It uses TCL scripting language as a user interface. The core of the ns-2 is implemented in C++. (Wu et al. 2001)

5.2.4 Ameba simulator

The Ameba (Ad hoc network simulation environment based on Beowulf cluster platform) simulator developed in the AURA project at the VTT Electronics is targeted to simplify the development and execution of network simulations. It was planned to use the PDNS as a simulation engine, but unfortunately at the time when Ameba was developed PDNS did not fully support wireless simulations. So instead of PDNS Ameba uses a sequential ns-2 for simulation execution. Ameba has a Java-based graphical user interface in which user can develop simulation models without being familiar with TCL or any other programming language whatsoever. Ameba generates the TCL files required by ns-2 automatically based on the information user has given. It could be said that Ameba is a graphical user interface for ns-2 simulator.

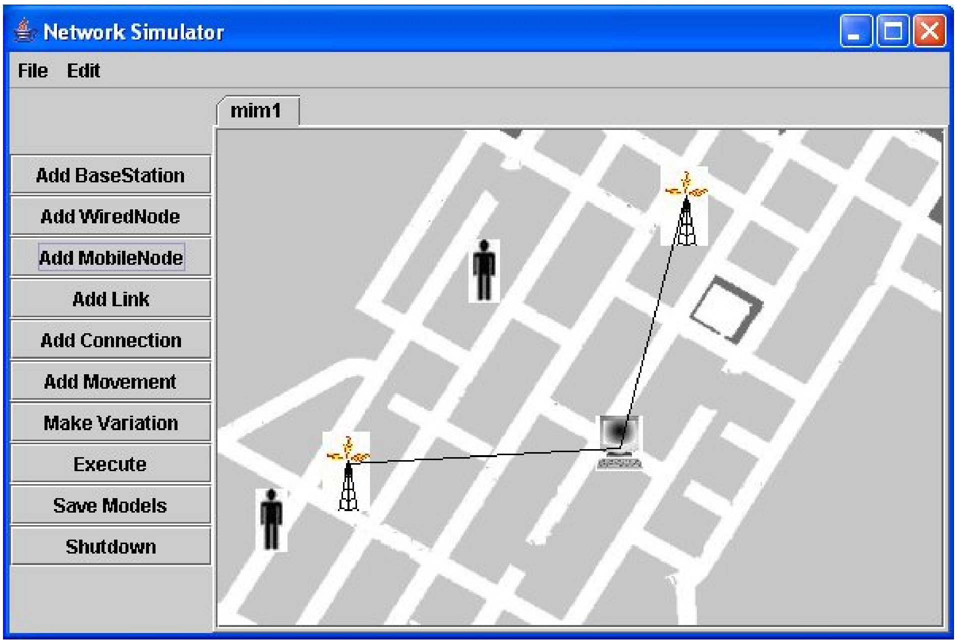

Figure 5.3. Screenshot from the Ameba simulator user interface.

Figure 5.3 shows a screenshot from the Ameba GUI. There is currently one simulation model named “mim1” opened. This model seems to have two mobile nodes (the human like figures), two base stations (the towers) and one wired node (the computer like figure). The base stations are connected to the wired node by links (the black lines). The map picture behind them helps in setting the nodes to the real world corresponding locations.

The several “Add” buttons on the left are used in adding (or editing/removing) objects to simulation model. Each button opens a new dialog window in which the user can add a new object, edit existing object or remove existing object. For example, in order to add a new base station into model, user must push the “Add BaseStation” button and give the name and coordinates of the new base station in the dialog window. Then the new base station appears on the map in the given location. The locations of different nodes can be changed either through the editing capabilities of the dialog windows or just dragging them to a new location on the map with mouse.

The same “Add” buttons can be found also from the “Edit” menu. There is also a possibility to set some general parameters of the simulation model, such as the simulation time and the length and the height of the simulation area. The “File” menu includes among others buttons to save and load simulation models.

Ameba was initially designed for the scenario of finding the best location for a new base station in wireless networks. The “Make Variation” button copies the currently active simulation model into a new tabbed window. Thus, after the base model is ready, it is easy to create a couple of variations from that by using this button. Then just do the changes to the variation models, for example to add a new base station, and execute the models by pressing the “Execute” button. Then the TCL files for each simulation model are generated and separate processes are launched to execute them with ns-2. These processes are processed sequential if only one processor is available. If more processors are available, several simulation models can be simulated parallel, each one on its own processor.

The master-slave model is used in parallelization. This means that one processor (the master) runs the Ameba GUI that manages the distribution of simulation processes and collects the simulation results. Other processors (the slaves) execute sequentially one simulation at time and send the results back to the master.

During the simulations the progress of simulations is displayed on the Ameba GUI. The results of the already simulated models are displayed. The results are sorted according to the average throughput of mobile nodes. After all simulations are done the results are also displayed on a web page, which contains in addition to the simulation results the details of each simulation model, their pictures, and links to the related ns-2 specific files (the model’s TCL script, trace files, etc…).

Ameba can be run on all platforms where Java Virtual Machine and ns-2 are available. Different kind of Unix platforms (Linux, etc…) are the most suitable ones for ns-2 but it has been seen to work also on Microsoft Windows platforms. The mpiJava library (Baker et al. 1999) is required for distributed execution. mpiJava requires a native MPI (Snir et al. 1996) environment such as MPICH (Gropp et al. 1996) or LAM (Burns et al. 1994). These are not available for Microsoft Windows platforms.

6. Discussions

In the project an efficient wireless networking scheme was researched and analysed for mobile terminals with advanced services and negotiated quality of service level. In the physical and data link layers it was observed that log-normal shadowing is very crucial for the spectral efficiency. In outdoor positioning advantages of wideband channel delay profiles were successfully utilised to locate terminals in UMTS network. For the indoor positioning a configuration tool was developed to make signal chart measurement faster with a more dense and accurate grid. The typical accuracy of positioning in indoor of 2 m was achieved. For vertical handover signalling schemes, key parameters mapping and classification for the related QoS issues were developed. In the QoS research an end-to-end QoS measurement tool was also developed to observe the performance differences of various terminals and operators of networks. On ad hoc and sensor networks power efficiency, routing and simulator tools were researched.

In each research topic several kind of idealisations was required to obtain results within a reasonable time. For example, in spectral efficiency research disturbances were assumed to obey Gaussian distribution based on the central limit theorem. However, we believe that idealisations made were reasonable and could give us a good view from the actual performance. Also the application and demonstration environments set their own limitations. Despite of that, the achieved results are supposed to be realistic and part of these were even verified via demonstrating these in real environment.

Wireless local area networks and ad hoc networking as its emerging evolution represent a new wireless revolution. They have potential to compete with wireless data technologies as cellular systems being offered by telecommunication carries. However, we are not likely to see a “winner-take-all” outcome because there seems to be a place for the cellular systems solutions as well. This is due to three facts. First, the structure of distribution areas of ad hoc networks with access points and gateways, and allocated frequencies will finally be as in the existing cellular systems because of cumbersome bandwidth requirement of routing protocols. Secondly, principles of seamless handovers will also finally be very similar as those in the existing cellular systems because of handover related signalling. Thirdly, increasing the user density to the cellular network level, the user payload data rate will dramatically drop even to a lower level than in the very optimised existing and emerging cellular networks due to the used medium access control (MAC) protocol. However, wireless local area and ad hoc networking will provide one significant advantage: they remove exorbitant overheads associated with acquiring radio spectrum, thereby lowering the barrier for the new entrants. This also drives product costs down. Furthermore, the increasing attractiveness of the integration of wireless ad hoc networks and cellular networks will guarantee the co-existence of both of them in the near future. Hence, it is very probable that the next generation of wireless networking is interoperation of networks instead of any new physical solution.

There may emerge several applications based on the research technologies: Spectral efficiency results can be used in optimisation of emerging networks. Positioning information can be used for location based services both indoor and outdoor environment. QoS monitoring tools may be used by users of cellular systems to verify their connection quality. Vertical handover research results will be utilised in emerging co-operation of UMTS and WLAN networks which might lead to countless applications.

The research of project topics will be continued in several VTT’s strategic technology theme projects during the next few years. Moreover, in order to reach long term results the span of this kind of projects should be long enough, preferably 5 than 2 years.

References

3GPP TS 23.060. 2002. General Packet Radio Service (GPRS) Service description - stage 2. Version 4.3.0 Release 4.

3GPP TS 23.107. 2002. Universal Mobile Telecommunications System (UMTS); Quality of Service (QoS) concept and architecture. Version 4.3.0 Release 4.

Abidi, A. A., Pottie, G. J. & Kaiser, W. J. 2000. Power-Conscious Design of Wireless Circuits and Systems. Proc. IEEE, Vol. 88, No. 10, pp. 1528–1545.

Alasalmi, J. 2002. Area Spectral Efficiency of Cellular Mobile Radio Systems. VTT Electronics. 43 p.

Alouini, M.-S. & Goldsmith, A. J. 1999. Area Spectral Efficiency of Cellular Mobile Radio Systems. IEEE Trans. Veh. Technol., Vol.48, No. 4, pp. 1047–1066.

Alouini, M.-S. & Goldsmith, A. J. 1999b. Capacity of Rayleigh Fading Channels Under Different Adaptive Transmission and Diversity-Combining Techniques. IEEE Trans. Veh. Technol., Vol. 48, No. 4, pp. 1165–1181.

Andersson, C. 2001. GPRS and 3G Wireless Applications: professional developer's guide. Wiley Computer Publishing. USA.

Bagrodia, R., Meyer, R., Takai, M., Chen, Y., Zeng, X., Martin, J. & Song, H. Y. 1998. Parsec: A Parallel Simulation Environment for Complex Systems. IEEE Computer, Vol. 31, No. 10, pp. 77–85.

Bagrodia, R. L. & Liao, W.-T. 1994. Maisie: A Language for the Design of Efficient Discrete-event Simulations. IEEE Transactions on Software Engineering, Vol. 20, No. 4, pp. 225–238.

Bahl, P. & Padmanabhan, V. 2000. RADAR: An In-Building RF-based User Location and Tracking System. IEEE infocom2000: Reaching the Promised Land of Communications, Tel Aviv, Israel. IEEE. Pp. 775–784.

Bajaj, L., Takai, M., Ahuja, R., Tang, K., Bagrodia, R. & Gerla, M. 1999. GloMoSim: A Scalable Network Simulation Environment. Technical Report 990027. UCLA, Computer Science Department.http://citeseer.nj.nec.com/225197.html

Baker, M., Carpenter, B., Fox, G., Ko, S. H. & Lim, S. 1999. mpiJava: An Object-Oriented Java interface to MPI. Presented at International Workshop on Java for Parallel and Distributed Computing, IPPS/SPDP.http://grids.ucs.indiana.edu/ptliupages/projects/HPJava/papers/ipps99/paper/paper.html

Bambos, N. 1998. Toward Power-Sensitive Network Architectures in Wireless Communications: Concepts, Issues, and Design Aspects. IEEE Personal Commun., Vol. 5, No. 3, pp. 50–59.

Biglieri, E., Proakis, J. G. & Shamai (Shitz), S. 1998. Fading channels: information-theoretic and communications aspects. IEEE Transactions on Information Theory, Vol. 44, No. 6, pp. 2619–2692.

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. & W. Weiss. 1998. An Architecture for Differentiated Services. RFC 2475.

Bluetooth SIG. 2001. Specification of the Bluetooth System, Version 1.1. URL: http://www.bluetoothsig.org/docs/Bluetooth_V11_Core_22Feb01.pdf

Braden, R., Zhang, L., Berson, S., Herzog, S. & Jamin, S. 1997. Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification. RFC 2205.

Breslau, L., Estrin, D., Fall, K., Floyd, S., Heidemann, J., Helmy, A., Huang, P., McCanne, S., Varadhan, K., Xu, Y. & Yu, H. 2000. Advances in Network Simulation. IEEE Computer, Vol. 33, No. 5, pp. 59–67.

Bruce, D. 1997. What makes a good domain-specific language? APOSTLE, and its approach to parallel discrete event simulation. POPL'97 workshop on domain-specific languages (DSL´97) http://citeseer.nj.nec.com/bruce97what.html

Burns, G. D., Daoud, R. B. & Vaigl, J. R. 1994. LAM: An Open Cluster Environment for MPI. Proceedings of Supercomputing Symposium ´94.

Cano, J. C. & Manzoni, P. 2001. Evaluating the Energy-Consumption Reduction in a MANET by Dynamically Switchin-off Network Interfaces. In: Proceedings of the 6th IEEE Symposium on Computers and Communications. Pp. 186–191.

Catreux, S., Driessen, P. F. & Greenstein, L. J. 2001. Attainable Throughput of an Interference-Limited Multiple-Input Multiple-Output (MIMO) Cellular System. IEEE Trans. Commun., Vol. 49, No. 8, pp. 1307–1311.

Catreux, S., Driessen, P. F. & Greenstein, L. J. 2002. Data Throughputs Using Multiple-Input Multiple-Output (MIMO) Techniques in a Noise-Limited Cellular Environment. IEEE Trans. Wireless Commun., Vol. 1, No. 2, pp. 226–235.

Chang, X. 1999. Network simulations with OPNET. Proceedings of the 1999 Winter Simulation Conference. http://www.informs-cs.org/wsc99papers/043.PDF

Chen, B., Jamieson, K., Balakrishnan, H. & Morris, R. 2001. Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks. In: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, July 2001, Rome, Italy. Pp. 85–96.

Chiani, M., Win, M. Z. & Zanella, A. 2003. On the capacity of spatially correlated MIMO Rayleigh-fading channels. IEEE Transactions on Information Theory, Vol. 49, No. 10, pp. 2363–2371.

Cowie, J. H., Nicol, D. M. & Ogilski, A. T. 1999. Modeling the Global Internet. IEEE Computing in Science & Engineering, Vol. 1, No. 1, pp. 42–50.

Cox, D. C. 1982. Co-Channel Interference Considerations in Frequency-Reuse Small-Coverage Area Radio Systems. IEEE Trans. Commun., Vol. 30, No. 1, pp. 135–142.

Crane, R. K. 1981. Fundamental Limitations Caused by RF Propagation. Proc. IEEE, Vol. 69, No. 2, pp. 196–209.

El Gamal, A. & Cover, T. M. 1980. Multiple User Information Theory. Proc. IEEE, Vol. 68, No. 12, pp. 1466–1483.

Ertel, R. B., Cardier, P., Sowerby, K. W., Rappaport, T. S. & Reed, J. H. 1998. Overview of Spatial Channel Models for Antenna Array Communications. IEEE Personal Commun, Vol. 5, No. 1, pp. 10–22.

Farrokhi, F. R., Lozano, A., Foschini, G. J. & Valenzuela, R. A. 2002. Spectral Efficiency of FDMA/TDMA Wireless Systems with Transmit and Receive Antenna Arrays. IEEE Trans. Wireless Commun., Vol. 1, No. 4, pp. 591–599.

Feeney, L. M. 1999. A Taxonomy for Routing Protocols in Mobile Ad Hoc Networks. SICS Technical Report T99/07. Swedish Institute of Computer Science.

Fujimoto, R. M. 1990. Parallel Discrete Event Simulation. Communications of the ACM, Vol. 33, No. 10, pp. 30–53.

Gesbert, D., Bölcskei, H., Gore, D. A. & Paulraj, A. J. 2000. MIMO Wireless Channels: Capacity and Performance Prediction. Proc. of IEEE Global Telecommun. Conf., San Fransisco, CA, USA, 27 Nov.–1 Dec. 2000. Pp. 1083–1088.

Golden, G. D., Foschini, G. J., Valenzuela, R. A. & Wolniansky, P. W. 1999. Detection Algorithm and Initial Laboratory Results Using V-BLAST Space-Time Communication Architecture. Electronics Letters, Vol. 35, No. 1, pp. 14–16.

Goldsmith, A. J. 1997. The Capacity of Downlink Fading Channels with Variable Rate and Power. IEEE Trans. Veh. Technol., Vol. 46, No. 3, pp. 569–580.

Goldsmith, A. J. & Varaiya, P. P. 1997. Capacity of Fading Channels with Channel Side Information. IEEE Trans. Inf. Theory, Vol. 43, No. 6. pp. 1986–1992.

Goldsmith, A. J. & Wicker, S. B. 2002. Design Challenges For Energy-Constrained Ad Hoc Wireless Networks. IEEE Wireless Communications, Vol. 9, No 4, pp. 8–27.

Goodman, D. & Mandayam, N. 2000. Power Control for Wireless Data. IEEE Personal Commun., Vol. 7, No. 2, pp. 48–54.

Gradshteyn, I. S. & Ryzhik, I. M. 2000. Table of Integrals, Series, and Products, Academic Press: San Diego, 6. ed.

Grilo, A., Macedo, M. & Nunes, M. 2003. A Scheduling Algorithm for QoS Support in IEEE802.11e Networks. IEEE Wireless Communications, pp. 36–43.

Gropp, W., Lusk, E., Doss, N. & Skjellum, A. 1996. A high-performance, portable implementation of the MPI message passing interface standard. Parallel Computing, Vol. 22, No. 6.

Günther, C. G. 1996. Comment on “Estimate of Channel Capacity in Rayleigh Fading Environment”. IEEE Trans. Veh. Technol., Vol. 45, No. 2, pp. 401–403.

Haagerup, U. & Thorbjørnsen, S. 2003. Random matrices with complex Gaussian entries. April 2003, unpublished material.

Hansen, R. C. 1981. Fundamental Limitations in Antennas. Proc. IEEE, Vol. 69, No. 2, pp. 170–182.

Harmuth, H. F. 1981. Fundamental Limits for Radio Signals with Large Bandwidth. IEEE Trans. Electromagnetic Compatibility, Vol. 23, No. 1, pp. 37–43.

Hashemi, H. 1993. The Indoor Radio Propagation Channel. Proc. IEEE, Vol. 81, No. 7, pp. 943–968.

Hatfield, D. N. 1977. Measures of Spectral Efficiency in Land Mobile Radio. IEEE Trans. Electromagnetic Compatibility, Vol. 19, No. 3, pp. 266–268.

Hollenbeck, K. J. 1998. INVLAP.M: A matlab function for numerical inversion of Laplace transforms by the de Hoog algorithm. Unpublished work.

de Hoog, F. R., Knight, J. H. & Stokes A. N. 1982. An improved method for numerical inversion of Laplace transforms. SIAM Journal of Scientific and Statistical Computation, Vol. 3, No. 3, pp. 357–366.

Huang, P. 1999. Enabling Large-scale Network Simulations: A Selective Abstraction Approach. Doctoral dissertation. University of Southern California. http://citeseer.nj.nec.com/huang99enabling.html

Intanagonwiwat, C., Govindan, R. & Estrin, D. 2000. Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In: Proceedings of IEEE/ACM Mobicom 2000 conference. Pp. 56–67.

Jayaweera, S. K. & Poor, H. V. 2003. Capacity of multiple-antenna systems with both receiver and transmitter channel state information. IEEE Transactions on Information Theory, Vol. 49, No. 10, pp. 2697–2709.

Kang, M. & Alouini, M.-S. 2003. Impact of correlation on the capacity of MIMO channels. In: Proceedings of International Conference on Communications, Anchorage, USA, May 2003. Pp. 2623–2627.

Kasapi, A., Da Torre, S. B., Roger, A.-F., Kerr, A. & Nolan, A. 2000. Massive Scale Air Interface Reciprocity (Motion) Survey of a PHS Network. Proc. of 34th Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, Ca, USA, 29 Oct. – 1 Nov., 2000. Pp. 297–300.

Kermoal, J. P., Schumacher, L., Pedersen, K I., Mogensen, P. E. & Frederiksen, F. 2002. A Stochastic MIMO Radio Channel Model with Experimental Validation. IEEE J. Sel. Areas in Commun,Vol. 20, No. 6, pp. 1211–1226.

Kotelba, A. 2004. Capacity of mobile wireless radio channels under different adaptive transmission schemes. Submitted to IEEE International Conference on Communications. Paris, June 2004.

Latvala, J., Syrjärinne, J., Niemi, S. & Niittylahti, J. 1999. Patient Tracking in a Hospital Environment using Wireless Stations and Extended Kalman Filtering. 1999 Middle East Conference on Networking, Beirut, Libanon. 5 p.

Lee, W. C. Y. 1990. Estimate of Channel Capacity in Rayleigh Fading Environment. IEEE Trans. Veh. Technol., Vol. 39, No. 3, pp.187–189.

Liu, X. (J.). 2001. Parallel Simulation of Large-Scale Wireless Ad Hoc Networks. Research proposal for doctoral dissertation. Darthmouth College. http://citeseer.nj.nec.com/liu01parallel.html

Low, Y.-H., Lim, C.-C., Cai, W., Huang, S.-Y., Hsu, W.-J., Jain, S. & Turner, S. J. 1999. Survey of Languages and Runtime Libraries for Parallel Discrete Event Simulation. Simulation, 72(3). http://citeseer.nj.nec.com/low99survey.html

Marzetta, T. L. & Hochwald, B. M. 1999. Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading. IEEE Trans. Inf. Theory, Vol. 45, No. 1, pp. 139–157.

Matinmikko, M. 2002. Measures of Power and Spectral Efficiency of Cellular Mobile Radio Systems. AURA Technical Report (unpublished). Oulu, Finland: VTT Electronics. 52 p.

Molkdar, D. 1991. Review on Radio Propagation into and within Buildings. IEE Proc. H, Vol. 138, No. 1, pp. 61–73.

Mämmelä, A. & Järvensivu, P. 2002. Cellular Communications Channels. In: Proakis, J. G. (ed.) The Wiley Encyclopedia of Telecommunications and Signal Processing. New York, John Wiley & Sons.

Nakano, K., Olariu, S. & Zomaya, A. Y. 2001. Energy-Efficient Permutation Routing in Radio Networks. IEEE Transactions on Parallel and Distributed Systems, Vol. 12, No 6. pp. 544–557.

Parsons, J. D. 2000. The Mobile Radio Propagation Channel. 2nd ed. Chichester, John Wiley & Sons. 418 p.

Perkins, C. E. 2001. Ad Hoc Networking. Reading, Massachusetts: Addison-Wesley. 370 p.

Perkins, C. 2002 IP Mobility Support for IPv4. RFC 3344.

Porcarelli, S., Di Giandomenico, F., Bondavalli, A., Barbera, M. & Mura, I. 2003. Service-level availability estimation of GPRS. IEEE Transactions on Mobile Computing, Vol. 2, No. 3, pp. 233–247.

Pottie, G. J. 1995. System Design Choices in Personal Communications. IEEE Personal Commun., Vol. 2, No. 5, pp. 50–67.

Proakis, J. G. 1995. Digital Communications, McGraw-Hill, 3rd ed.

Qiu, X. & Chawla, K. 1999. On the Performance of Adaptive Modulation in Cellular Systems. IEEE Trans. Commun., Vol. 47, No. 6, pp. 884–895.

Rappaport, T. S. 1996. Wireless Communications: Principles and Practice. Upper Saddle River, New Jersey, Prentice Hall. 641 p.

Recommendation ITU-R M.1225. 1997. Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000. 60 p.

Riley, G. F., Fujimoto, R. M. & Ammar, M. H. 1999. A Generic Framework for Parallelization of Network Simulations. 7th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems.

Rosen, E., Viswanathan, A. & Callon, R. 2001. Multi-Protocol Label Switching Architecture. RFC 3031.

Safwat, A., Hassanein, H. & Mouftah, H. 2002. Energy-Aware Routing in Wireless Mobile Ad Hoc Networks. In: Proceedings of the International Conference on Wireless Networks, June 2002, Las Vegas, Nevada, USA.

Saleh, A. A. M. & Valenzuela, R. A. 1987. A Statistical Model for Indoor Multipath Propagation. IEEE J. Sel. Areas in Commun., Vol. 5, No. 2, pp. 128–137.

Saunders, S. R. 1999. Antennas and Propagation for Wireless Communication Systems. Chichester, John Wiley & Sons. 409 p.

Scott, K. & Bambos, N. 1996. Routing and Channel Assignment for Low Power Transmission in PCS. In: Proceedings of ICUP'96, Vol. 2, October 1996. Pp. 498–502.

Shah, R. C. & Rabaey, J. M. 2002. Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. In: Proceedings of IEEE WCNC, March 2002, Orlando Florida, USA.

Shannon, C. E. 1958. Channels with side information at the transmitter. IBM Journal on Research and Development, Vol. 2, October 1958, pp. 289–293.

Simon, M., Hinedi, S. & Lindsey, W. 1995. Digital Communication Techniques, Signal Design and Detection, Prentice Hall: New Jersey.

Snir, M., Otto, S., Huss-Lederman, S., Walker, D. & Dongarra, J. 1996. MPI: The Complete Reference. Cambridge, Massachusetts, USA: The MIT Press.

Sukuvaara, T. 2002. Energy Efficient Routing Protocols in Ad-hoc Mobile Networking: State of the Art Review. AURA Technical Report (unpublished). Oulu, Finland: VTT Electronics. 29 p.

Telatar, E. 1999. Capacity of multi-antenna Gaussian channels. European Transactions on Telecommunications, Vol. 10, No. 6, pp. 585–595.

Thomson, S. & Narten, T. 1998. IPv6 Stateless Address Autoconfiguration. RFC 2462.

Thomson, S. & Narten, T. 2001. Privacy Extensions for Stateless Address Autoconfiguration in IPv6. RFC 3041.

Toh, C.-K. 2001. Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad Hoc Networks. IEEE Communications Magazine, Vol. 39, No 6, pp. 138–147.

Toh, C.-K. 2002. Ad Hoc Mobile Wireless Networks: Protocols and Systems. Upper Saddle River, New Jersey: Prentice Hall. 302 p.

Toh, C.-K., Cobb, H. & Scott, D. A. 2001. Performance Evaluation of Batter-Life-Aware Routing Schemes for Wireless Ad Hoc Networks. In: Proceedings of IEEE ICC 2001, June 2001, Helsinki, Finland.

UMTS 30.03 V.3.2.0. 1998. Universal Mobile Telecommunications System (UMTS); Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS. TR 101 112 V3.2.0 (1998-04). 84 p.

Waldorf, J. & Bagrodia, R. 1994. MOOSE: A Concurrent Object-Oriented Language for Simulation. International Journal of Computer Simulation, Vol. 4, No. 2.

Wang, Z. & Giannakis, G. B. 2002. Outage mutual information of space-time MIMO channels. In: Proceedings of the 40th Allerton Conference, October. Pp. 885–894.

Winters, J. H., Salz, J. & Gitlin, R. D. 1994. The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems. IEEE Trans. Commun., 42, 2/3/4, pp. 1740–1751.

Wroclawski J. 1997. The Use of RSVP with IETF Integrated Services. RFC 2210.

Wu, H., Fujimoto, R. M. & Riley, G. 2001. Experiences parallelizing a commercial network simulator. Proceedings of the 2001 Winter Simulation Conference. http://citeseer.nj.nec.com/473173.html

Wu, J., Gao, M. & Stojmenovic, I. 2001. On Calculating Power-Aware Connected Dominating Sets for Efficient Routing in Ad Hoc Wireless Networks. In: Proceedings of International Conference on Parallel Processing, 2001. Pp. 346–354.

Wu, S.-L., Tseng, Y.-C. & Sheu, J. P. 2000. Intelligent Medium Access for Mobile Ad Hoc Networks with Busy Tones and Power Control. IEEE Journal on Selected Areas in Communications, Vol. 18, No 9, pp. 1647–1657.

Wyner, A. D. 1981. Fundamental Limits in Information Theory. Proc. of the IEEE, Vol. 69, No. 2, pp. 239–251.

Zorzi, M. & Rao, R. R. 1997. Energy-Constrained Error Control for Wireless Channels. IEEE Personal Commun., Vol. 4, No.6, pp. 27–33.

Published by

Series title, number and report code of publication

VTT Research Notes 2248
VTT–TIED–2248

Author(s)
Frantti, Tapio, Jurvansuu, Marko & Mämmelä, Aarne

Title
Efficient wireless networking with advanced services and negotiated QoS

Abstract
In the AuRa project an efficient wireless networking scheme for mobile terminals with advanced services and negotiated quality of service (QoS) level was researched and developed. The project consisted of various topics in the various communication layers of the OSI model. In the physical and data link layers area spectral efficiency was researched It was observed that log-normal shadowing is very crucial for the spectral efficiency. In outdoor positioning, advantages of wideband channel delay profiles were utilised so that database correlation method (DCM) algorithm can locate the terminal using only one base station hereby avoiding the problem common to observed time difference of arrival algorithm (OTDOA). For the indoor positioning a configuration tool was developed to make signal chart measurement faster with more dense and accurate grid. Hence, the typical accuracy of positioning in indoor of 2 m was achieved. In vertical handover topics signalling schemes as well as key parameters mapping and classification for the related QoS issues were developed in order to make vertical handover possible between the WLAN and UTMS networks. In the QoS research an end-to-end QoS measurement tool was also developed. From the measurements significant variations in end-to-end latency time of General Packet Radio Systems (GPRS) were observed between different operators and terminal types.

Network layer topics included power efficiency research on ad hoc and sensor networks taking into consideration mobility of nodes and the number of sequential hops in signalling and data delivery. Furthermore, methods and tools for parallel networking simulation were researched and the graphical user interface tool for the NS2 simulation tool was built with the aim of hiding the complexity of NS2 tool's scripts and distribution of simulation runs.

Keywords
Quality of Service (QoS), wireless network routing, positioning technologies, spectral efficiency, cellular systems

Activity unit
VTT Electronics, Kaitoväylä 1, P.O.Box 1100, FIN–90571 OULU, Finland

ISBN
951–38–6473–1 (URL: http://www.vtt.fi/inf/pdf/)

Date	Language	Pages
June 2004	English	110 p.

Name of project	Commissioned by
MIPO.MMC MIPO_CES	

Series title and ISSN	Sold by
VTT Tiedotteita – Research Notes 1455–0865 (URL: http://www.vtt.fi/inf/pdf/)	VTT Information Service P.O.Box 2000, FIN–02044 VTT, Finland Phone internat. +358 9 456 4404 Fax +358 9 456 4374

VTT TIEDOTTEITA – RESEARCH NOTES

VTT ELEKTRONIIKKA – VTT ELEKTRONIK – VTT ELECTRONICS

1933 Ihme, Tuomas, Kumara, Pekka, Suihkonen, Keijo, Holsti, Niklas & Paakko, Matti. Developing application frameworks for mission-critical software. Using space applications as an example. 1998. 92 p. + app. 20 p.

1965 Niemelä, Eila. Elektroniikkatuotannon joustavan ohjauksen tietotekninen infrastruktuuri. 1999. 42 s.

1985 Rauhala, Tapani. Javan luokkakirjasto testitapauseditorin toteutuksessa. 1999. 68 s.

2042 Kääriäinen, Jukka, Savolainen, Pekka, Taramaa, Jorma & Leppälä, Kari. Product Data Management (PDM). Design, exchange and integration viewpoints. 2000. 104 p.

2046 Savikko, Vesa-Pekka. EPOC-sovellusten rakentaminen. 2000. 56 s. + liitt. 36 s.

2065 Sihvonen, Markus. A user side framework for Composite Capability / Preference Profile negotiation. 2000. 54 p. + app. 4 p.

2088 Korva, Jari. Adaptiivisten verkkopalvelujen käyttöliittymät. 2001. 71 s. + liitt. 4 s.

2092 Kärki, Matti. Testing of object-oriented software. Utilisation of the UML in testing. 2001. 69 p. + app. 6 p.

2095 Seppänen, Veikko, Helander, Nina, Niemelä, Eila & Komi-Sirviö, Seija. Towards original software component manufacturing. 2001. 105 p.

2114 Sachinopoulou, Anna. Multidimensional Visualization. 2001. 37 p.

2129 Aihkisalo, Tommi. Remote maintenance and development of home automation applications. 2002. 85 p.

2130 Tikkanen, Aki. Jatkuva-aikaisten multimediasovellusten kehitysalusta. 2002. 55 s.

2157 Pääkkönen, Pekka. Kodin verkotettujen laitteiden palveluiden hyödyntäminen. 2002. 69 s.

2160 Hentinen, Markku, Hynnä, Pertti, Lahti, Tapio, Nevala, Kalervo, Vähänikkilä, Aki & Järviluoma, Markku. Värähtelyn ja melun vaimennuskeinot kulkuvälineissä ja liikkuvissa työkoneissa. Laskenta-periaatteita ja käyttöesimerkkejä. 2002. 118 s. + liitt. 164 s.

2162 Hongisto, Mika. Mobile data sharing and high availability. 2002. 102 p.

2201 Ailisto, Heikki, Kotila, Aija & Strömmer, Esko. Ubicom applications and technologies. 2003. 54 p.

2213 Lenkkeri, Jaakko, Marjamaa, Tero, Jaakola, Tuomo, Karppinen, Mikko & Kololuoma, Terho. Tulevaisuuden elektroniikan pakkaus- ja komponentti-tekniikat. 2003. 78 s. + liitt. 4 s.

2238 Kallio, Päivi, Niemelä, Eila & Latvakoski, Juhani. UbiSoft - pervasive software. 2004. 68 p.

2248 Efficient wireless networking with advanced services and negotiated QoS. Ed. by Frantti, Tapio, Jurvansuu, Marko & Mämmelä, Aarne. 2004. 110 p.

Tätä julkaisua myy	Denna publikation säljs av	This publication is available from
VTT TIETOPALVELU PL 2000 02044 VTT Puh. (09) 456 4404 Faksi (09) 456 4374	VTT INFORMATIONSTJÄNST PB 2000 02044 VTT Tel. (09) 456 4404 Fax (09) 456 4374	VTT INFORMATION SERVICE P.O.Box 2000 FIN–02044 VTT, Finland Phone internat. + 358 9 456 4404 Fax + 358 9 456 4374

ISBN 951–38–6473–1 (URL: http://www.vtt.fi/inf/pdf/)
ISSN 1455–0865 (URL: http://www.vtt.fi/inf/pdf/)