

Study and Performance Evaluation of Multimedia Broadcast / Multicast Service in GERAN

Michal Hronec

April 2004

University of Joensuu
Computer Science
Pro gradu

ABSTRACT

University of Joensuu

RESEARCH RELEASE

This thesis describes Multimedia Broadcast/Multicast Service (MBMS). MBMS is a bearer service in a digital cellular radio communication network that supports resource-efficient point-to-multipoint transmission of multimedia data. Services provided by a mobile network operator or some external service provider can use MBMS's capabilities to deliver identical data to multiple users simultaneously. It resembles IP multicast and it shall be compatible with it, though there are several important differences demanded by the cellular network environment. The differences and requirements for introduction of MBMS into GSM/EDGE radio access network are examined. A network level simulator was extended with abilities to simulate point-to-multipoint connections to estimate MBMS's performance and impact on other services in the simulated network. The simulation results show that MBMS uses radio resources efficiently and that the impact on other services is minimal in comparison with alternative point-to-point delivery of the same data. The performance regarding connection throughput is close to the maximum throughput at the given modulation and coding scheme for 99% of the receiving users. However, high (99% percentile is around 20%) RLC SDU error ratio shows that existing mechanisms alone are not enough to fulfil the QoS requirements for streaming and background type of traffic, which should be delivered by MBMS. Among the possible solutions to lower the RLC SDU error ratio are simple repetition redundancy, outer coding using Reed-Solomon codes and uplink feedback.

Author: Michal Hronec

Student number: 149116

Title: Study and Performance Evaluation of Multimedia Broadcast / Multicast Service in GERAN

Faculty / Department: Faculty of Science / Computer Science

Nr. of pages: 59

Time: 18 months

Type of work: pro gradu

Key words: MBMS, multicast, point-to-multipoint, performance evaluation, GERAN, cellular network simulation

Notes: _____

ACKNOWLEDGEMENTS

I would like to thank my supervisor from the University of Joensuu Prof. Jussi Parkkinen and my supervisor from the Nokia Research Center Guillaume Sébire for providing comments to the draft paper, Eero Nikula, Shkumbin Hamiti, Martti Moisio, Johanna Pekonen, Janne Parantainen, David Bohatý and Jaroslav Uher for their comments, advice and information on cellular radio communication networks, simulation systems and environment.

Special thanks go to Merja Hyttinen, Marja-Liisa Makkonen and Eeva Saukkonen for their help and advice with study-, work- and life-related paper handling.

Further I would like to thank my sister, my girlfriend and my parents for their moral support.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
Abbreviations.....	vi
1. Introduction.....	1
2. Cellular Multicast Systems	2
2.1 Overview of Cellular Systems.....	2
2.1.1 GSM.....	3
2.1.2 GPRS.....	6
2.1.3 UMTS.....	11
2.2 Overview of IP Multicasting	14
2.2.1 Multicast Groups.....	14
2.2.2 IP Multicast Routing	15
2.3 Multimedia Broadcast/Multicast Service in Cellular Systems	16
2.3.1 Differences Between MBMS and IP Multicast.....	16
2.3.2 Requirements	17
2.3.3 Broadcast Mode	17
2.3.4 Multicast Mode	18
2.3.5 Architecture.....	19
2.3.6 MBMS Context And Delivery Path.....	21
2.3.7 Data Transfer in the Core Network.....	21
3. Multimedia Broadcast/Multicast Service in GERAN	23
3.1 Requirements	23
3.1.1 Joining Requirements.....	23
3.1.2 Notification Requirements	24
3.2 Functions	24
3.3 MBMS Context.....	25
3.4 Notification.....	25
3.4.1 Notification on (P)BCCH.....	26
3.4.2 Notification on (P)CCCH / (P)NCH	26
3.4.3 Notification on (P)PCH.....	26
3.4.4 Counting.....	27
3.5 Data Transfer	27
3.5.1 Reed-Solomon Codes.....	28
3.5.2 Feedback in Uplink	29
4. Evaluation of Multimedia Broadcast/Multicast Service in GERAN.....	30
4.1 Description of the Simulator.....	30
4.2 Simulation Model Overview.....	31
4.2.1 Point-to-multipoint Data Transmission Model	32
4.3 Simulation Assumptions	34
4.3.1 Simulation Scenarios.....	34
4.4 Analysis of the Results	38
4.4.1 Conclusions	43

5. Discussions.....	44
References.....	45
6. Appendix 1 – Simulation Results	49
6.1 Connection Throughput.....	49
6.2 RLC SDU Error Ratio	55
6.3 RLC SDU Error Ratio Map For PTM Scenario.....	58

ABBREVIATIONS

The following list contains all abbreviations and acronyms used in this thesis. Abbreviations that contain version numbers, like IPv6 for Internet Protocol Version 6, are not listed.

2G	2 nd generation
3G	3 rd generation
3GPP	3 rd Generation Partnership Project
8-PSK	Octagonal Phase Shift Keying
AMR	Adaptive Multi-Rate
ARQ	Automatic Repeat Request
AuC	Authentication Centre
BG	Border Gateway
BM-SC	Broadcast/Multicast Service Centre
BS	Base Station
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CBCH	Cell Broadcast Channel
CBS	Cell Broadcast Service
CN	Core Network
CS	circuit-switched
DRX	Discontinuous Reception
ECSD	Enhanced Circuit-Switched Data
EDGE	Enhanced Data Rates for GSM Evolution
EGPRS	Enhanced General Packet Radio Service
EIR	Equipment Identification Register
FDMA	Frequency Domain Multiple Access
FTP	File Transfer Protocol
GERAN	GSM/EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GMSC	Gateway Mobile Services Switching Centre
GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service
GRA	GERAN Registration Area
GSM	Global System for Mobile Communications
GTP	GPRS Tunnelling Protocol
HLR	Home Location Register
HSCSD	High-Speed Circuit-Switched Data
IGMP	Internet Group Management Protocol
IMSI	International Mobile Subscriber Identification
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LAN	Local Area Network
LLC	Logical Link Control
MAC	Medium Access Control
MBMS	Multimedia Broadcast and Multicast Service
MCS	Modulation and Coding Scheme
ME	Mobile Equipment
MLD	Multicast Listener Discovery

MM	Mobility Management
MS	Mobile Station
MSC	Mobile Services Switching Centre
NAK	Negative Acknowledgement
NMS	Network Management Subsystem
NSS	Network Switching Subsystem
PACCH	Packet Associated Control Channel
PAGCH	Packet Access Grant Channel
PBCCH	Packet Broadcast Control Channel
PCCCH	Packet Common Control Channel
PDCP	Packet Data Convergence Protocol
PDN	Public Data Network
PDP	Packet Data Protocol
PDTCH	Packet Data Traffic Channel
PLMN	Public Land Mobile Network
PNCH	Packet Notification Channel
PPCH	Packet Paging Channel
PS	packet-switched
PSI	Packet System Information
PSTN	Public Switched Telephone Network
p-t-m	point-to-multipoint
p-t-p	point-to-point
QoS	Quality of Service
R99	Release 1999
RAB	Radio Access Bearer
RAN	Radio Access Network
RLC	Radio Link Control
RNC	Radio Network Controller
RNS	Radio Network Subsystem
RP	Rendezvous Point
RS	Reed-Solomon
SAIC	Single Antenna Interference Cancellation
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SIM	Subscriber Identity Module
SMS	Short Message Service
TBF	Temporary Block Flow
TDMA	Time Domain Multiple Access
TMGI	Temporary Multicast Group Identification
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visitor Location Register
VPLMN	Visited Public Land Mobile Network
WCDMA	Wideband Code Division Multiple Access

1. INTRODUCTION

Mobile communications have become reality of our everyday lives. For many people and businesses, it is perceived as essential and a matter-of-course. The most limiting factor of radio telecommunication systems is the bandwidth of the radio resource – the frequency bands. Mobile network operators may pay huge amounts of money for licences giving them a temporary right to use certain frequency bands for telecommunication. It is therefore essential that this scarce resource be used as efficiently as possible. This means to serve as many users as possible with acceptable quality of service. In other words, the capacity of a mobile network must be as high as possible without sacrificing on user satisfaction. Significant milestones have been reached that allows for boosting mobile networks' capacity e.g. in GSM:

- half-rate speech enables to double the number of simultaneous speech connections by occupying only half of the traffic channel,
- adaptive multi-rate (*AMR*) enabling optimised transfer of speech data by adapting the channel coding and codec mode to current radio conditions,
- general packet radio service (*GPRS*) enables efficient usage of radio resources during the transfer of packet data,
- enhanced data rates for global evolution (*EDGE*) brings three times higher transfer rate per one timeslot by coding three bits in one modulated symbol.

New solutions are also being studied and introduced e.g. single antenna interference cancellation (*SAIC*) – a technique to suppress channel interference using digital signal processing on receivers using only one antenna [3GP03d] – and Multimedia Broadcast/Multicast Service (*MBMS*).

MBMS is a new bearer service within cellular radio communication system aimed at efficient use of radio resources through delivery of the same data, ideally on a single flow, to several users. The specification of *MBMS* is under progress within the *3rd Generation Partnership Project (3GPP)* [3GP03a] standardisation body.

This thesis examines the functions and requirements of *MBMS* and evaluates its performance in *GSM/EDGE radio access network (GERAN)*. It is divided into several chapters: the second chapter contains an overview of cellular systems, IP multicast and an overall description of *MBMS*. The third chapter describes the integration of *MBMS* within *GERAN*, its impacts on radio protocols and needs for new procedures. The fourth chapter contains the performance evaluation of *MBMS* in *GERAN*. It includes general information about the simulation tool used, the simulation model, assumptions and scenarios as well as analysis of the simulation results.

2. CELLULAR MULTICAST SYSTEMS

This chapter contains an overview of cellular systems targeted for provision of MBMS. First an overview of cellular systems is given. Then follows the description of its most successful example: GSM and its enhancement GPRS. After that, a short outline of UMTS is presented. It is followed by basic principles of IP multicast, which inspired the creation of MBMS. Finally, the overview of MBMS itself closes this chapter.

2.1 Overview of Cellular Systems

A cellular mobile communications system consists of a large number of low-power transmitter/receivers called *base transceiver stations (BTS)* or *base stations (BS)*. The area covered by the cellular system is divided into cells (see Figure 2.1), which are the basic geographic service areas. Each cell is served by one base station. A BTS can serve only a certain amount of mobile phones further denoted as *mobile stations (MS)*. Thus to serve subscribers in densely populated areas, such as cities, the cell sizes must be appropriately smaller. Conversely, in rural areas the cells can be larger up to a limit imposed by the underlying technology. For example in GSM, the cells can have a maximum radius of 35 km [Mou92].

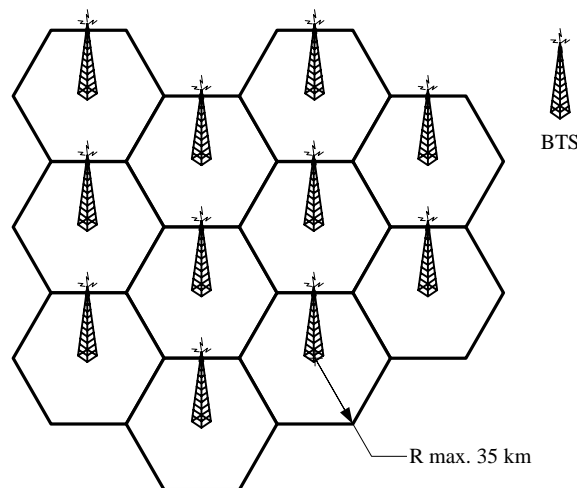


Figure 2.1: Model of a cellular communications system

When the user is making a call through his mobile station, the BTS reserves a temporary bi-directional channel for communication with this particular MS. The direction from MS to BTS is called *uplink*; the other direction is *downlink*. For such bi-directional channel, two radio frequencies are used in order to avoid interference between transmitter and receiver part of the transceiver, which is the essential part of every BTS and mobile station. Because the number of available channels in a cell and in the whole cellular system is limited, the channels are released after the call ends and can be reused by another MS.

Unfortunately, not all frequencies reserved for a cellular system can be used in all cells. Using the same frequency for a channel in two adjacent cells at the same time results in interference that severely disturbs communication in both channels making the communication impossible. Interference

can be reduced if the two interfering transmitters are far away enough because the signal strength decreases rapidly with increasing distance. Thus, the frequencies can be reused in non-adjacent cells.

Another way to limit the level of interference in the cell is to control the power output of BTS and MS so that the received signal is strong enough to ensure reliable communication but no stronger to limit the interference. To achieve this goal, the mobile and the base station regularly measure the received signal strength and inform each other. The BTS then instructs the mobile phone via signalling messages to decrease or increase its output power according to these measurements. This process is called *power control*.

If a mobile station moves between cells during a call, it is automatically handed over from cell to cell so that the call is not interrupted or at least not dropped. This process, called *handover*, is controlled by the network. It requires the mobile station to regularly measure and report the signal level of neighbouring cells. Based on these measurements the network selects an appropriate neighbouring cell so that the disturbance of the call caused by the handover is minimal.

2.1.1 GSM

The *Global system for mobile communications (GSM)* is a digital cellular mobile communication system. With analogue predecessors of GSM being the first generation of personal radio communication systems, the GSM is the most successful and best known representative of the second generation (2G).

The interfaces between various components of the GSM have been standardised and published in order to enable the competition between hardware vendors and to ensure the interoperability between equipment from different vendors. For the subscribers this means that they can travel in different countries while still being able to use their mobile phone provided the frequency bands are adequate. This is called *roaming* and its availability in particular country depends only on the international treaties of operators, ideally not on the hardware [Mou92].

2.1.1.1 Architecture

The general structure of the GSM (circuit-switched) is in Figure 2.2.

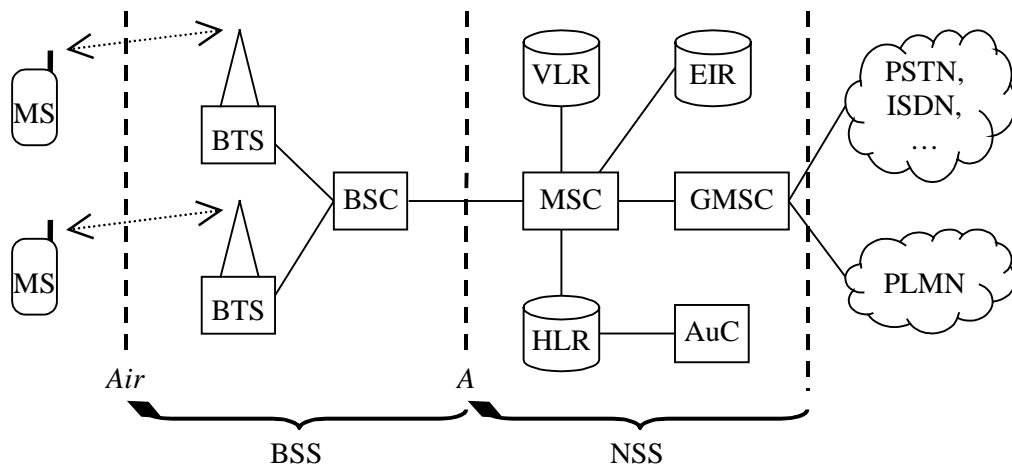


Figure 2.2: GSM structure

The GSM consists of several subsystems. The users are generally aware of only one part of the whole system – the mobile station. A mobile station has two parts: *mobile equipment (ME)* and *subscriber identity module (SIM)*, which holds information about the subscriber, his phone book, secret ciphering key and other things depending on its capabilities and memory size.

The subsystem responsible for dealing with radio aspects of GSM is the *base station subsystem (BSS)* which contains a *base station controller (BSC)* and one or more BTSs. The radio transmitter and receiver are parts of the base transceiver station standing on the other side of the radio link with MS. This radio interface is more closely described in the next subsection. The radio coverage of the antenna of BTS defines the cell area. Several BTSs are connected to the BSC. It is the central component of the BSS. It takes care of radio resource management (e.g. radio channel allocation, handover control), mobility management (e.g. paging, location registration and updates) and traffic connection establishment between MS and *network switching subsystem (NSS)* commonly referred to as *core network (CN)*.

The BSC is connected with the network switching subsystem through the A interface to the *mobile services switching centre (MSC)* in circuit-switched domain. There are several BSCs connected to one MSC. The area controlled by these BSCs is the *service area* of the MSC. The area covered by an MSC is a *location area*. The MSC is a large switch. It co-ordinates the setting-up of calls and their routing. The NSS has interfaces to connect to external networks such as *public switched telephone network¹ (PSTN)*, *integrated services digital network (ISDN)* and also to other *public land mobile networks (PLMN)* and transport networks. Because connections with external networks often require some data adaptation, there is a special node in the NSS for this task: the *gateway mobile services switching centre (GMSC)*.

In the NSS there are also databases containing information about subscribers and their equipment:

- *Home location register (HLR)* – stores information about subscribers such as services provided to them, their number and current approximate location. Part of HLR is the *authentication centre (AuC)* containing security data for subscriber authentication.
- *Visitor location register (VLR)* – holds information about all subscribers currently situated in the service area of the MSC. It has more accurate information on their location than HLR and it has also information about visitors from other mobile networks. This enables roaming and ensures correct charging.
- *Equipment identification register (EIR)* – holds information about mobile stations, which are allowed to connect to the network. This way if a mobile is stolen it can be denied access thus making it useless.

The last part of GSM is the *network management subsystem (NMS)*. It is connected to all other components of the network yet for them it is invisible. It serves the operator as a powerful tool to configure, monitor and control the whole network. It collects numerous statistics of the network load and usage that can be used for its further optimisation and expansion.

2.1.1.2 Radio Interface

The BSS implements the radio interface, which is the most important part of GSM because the spectral efficiency of the cellular system depends completely on this interface. The full specification of this

¹ Ordinary analog circuit-switched telephone network

interface and its worldwide standardisation also makes it possible to achieve one of the main goals of GSM – international roaming [Mou92].

Today, GSM operates in 850 MHz, 900 MHz, 1800 MHz and 1900 MHz frequency bands and uses a combination of *frequency domain multiple access (FDMA)* and *time domain multiple access (TDMA)* (see Figure 2.3). In 900 MHz frequency band there are 124 frequency channels, each 200 kHz wide. The traffic is sent in bursts of symbols (in GSM, which employs the GMSK¹ modulation scheme, one symbol equals one bit). In time domain each burst fits into one time slot, with the duration of 0.577 ms (exactly 15/26 ms). Eight time slots form a *TDMA frame*. The sequence of time slots in consecutive frames forms a *basic physical channel*. As there are 8 time slots per TDMA frame, there exist 8 basic physical channels per one frequency channel, which are used to transfer user data and signalling information.

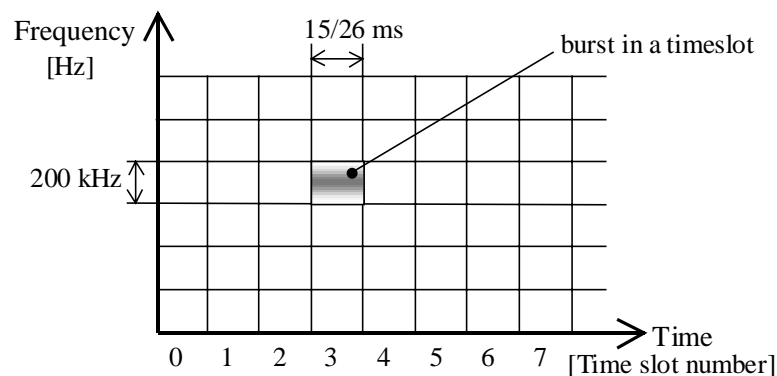


Figure 2.3: Multiple access scheme in GSM

A collection of 26 successive TDMA frames forms a *26-multiframe*, which contains traffic channels for transferring voice and user data. Another cyclic structure, that contains 51 TDMA frames in a row, is known as the *51-multiframe* bearing channels used to carry signalling information.

Before transmission speech is digitised and compressed using a speech codec. The resulting data bits are encoded before they are transmitted in the bursts. This encoding is used to tackle the degradation due to the propagation in a radio channel. It offers robust error correction after the data is received. If after the error correction, some errors are still detected, either retransmission is carried out, or the data is discarded. During the encoding there are some redundant bits added to the data, which effectively lowers the data rate.

2.1.1.3 Services

GSM provides several services to the subscribers [Mou92]:

- Telephony – speech is digitally encoded in the mobile station and transmitted through the GSM network as a digital stream. When the callee is using an ordinary phone, the digital stream is converted to analog signal in GMSC (see Figure 2.2). Conversely, if someone from the PSTN is calling a GSM subscriber, the analog signal is converted to digital stream in the GMSC. The GSM network can also communicate with ISDN.
- Fax – subscriber has two numbers: one for speech and the other for fax.

¹ Gaussian Minimum Shift Keying

- Data transfer – initially there were only very low bit rates supported: 2.4 kbit/s, 4.8 kbit/s and 9.6 kbit/s. Later, with the introduction of *high-speed circuit-switched data (HSCSD)*, higher bit rates were possible, e.g. 14.4 kbit/s and multiples. This service enables connections with packet switched public data networks such as Internet. The mobile station can be connected to a computer and acts like a modem.
- SMS¹ – subscribers can send short text messages to each other. The length of a message is limited to 160 characters. The subscriber may request a confirmation of delivery of the message. SMS can be sent or received even during a phone call. In addition to point-to-point messages, the network is capable of broadcasting short text messages sent by the operator to certain (operator defined) areas. This is known as the *cell broadcast service (CBS)*. These messages can be used for example for road traffic information or news updates.

The actual services, that are available to a subscriber, depend on the contents of his subscription, on the capabilities of the particular network he is using (owned by the operator) and on the capabilities of his mobile station. All the services use circuit-switched (*CS*) connections (except SMS), thus the subscriber pays for the connection time no matter how much data has been transferred, or whether the data has been correctly received or not.

The subscribers must authenticate to the network before using any of its services. The MS does the authentication transparently to the user. The speech and data service is ciphered on the radio link, the secret key is stored in the SIM. Ciphering is intended to avoid any unauthorised acquisition of data by a third party.

2.1.2 GPRS

General packet radio service or in short *GPRS* is a *packet-switched (PS)* data network addition to GSM. It offers robust access to external public data networks (*PDN*) such as Internet. In GPRS there are no circuits established for data transfers. Radio resources are reserved only for the duration of the data transfer and released thereafter, hence avoiding the “waste” inherent to a circuit-switched connection. That a GPRS user may remain connected to the packet-switched core network even without anything to download/transfer is commonly referred to as “always-on”. A user is charged for the amount of data transferred rather than the duration of the call. For the operator GPRS offers efficient resource usage, because the resources are allocated for each packet separately and after its transmission they are freed and can be reused by other user. It is also possible to allocate different capacity for uplink and downlink. [3GP02h, Hal02]

There are three classes of modes of operation a GPRS capable MS can support [3GP02a]:

- class A: MS simultaneously uses CS and PS services,
- class B: MS uses both CS and PS services but not at the same time,
- class C: MS uses either CS or PS services or both but the selection must be done manually.

2.1.2.1 Architecture

GPRS integration into GSM network requires upgrades in BSS [Hei00]. A *packet control unit* has to be added into BSS so that the BSS is able to handle the data packets on the radio interface properly. Aside from this the base station subsystem is shared unchanged with GSM and also some components

¹ Short Message Service

and interfaces in the network switching subsystem. New components, that implement the packet-switched part of NSS are *serving GPRS support node (SGSN)* and *gateway GPRS support node (GGSN)*. Their integration into the GSM architecture is depicted in Figure 2.4.

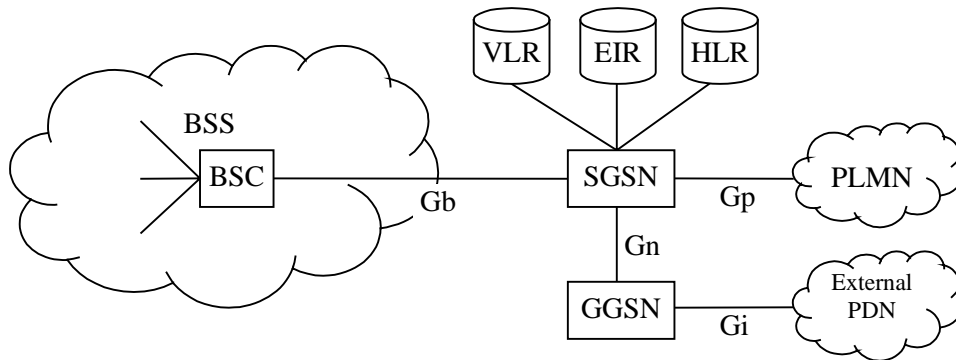


Figure 2.4: Integration of GPRS into GSM

The SGSN is a powerful packet switch that connects to one or more base station controllers. It has a similar role as the MSC in circuit-switched domain; its service area¹ is called *routing area*. It co-operates with VLR and HLR to keep track of mobile stations in its routing area and handles their registration and authentication.

SGSN also ciphers the transmitted data to MS (unlike in GSM, where the BTS takes care of ciphering) to prevent unauthorised access to them. It opens and closes connections to external data networks via GGSN. Furthermore it collects charging information and data for statistics. It also stores *PDP² contexts* for each data connection. The PDP context contains information about the endpoint MS of the communication especially its assigned IP address and *quality of service (QoS)* parameters. When the user moves to a routing area served by another SGSN the PDP context “travels” with him to the new SGSN and is deleted in the old one during the routing area update procedure.

The tracking of mobile stations is needed in order to provide connection to the MS, which can be anywhere in the PLMN. Tracking is the main function of *mobility management (MM)* done in SGSN. From the mobility management point of view the MS can be in these states [Hal02]:

- Idle – the location of MS is unknown and thus it is not reachable by the network. MS is in this state for example when powered off or when it is not allowed to attach to GPRS because it is blacklisted in EIR.
- Ready – the location of MS is known with cell accuracy. The MS reaches this state by a procedure called *GPRS attach*, which is done automatically after power on and it is transparent to the user. When the MS wants to send or receive data it must activate a PDP context first. When the MS wants to disconnect from the network it performs a *GPRS detach* procedure, and changes state to idle. GPRS detach can also be initiated by the network.
- Standby – the MS is attached to the network but its location is known only with routing area accuracy. When it sends any data, it changes state to ready. It may remain in standby state and receive p-t-m data and paging messages. MS enters standby state from ready state after a timeout.

¹ One SGSN can serve several routing areas.

² Packet Data Protocol, generic name describing any packet protocol.

The GGSN is a router, sometimes integrated with firewall. It is the interfacing network element between the GPRS packet-switched core network (*PS CN*) and external data networks or foreign PLMN. Because it routes data packets to the MS it needs to know which MS is served by which SGSN. This information can be obtained from the HLR. Upon activating the data service by GPRS attach procedure the GGSN assigns an IP address to the mobile station and stores it in its PDP context, so that it can be addressed from the external networks.

The GPRS support nodes (SGSN and GGSN) are part of the IP-based backbone network known as PLMN [Vai02]. Above IP protocol is *GPRS tunnelling protocol (GTP)*, which encapsulates all the packets in order to deliver them to mobile stations. GTP is also used in inter-PLMN transport networks to deliver data to subscribers roaming in foreign PLMNs (also called *visited PLMN – VPLMN*).

2.1.2.2 Radio Interface

In the radio interface, three new coding schemes¹ have been introduced when GPRS was introduced (in Release 97) to enable higher data rates in adequate channel conditions at the cost of reduced data protection against transmission errors.

Table 2.1: Coding schemes in GPRS

Scheme	Data rate [kbit/s]
CS-1	9.05
CS-2	13.4
CS-3	15.6
CS-4	21.4

Each of them offers less data protection than the previous one as they are listed in the Table 2.1 and the data rate increases accordingly. The CS-4 coding scheme offers no data protection at all and reliable data delivery is accomplished by retransmissions only. Which coding scheme is used for particular data transmission is determined by the BSS according to current radio link conditions: this process is called *link adaptation*.

A new frame structure – the *52-multiframe* consisting of 52 TDMA frames – has been added and is used to group *packet data traffic channels (PDTCH)* – logical channels carrying packet data [3GP02i].

2.1.2.3 EDGE, GERAN

The evolution of GSM/GPRS reached the so-called *enhanced data rates for global evolution (EDGE)* in 3GPP Release 99. EDGE introduces an additional modulation to GMSK, known as *8-PSK*² modulation, which effectively triples the data throughput, because one symbol now carries three bits instead of one. EDGE constitutes of two parts: *ECSD (Enhanced Circuit-Switched Data)* in the CS domain, and *EGPRS (Enhanced GPRS)* in the PS domain.

Not only does EGPRS allow from higher throughput but it also defines link quality control as a combination of link adaptation and incremental redundancy. It enables flexible adaptation of coding scheme used to encode data into sequence of bits, that are transferred by radio waves, to current radio

¹ CS-1 was already in GSM.

² Octagonal Phase Shift Keying

conditions. EGPRS also offers incremental redundancy (called *Hybrid Type II ARQ*) that consists in sending additional redundancy upon block errors, hence increasing the probability of successful decoding after retransmissions through combining with prevailing (re)transmissions.

The combination of GSM and EDGE lead to the term *GSM/EDGE radio access network* or *GERAN*, to be compared with *UMTS terrestrial radio access network (UTRAN)*.

2.1.2.4 PS Domain Radio Interface Protocols in GERAN

The protocols used in GSM/EDGE are organised in layers and planes. There is user plane and control plane. Protocols in user plane are used to convey user data whereas protocols in control plane are used to transport signalling and control information. Each layer in the protocol stack has precisely defined function and interface with its upper and lower layer.

In the transmission chain the data goes through several protocol layers (see Figure 2.5). The data originates in the application layer (not shown) and descends through the stack of protocols down to physical layer (PHY), where it gets transmitted to the receiver. During this descent the data is handed from layer to layer in *data units*. Data unit entering a layer from layer above it is called *service data unit (SDU)*. SDU might get segmented into several smaller data units during the processing in the current layer and a header might be appended to it. The result of the processing is one or more *protocol data units (PDU)*, which are passed to lower layer or transmitted if they come from the lowest layer in the stack.

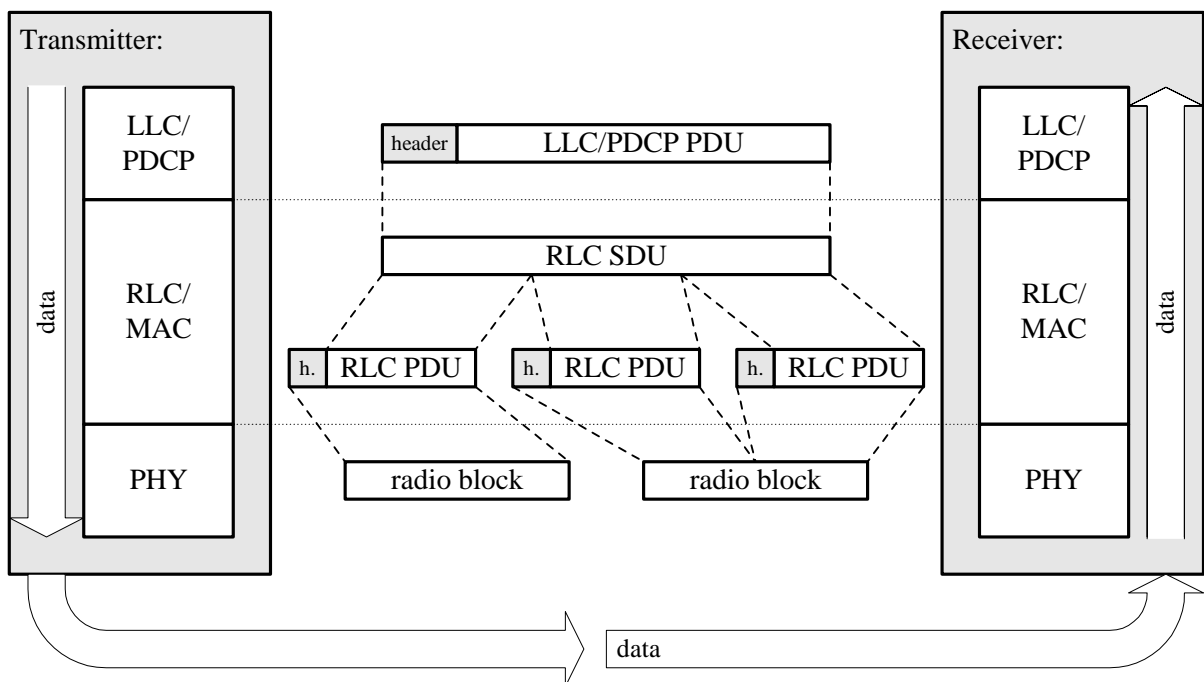


Figure 2.5: Protocol layers and data units

In the receiver the data goes through a reversed process up the protocol stack, the data units are reassembled to larger data units and finally delivered to the receiver's application layer.

The protocols used in the radio interface Um in GERAN are shown in Figure 2.6 (for Gb mode) and Figure 2.7 (for Iu mode). Only user plane protocols are described, which were involved in the simulations (see section 4.2). The mode, in which MS and GERAN work, depends on the interfaces used to connect GERAN to the core network. If GERAN is connected through A and Gb interface, Gb mode is used (in PS domain), otherwise it must¹ be connected via Iu interface and Iu mode is used (see also Figure 2.8). In further text the differences in mode are not taken into account because the differences in using GERAN via Gb or Iu interface are not considered on RAN level.

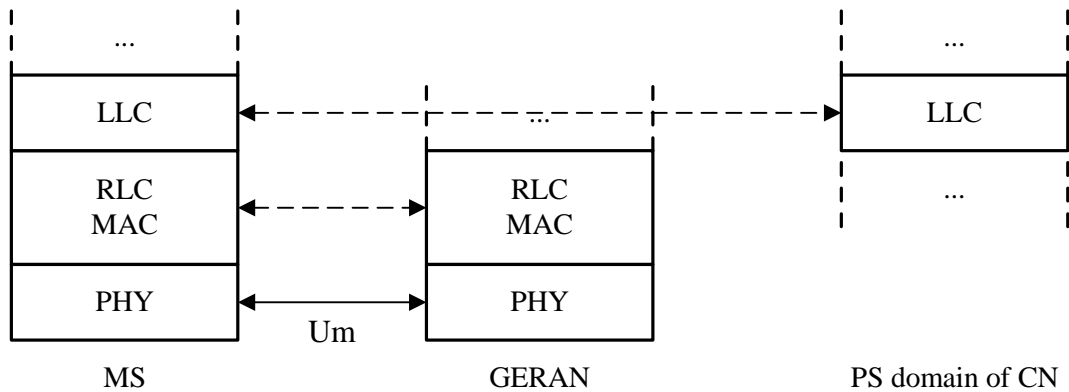


Figure 2.6: User plane radio interface protocols in Gb mode

The *logical link control (LLC)* protocol belongs to layer 3 (in Gb mode). It provides reliable ciphered logical link between MS and SGSN. This link is maintained even when MS changes cells, as long as the cells are served by the same SGSN [3GP02g].

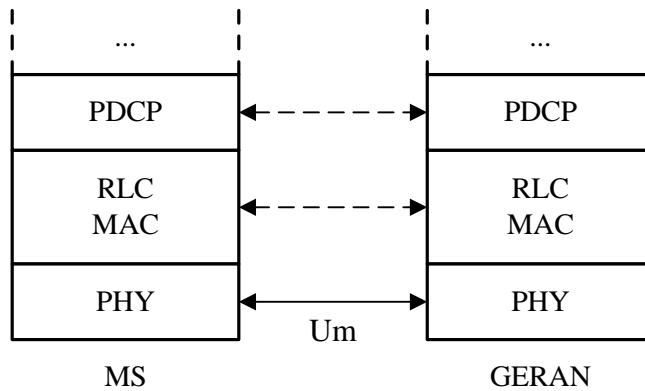


Figure 2.7: User plane radio interface protocols in Iu mode

The *packet data convergence protocol (PDCP)* is part of layer 3 in Iu mode. PDCP takes care of header compression and decompression of IP packets. It does not compress the payload of IP packets.

The *radio link control (RLC)* protocol is a sublayer of layer 2. RLC functionality depends on the mode it operates in:

¹ According to [3GP03e].

- Transparent – offers data transfer without adding any protocol overhead, the functionality of RLC is bypassed.
- Unacknowledged – on the transmitting side RLC offers segmentation of upper layer PDUs into RLC PDUs and their unreliable delivery. In the receiver, RLC concatenates the received RLC PDUs and reassembles the upper layer PDUs. RLC ciphers the payload of RLC PDUs.
- Acknowledged – offers the same functionality as in unacknowledged mode except the delivery of RLC PDUs is reliable. Transmission reliability is ensured by selective retransmission of erroneously received RLC PDUs. In this mode RLC employs link adaptation mechanism to choose the optimal modulation and coding scheme for the current radio link quality, which is periodically reported by the mobile station.

Medium access control (MAC) protocol is the other part of layer 2. It is below RLC layer and it provides means for sharing a common transmission medium among several data flows (from one or more users) simultaneously. To the upper layers it offers unreliable data transfer. For multiplexing several data flows, MAC establishes *temporary block flows (TBF)*, where MAC PDUs are conveyed. MAC handles specific control signalling associated with certain data transmission (on PDTCH). This signalling is sent on *packet associated control channel (PACCH)* and it consists of signal measurement reports (from MS) and broadcast information (from network).

The physical layer (PHY) – layer 1 – offers logical channels to upper layers. It converts data on the logical channels into radio signal, which is then transmitted via electromagnetic waves to the receiver. Physical layer uses forward error correction coding to provide protection of the transmitted data against errors. It is also capable of detecting unrecoverable transmission errors, which are then reported to the upper layers. Physical layer also detects and reports congestion of the physical link [3GP02h].

2.1.3 UMTS

The *Universal mobile telecommunication system (UMTS)* is the third generation (3G) of mobile networks, which is expected to be truly global. It is largely based on the existing GSM/GPRS architecture and was specified in 3GPP Release 1999 (*R99*). UMTS R99 is composed of UTRAN and the 3G core network.

The main differences between UTRAN and GERAN lie in the radio access network (*RAN*). UTRAN offers wideband radio access through *wideband code division multiple access (WCDMA)* technology. As of 3GPP Release 5, UMTS contains not only UTRAN, but also GERAN. Both UTRAN and GERAN connect to the same 3G core network through the Iu interface (while GERAN also connects to the 2G core network through the A and Gb interfaces). The interworking between the two radio access networks and core networks is shown in Figure 2.8.

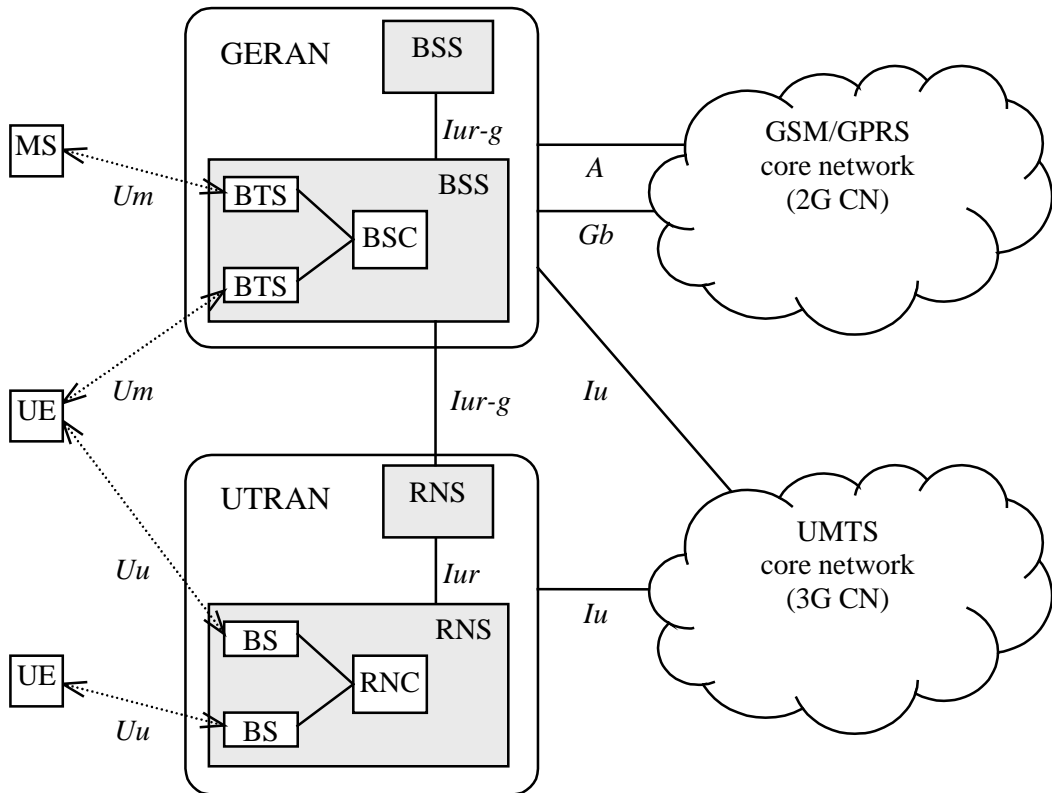


Figure 2.8: Interworking of 2G and 3G radio access networks and core networks

With the A interface GERAN connects to the CS part of the 2G CN. The Gb interface is for connection with the PS part of the 2G CN. In UMTS, the Iu interface has a comparable role to connect to the CN, but contains two parts: Iu-CS for connecting with CS part of 3G CN and Iu-PS for connecting with PS part of 3G CN (see Figure 2.9).

The UTRAN is divided into *radio network subsystems (RNS)* (the peer of BSS in GERAN) [Kaa01]. Each RNS consists of several base stations (also called *Node-B* in standardisation documents) connected to one *radio network controller (RNC)* that has similar role to the BSC in GSM. A mobile station connected to UTRAN is commonly referred to as a *user equipment (UE)*.

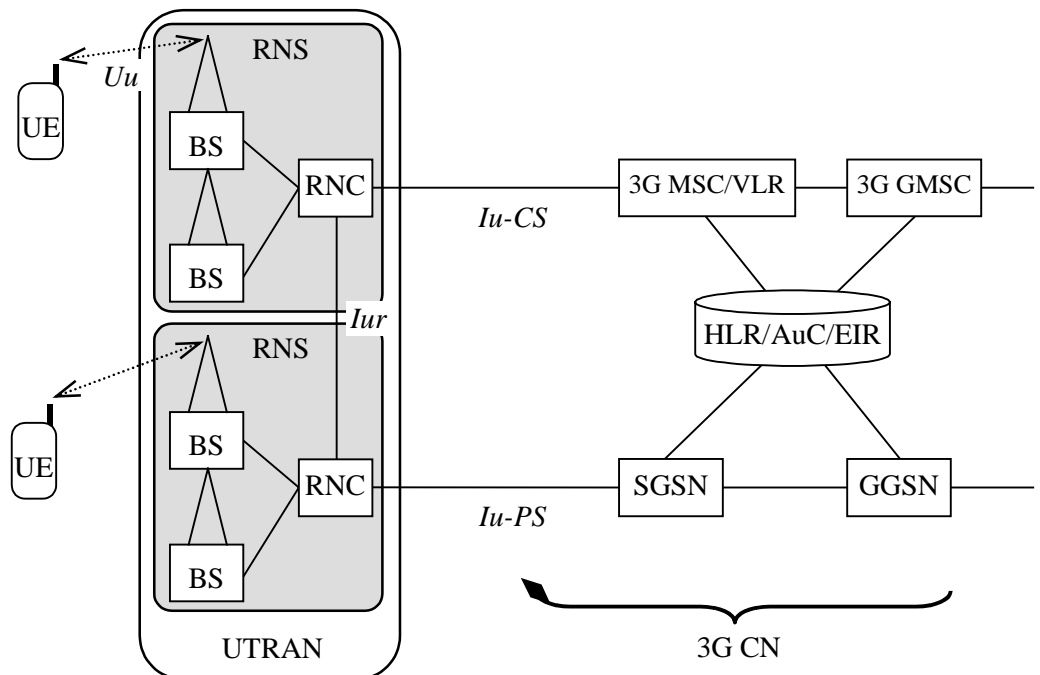


Figure 2.9: Architecture of UMTS with UTRAN

According to the nature of traffic processed in the UMTS core network, two domains are distinguished: circuit-switched and packet-switched. Figure 2.9 shows the basic structure of UMTS with UTRAN. The architecture of the core network does not differ much from GSM/GPRS; the major changes are in its components, interfaces and protocols.

The quality of services, which the cellular network offers to the users, depends on the perceived connection delay and data transfer rate. To describe the QoS requirements of various services offered by UMTS, four QoS classes have been defined [Kaa01]:

- Conversational class – minimum fixed-duration delay with certain guaranteed bit rate. Same amount of data in both uplink and downlink direction (symmetric traffic). This is the most demanding class. It is suited for real-time traffic such as voice calls.
- Streaming class – minimum variable-duration delay with guaranteed bit rate, asymmetric traffic. The experienced delay can vary because the network buffers the data on its way. For example streaming audio (Internet radio) could use this class.
- Interactive class – moderate variable-duration delay, asymmetric traffic. The bit rate is not guaranteed. This class is suited for web browsing.
- Background class – large variable-duration delay, asymmetric traffic and not guaranteed bit rate. This is the least demanding class, which could be used for file transfer such as e-mail download.

2.2 Overview of IP Multicasting

In this section IP multicasting is introduced. The following paragraphs discuss its main features and limitations, architecture, addressing and routing.

IP multicast is an addressing and forwarding scheme that allows delivery of datagrams sent by one source host to multiple destination hosts (receivers, listeners) simultaneously. It is based on UDP¹ and thus it is unreliable connectionless service. If multicasting is supported by the underlying hardware it is much more efficient way of delivering information to multiple targets than using conventional unicast because the packets are sent only once and distributed by the multicast capable routers to the receiving hosts. Unfortunately, in the current Internet there are not many such routers though this situation will change with the spread of IPv6, which supports multicast natively. The efficiency of multicasting grows with the growing number of receivers (in comparison to unicast). This property makes it attractive for various applications, such as distributed computation/simulation systems, Internet multimedia broadcasting (e.g. music, video), digital TV broadcasting, audio/video/data conferencing, querying and replication of distributed databases, etc. [Com00]

2.2.1 Multicast Groups

A set of receivers is called a *multicast group* [Com00]. Groups are distinguished and addressed by special IP addresses. There is one IP address for the whole multicast group. In the currently used IP version 4, class D addresses are reserved for multicast groups.

The number of members of the group can change over time; i.e. any host may join or leave the group at any time. The group can also have no members at all and still it would be possible to send datagrams to it. Further there are no limits imposed by the protocols on the location or number of members in a group. The local multicast router does not even know how many group members there are on the local network; it suffices to know that there is at least one. To get this information the multicast router asks all the hosts on its local network periodically. Any host can be a member of more than one group at a time. The source of multicast traffic does not need to be a member of the group it sends datagrams to.

In the current IPv4 the multicasting capability has been added later by extension as described in [Dee89]. For group management in IPv4, *Internet group management protocol (IGMP)* is used or one of its later variants IGMPv2, IGMPv3. In IPv6, which supports IP multicast from the beginning, the group management protocol is called *multicast listener discovery (MLD)*. Version 1 of this protocol has been standardised and now there is work in progress on its second version MLDv2. In local networks IGMP is used between hosts wishing to join or leave a group and their multicast router.

When a host wishes to receive multicast traffic of certain group it sends a *join request* to the local multicast router (also called *designated router* [Vai02]) that records group memberships of all host on the local network it is connected to. When a host no longer wants to listen to the group traffic it sends a *leave group* message. In addition to this, the designated router periodically sends multicast group membership queries. When no host responds for a group within some time the router assumes there are no receivers for the group and stops forwarding the group traffic to the local network. The multicast routers inform each other about the group members in their networks.

¹ User Datagram Protocol

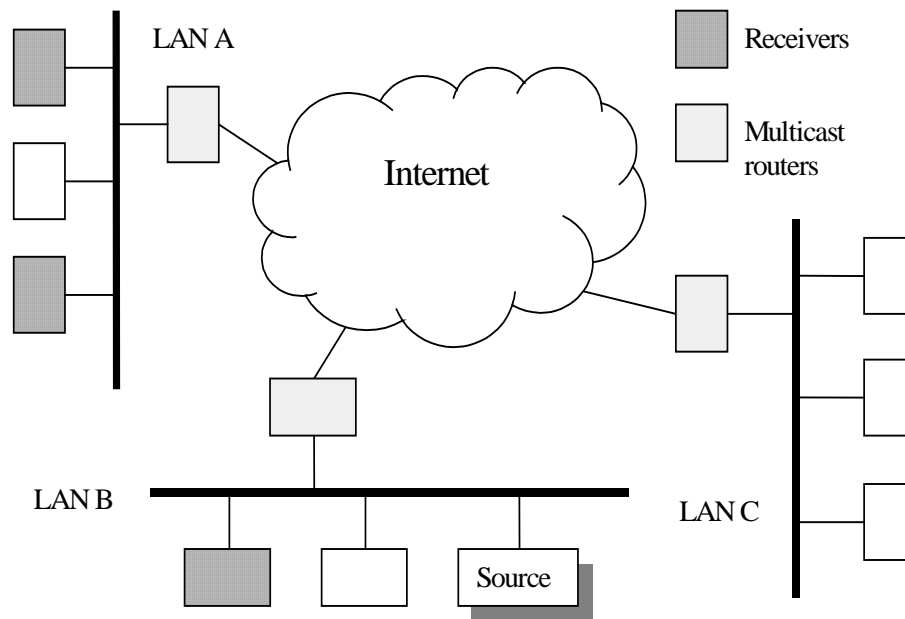


Figure 2.10: Architecture of networks using IP multicast

2.2.2 IP Multicast Routing

Multicast routers communicate with hosts on their networks using IGMP (or MLD) and co-operate with other multicast routers in the Internet using some multicast routing protocol to build a forwarding tree between sources and group members (see Figure 2.10 for an example of arrangement of source and receivers). There are currently several protocols, unfortunately none of them is standard. Multicast routing is quite complex a task and the best way to do it depends on several factors such as e.g. inter-network topology, reliability, quality and available bandwidth of internetwork links, traffic load and requirements put on the multicast service itself (e.g. quality of service or ensuring reliable multicast). There is no single protocol that could be used to deliver maximum performance under all circumstances.

Simple idea would be to use unicast (conventional) routing also for multicast traffic. But this does not work because there are several important differences between these two types of routing:

- routes can change not only when network topology changes or breakdowns happen but also when an application on a host joins or leaves a multicast group,
- multicast routing requires a router to examine more than the destination address. It must also observe from which network the packet came (source network ID) so that it does not transmit it unnecessarily to where it came from.

The multicast routing information held in routers forms a multicast distribution tree. There are two types of them:

1. Source trees – root of the tree is the source of multicast traffic. Branches form a shortest path spanning tree to the receivers. Routing table entries in routers are identified by a pair (source's network id, multicast group).

2. Shared trees – root of the tree is one selected router called *core router* or *rendezvous point (RP)* (depending on the protocol). Every source wishing to send traffic to the group sends its data to RP, from which the data is multicast to the receivers along paths forming shortest path spanning tree. The same tree is used by all sources, which send traffic to the particular group – the tree is shared by all sources. The entries in routing tables are identified by a pair (***, *multicast group*) where *** means any source network id. Having only one entry per multicast group results in much smaller routing tables in comparison to source trees, but the total routing path from source to a receiver might not be optimal.

2.3 Multimedia Broadcast/Multicast Service in Cellular Systems

Multimedia broadcast/multicast service (MBMS) is a bearer service of a cellular network that supports resource-efficient point-to-multipoint (*p-t-m*) transmission of multimedia data [3GP02b], but also point-to-point (*p-t-p*). Other services provided by operator or some external provider can use MBMS's capabilities to deliver its contents to multiple listeners simultaneously. Not only does it resemble IP multicast but it shall also be compatible with it, though there are several important differences due to the cellular network environment. They will be the subject of the next section.

Although there already is one broadcast capable service, namely the Cell Broadcast Service introduced briefly in subsection 2.1.1.3, it cannot be used for high bit rate data transfers required to deliver multimedia contents. As the interest in multimedia broadcasting has been rising among the operators in recent years and the current support for IP multicast in GPRS is inefficient [Vai02] and unsuitable for the mobile environment, there is a need for a new service, MBMS. In the following sections its requirements, features and overview of architecture will be described as it is outlined in specifications [3GP02b] and [3GP02c].

2.3.1 Differences Between MBMS and IP Multicast

MBMS differs from IP multicast in several aspects:

- It is unidirectional. The listeners cannot be the source of multicast traffic, the source is usually in the operator's network or it is some external content provider (see also Figure 2.12).
- The MBMS reception is limited to a *broadcast or multicast service area*, which is specified by the content provider. It is a geographical region such as town, city district, stadium, highway, etc. It can be as small as one cell and as large as the whole PLMN. The multicast service area is defined for each multicast group separately. The operator controls which cells can be used as service areas for MBMS through the network management system.
- MBMS provides several levels of quality of service. MBMS traffic will be delivered to the listeners in a quality as close to the one intended by the content provider as the network load and radio conditions permit. The perceived quality will be also limited by the capabilities of the user's mobile station.
- There are two distinct modes of operation – broadcast and multicast, which are described in the following sections.
- MBMS transmits data in sessions. Each service can have at most one active session in its service area. Service areas can be further divided into local service areas, where the contents of the session in each local service area can differ from each other. This way for example local news can be broadcast at the same time in the whole service area, yet be specific for the region.

- The network actively informs the subscribers about ongoing sessions and those about to begin.
- Certain data transmissions (mostly in multicast mode), for which the listeners are charged, are ciphered on the application level prior to broadcasting them through the radio to prevent unauthorised access to them.

2.3.2 Requirements

MBMS is to be provided through both GERAN and UTRAN. It should reuse the existing infrastructure and procedures as much as possible. Both broadcast and multicast mode shall use the same low-level radio access bearers. Its major goal is to make efficient use of the radio resources. So if there are several listeners of the same data in the cell they should share one radio channel to receive the data. For multicast mode of operation, it shall be possible to detect the presence of listeners of particular session in the cell and multicast it only if there are some. [3GP03f]

MBMS should provide transparent data delivery for the content providers. The content provider can be the operator itself or an external content provider that will be charged for using the operator's network. [Vai02]

To each service provided by an operator is associated a certain QoS class and level i.e. QoS requirements. The network is in charge of guaranteeing these requirements will be met, accounting for e.g. mobile stations' capabilities, radio conditions, cell load, etc.

Another requirement is that it shall be possible – in the standard to allow for transmitting several different MBMS sessions in a service area but also for receiving more than one session on a capable enough mobile station. MBMS reception should be possible when the mobile station is idle and if possible during a call or during an active data connection. During the reception of MBMS session it should be possible to receive information about other active sessions in the current location and it must be possible to receive paging messages for non MBMS services (e.g. voice call). Paging reception has precedence over MBMS reception. When the user moves between cells within the service area the data loss during the cell change should be minimal. [3GP03f]

MBMS users should be informed about current and future services by *service announcements*. These announcements must also inform the users, where the services will be available (in which service area). The services may be announced for example on the operator's web site, in newspapers, in teletext, by e-mail or SMS or by broadcast mode session. In any case, the user shall be able to discover the services offered in his current location and shall be able to choose whether to receive them or not. [3GP02b]

2.3.3 Broadcast Mode

In broadcast mode, the data is broadcast in the service area independently on whether there are any listeners in the individual cells or not. Users are not required to subscribe to broadcast mode as this service is intended to be provided free of charge to all users within the service area with MBMS compatible mobile stations. Thus, there is no need to encrypt the data. Still charging information is collected to charge the service provider.

Broadcast mode will be probably used for information services like advertisements, welcome messages, traffic news, etc. The network does not guarantee the delivery.

The provision of broadcast mode MBMS will be done in the following consecutive phases shown on the left-hand side in Figure 2.11. [3GP03g]

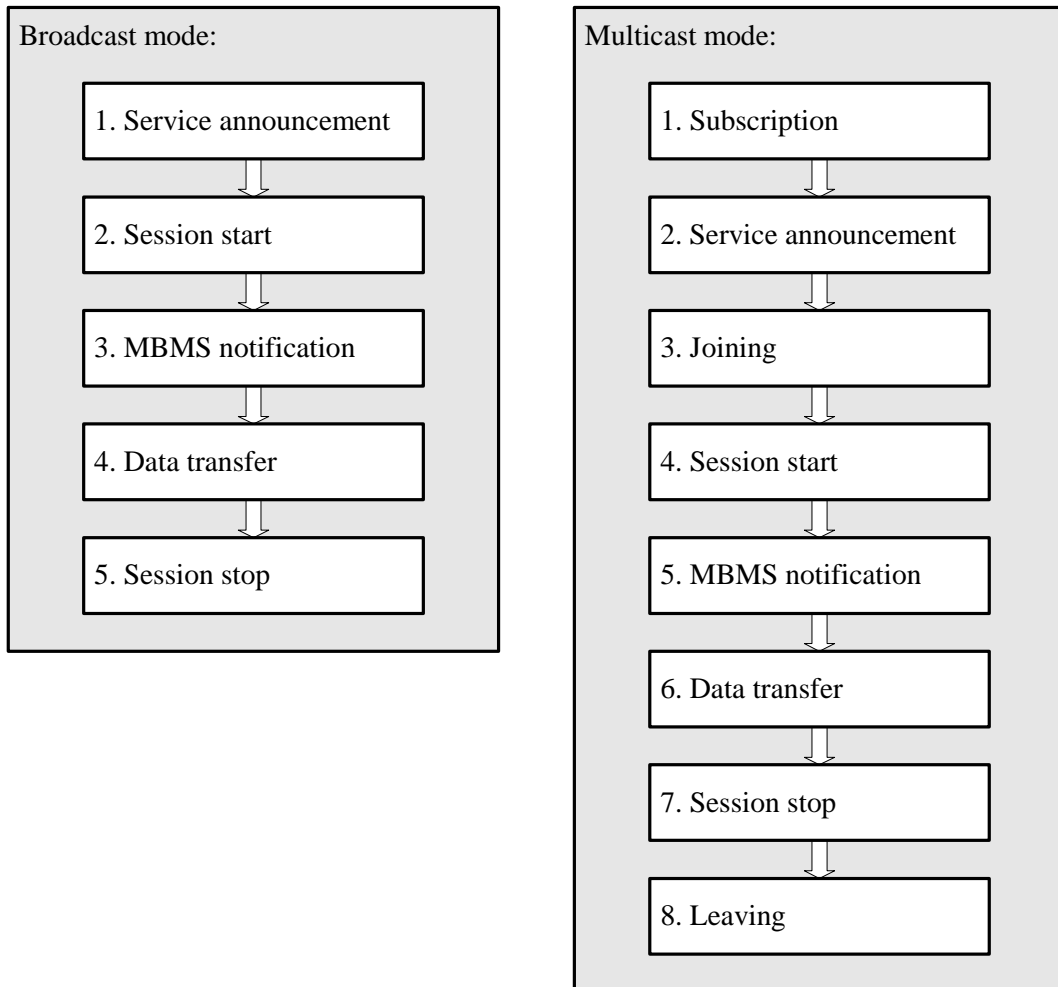


Figure 2.11: Steps of session reception in broadcast and multicast mode [3GP03g].

1. Service announcement – informs potential listeners about present and future services.
2. Session start – triggers network resources assignment for MBMS data transmission in the service area.
3. MBMS notification – informs mobile stations in the service area about currently beginning session and possibly about already on-going sessions too.
4. Data transfer – session data is broadcast in the whole service area.
5. Session stop – data transmission stops and network resources are released.

2.3.4 Multicast Mode

In this mode, each service has its associated multicast group. In addition to multicast groups, there exist also *multicast subscription groups*. A user, who subscribes to a multicast service, becomes member of the multicast subscription group associated with this service and is thus authorised to join the serv-

ice's multicast group. This enables him to receive session notifications for this service and receive the data. If he wishes to stop the reception, he leaves the multicast group. But if he wants to terminate the service itself, he has to unsubscribe from the multicast subscription group. The network will check if a user attempting to join a multicast group is a member of the associated multicast subscription group.

The provision of multicast mode MBMS involves the following consecutive steps depicted on the right-hand side in Figure 2.11. [3GP03g]

1. Subscription – establishes the relationship between the user and the service provider, which allows the user to receive the related multicast service.
2. Service announcement – informs potential listeners about present and future services.
3. Joining – subscriber joins the multicast group thus informing the network about his intention to receive the multicast session.
4. Session start – the network assigns resources for MBMS data transfer in the service area.
5. MBMS notification – informs mobile stations about currently beginning session. The MS also receives information about other, already ongoing services, to which the user is subscribed. The notification must be broadcast in the whole service area so that even those users, who have missed the first notification, can join the multicast group.
6. Data transfer – session data is multicast only in those cells of the service area, where listeners have been detected (either they joined there the session or moved into the cell from an adjacent one when already listening to the session.)
7. Session stop – data transmission is stopped and network resources are released.
8. Leaving – subscriber leaves the multicast group to stop the multicast mode data reception. This can also happen during the data transfer step.

At the beginning of the session, the network notifies users subscribed to a multicast mode service so that they can join the multicast group if they wish. They may also join later if the session is still ongoing. Similarly, they can leave the multicast group before the session is over to stop the reception.

The traffic content of multicast sessions is ciphered and users are charged for its reception. With ciphering involved, there is also the need for secure delivery of the ciphering keys to authorised users. Only members of multicast subscription group have decryption keys for the associated multicast group. Also charging information is collected for the content provider. Data delivery is not guaranteed. If a guaranteed delivery is required, it is left to the application layer to implement it.

2.3.5 Architecture

MBMS will be implemented in PS part of core network only. The preferred architecture proposed in the standard [3GP02c] is shown in Figure 2.12.

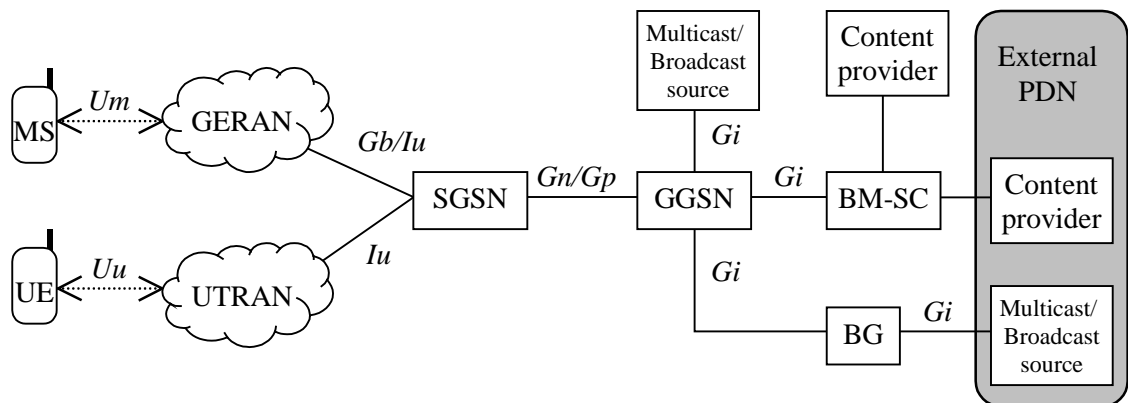


Figure 2.12: MBMS Architecture

The new component in the PS part of CN is the *broadcast/multicast service centre (BM-SC)*. It is the source of data for all multicast groups served by some content provider (for both broadcast and multicast mode). It determines the QoS level for the transmission based on content provider's requirements within operator specified bounds. It takes care of session scheduling and produces service announcements for the mobile stations. To content providers it offers functions to select the service area, to which the data will be delivered, and the mode of service. From the network's point of view, it acts as a gateway and thus it has a security function. To prevent malicious content insertion into the network it authenticates the content providers and checks integrity of data received from them. It also collects charging information.

BG stands for *border gateway*, which acts as a firewall between external multicast/broadcast source and the core network. The multicast/broadcast source may be also internal to the PLMN; in that case, it is connected to the GGSN directly. The difference between content provider and multicast/broadcast source is that content provider is charged for using the network whereas multicast/broadcast source is not. The latter one uses IP multicast on the standardised Gi reference point¹ while the interface between BM-SC and content provider is not standardised.

SGSN has a lot of functions, which can be used also in MBMS such as user individual mobility management and service control, generation of charging data for the subscribers and establishment and storing of *radio access bearers (RAB)* upon data arrival (see section 2.3.6).

GGSN plays a major role in routing of MBMS data, as it is the multicast core router for the PLMN in the sense of section 2.2.2. It reacts to joins from SGSNs and thus is involved in mobility management. It can tunnel the multicast mode data to subscribers roaming in a VPLMN. It may also carry out message screening for traffic incoming from an external multicast/broadcast source if it is not already done by border gateway.

Both GERAN and UTRAN shall provide means for efficient MBMS data delivery. For broadcast mode, p-t-m radio channels will be established in all cells of the service area regardless of the presence of listeners. For multicast mode, the RAN shall detect the listeners and multicast data only in those cells containing some listeners. Furthermore, it should decide whether to use p-t-p or p-t-m links depending on e.g. the number of active listeners in the cell. The mechanism for counting listeners on the network side is implementation specific. Typically, it would be sufficient – from a counting perspec-

¹ A reference point is similar to interface, but in contrast to an interface the information exchanged through a reference point is not specified [3GP02g].

tive to detect, whether the number of listeners is higher/lower than a certain threshold. A problem lies in the fact that using p-t-m links requires the BTS to multicast the data with maximum power to make sure all mobile stations will receive the data, in case cell coverage is desired. This increases the interference level in the cell. This is especially unacceptable in UTRAN, where the cell capacity depends on the interference level in the cell [Kaa01] and broadcasting with maximum power would decrease the cell capacity. On the other hand, when p-t-p links are used, power control is employed to limit the interference, but radio resources are wasted to deliver identical data to several listeners through separate channels.

2.3.6 MBMS Context And Delivery Path

Similarly to PDP contexts for data p-t-p connections, each MBMS session has its *MBMS context*. Because the context concerns multiple listeners, it contains only service specific information. It contains among other things the IP address of the associated multicast group and QoS parameters for the data transmission. The MBMS context is maintained by SGSN and GGSN involved in the data transfer.

Data delivery over the radio access network is realised by *radio access bearers (RAB)* – data paths from SGSN to mobile station. MBMS context together with MBMS RAB forms an *MBMS bearer*. MBMS bearer must be activated before the data delivery is possible. For broadcast mode it is activated by the network (BM-SC), for multicast mode it is activated by MS [Vai02].

2.3.7 Data Transfer in the Core Network

The data to be broadcast/multicast has to be delivered from the source (BM-SC, IP multicast source) to the base station controllers. There are two alternatives proposed in the standard [3GP02c] either to use IP unicast or IP multicast (both on top of GTP).

For multicast mode data, the network must decide to which BSCs to send the data according to the presence of listeners of particular service in its service area. The mechanism to discover whether there are listeners for a particular multicast service in particular cell belonging to the service area has not been specified yet. One possible solution is suggested in section 3.4.4. This mechanism will be used only when the potential listeners' mobile phones are in standby mode (see page 7) and thus their location is known only with routing area accuracy.

When IP unicast is used the GGSN makes copies of the IP packets for each GTP tunnel to SGSN and SGSN does the same for BSCs. On the other hand when IP multicast is used, there have to be established internal multicast groups for delivery of the GTP-tunnelled IP packets (across Gn interface). Then the GGSN can send the IP packets only once and they will be replicated along the multicast tree by intermediate routers on their way to SGSNs. For this to work the SGSNs must join the internal multicast group first (whose address is obtained from GGSN). Figure 2.13 shows an example procedure of this process:

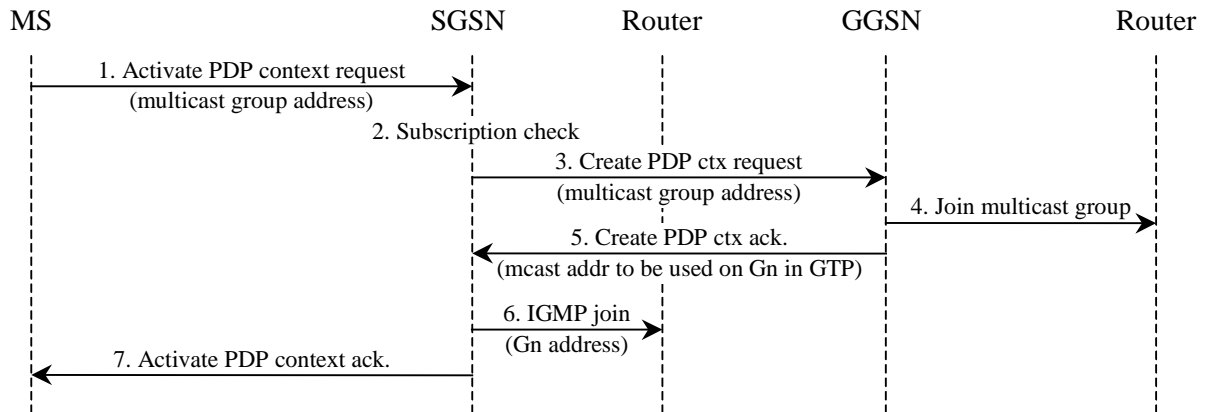


Figure 2.13: Example of possible activation procedure [3GP02c]

The obvious advantage of using IP multicast over unicast is efficient use of CN resources to transport MBMS data. The disadvantage is in the need for co-ordination of multicast addresses at Gn interface among the operators (to enable multicast mode MBMS reception to roaming users).

3. MULTIMEDIA BROADCAST/MULTICAST SERVICE IN GERAN

This chapter describes closer the suggested changes required to introduce MBMS into GERAN. It is based on existing specifications and proposals of various parties involved in the design of MBMS.

First, the GERAN specific requirements of MBMS are described in section 3.1. Then follows the list of new functions, which should be built into GERAN to support MBMS. After that follow sections dealing with MBMS context, notification and data transfer.

Similarly to IP multicast, in MBMS there will be multicast groups, which will play the role of logical targets for data delivery. Every multicast group will be addressed with *temporary multicast group identifier (TMGI)* within GERAN¹. TMGI assignment will be managed by BM-SC. When a mobile station joins a multicast group, it will receive the TMGI belonging to that group in the reply from BM-SC. From that point on, the MS can monitor the status of sessions belonging to this multicast group. The status of broadcast and multicast sessions is periodically broadcast in the cells in the form of notification messages. Notification is further described in section 3.4.

3.1 Requirements

The general MBMS requirements (described in section 2.3.2) form the base for several requirement assumptions on GERAN, which have been identified in [3GP03b and 3GP03f]:

1. A mechanism allowing the RAN to move mobile stations between cells is required.
2. MS controlled cell selection or reselection to ensure MBMS reception shall not be permitted.
3. Active MBMS session shall not affect handover and SGSN relocation.
4. Multicast mode transmissions should generally use p-t-m links (common resources) to deliver the session data. However to support simultaneous reception of other services (such as data/voice call) p-t-p links (dedicated resources) should be used.
5. To optimise the radio resource use, a mechanism for discovering multicast group members in idle mode is required to decide whether to start multicast mode transmission in a cell or not. Similarly, during the session transmission, the RAN should be able to discover whether there are still any interested listeners in the cell and if not, it stops the transmission.
6. Charging for MBMS should be transparent to RAN.
7. MBMS should allow low battery consumption of the MS.
8. MBMS should not prevent support for SGSN in pool.

3.1.1 Joining Requirements

It shall be possible to join at any time before or during the multicast session. During joining, the ciphering key could be transferred that is necessary for decrypting the contents of the session. For the RAN the whole joining procedure is transparent, i.e. does not require any RAN-specific mechanism.

¹ The same identifier should also be used within UTRAN to simplify inter-system cell reselections and handovers.

3.1.2 Notification Requirements

MBMS notification (see section 3.4) is broadcast in all cells of MBMS service area regardless of the presence of potential listeners. It may be transmitted periodically for the whole session duration so that mobile stations entering the cell after the session start can still receive the data. In cells, where the RAN is not sure whether there are any listeners or not, the notification will serve as a trigger for listener discovery (discussed in section 3.4.4). The RAN does not know mobile station's location at cell level if there is no connection established between the RAN and the mobile station. The notification message should be sent in such a way, that every MS, which has joined a multicast group, is able to receive the corresponding session (unless it is e.g. involved in a non-MBMS call). To this purpose, the notification delivery should reuse existing channels if possible. In addition, the notification should not require the MS to listen to some channel permanently. Rather it should allow the MS to listen in regular intervals and "sleep" otherwise, thus preserving battery life. This is a very similar principle to that of *discontinuous reception (DRX)* used when listening to paging messages. In addition to listening to MBMS notification the MS shall be able to monitor other channels such as paging channel or CBCH¹. Paging has however precedence over MBMS notification and the reception of notification is not guaranteed.

3.2 Functions

To support MBMS the GERAN should be extended with the following functions (based on [3GP02d and 3GP03b]):

- admission control for MBMS: when SGSN requests RAN to provide RABs for MBMS. Parameters could be: required capacity, QoS, priority. RAN knows available resources and can decide how to meet the parameters in current condition,
- selection between p-t-p and p-t-m RABs (implementation specific),
- p-t-m RAB establishment, modification and release upon request from CN,
- radio resource management for p-t-m communication: assign resources and select parameters for p-t-m RABs (based on parameters stored in MBMS context),
- notification: alert MSs that data transmission is started or on-going, indicate location of the shared p-t-m channel if it exists in the current cell and optionally parameters for listener discovery mechanism,
- discovery of listeners – joined MSs within cell (for particular multicast group),
- power control: set power level for p-t-m (suggested fixed), for p-t-p as for other services (existing power control),
- support of cell change and resynchronisation after that to minimise data loss,
- maintain MBMS context (create, update, remove): keep track of what MBMS sessions are active in each cell,
- replication of MBMS data stream to multiple cells (from BSC to BTSs),
- transfer of MBMS data over Uu/Um interface: header compression, segmentation/concatenation, multiplexing, coding,

¹ Cell Broadcast Channel. It is used to convey Cell Broadcast service messages.

- starting transmission on demand: when there are no listeners for an on-going session and a new listener appears (joins or comes from another cell).

3.3 MBMS Context

Every BSC within an MBMS service area will keep a data structure for holding information regarding MBMS: the MBMS context. For every active MBMS delivered by a BSC, there is one MBMS context. The contents of MBMS context could be:

- TMGI, (IP multicast group address and APN),
- broadcast/multicast mode flag,
- session is on-going flag (but session data not necessarily transmitted if in multicast mode),
- info about RAB as requested by SGSN,
- list of cells belonging to the service area of the particular MBMS, for every BTS it should contain:
 - cell ID,
 - session data is transmitted flag (does the traffic channel exist?) – if yes: parameters of the channel
 - counting process is active flag (waiting for responses),
 - wait time (parameter to spread responses when counting),
 - timestamp of last counting completed,
 - number of received valid responses during the last counting (= estimated idle listeners),
 - list of IMSIs of connected mode listeners – this information is obtained from core network and can be used for counting or to decide that counting is not needed.

3.4 Notification

Notification is a message sent to inform joined listeners (in case of multicast mode MBMS) or broadcast mode listeners about a starting or an on-going MBMS session. The notification mechanism must ensure that users in idle mode can be notified, hence common channels are being considered as appropriate bearers to convey the notification. At the time this thesis is written, 3GPP has not selected a notification mechanism in GERAN. The candidate solutions are described in the following subsections.

The notification indicates:

- Existence of broadcast or multicast session in the current cell. The session would be identified by a service identifier, e.g. TMGI.
- Whether counting (mechanism of listeners' discovery) is used or not, and if so the appropriate parameters for counting are provided to the MSs. See subsection 3.4.4.
- Optionally, set-up parameters of radio bearer, where the session data will be transmitted. These include at least timeslot allocation, temporary flow identifier, TBF starting time, frequency parameters, etc. similarly to packet downlink assignment message, described in detail in [3GP02f]. If these parameters are not present in the notification message, it means that the session is on-going but is not currently transmitted in the cell because RAN has decided there are not enough interested lis-

teners. If an MS enters such a cell and wants to receive the session, it must inform the RAN via uplink signalling about its interest. The parameters for the uplink signalling could be sent in the notification message as part of the mechanism to discover MBMS's listeners in the cell.

The notification messages must be sent repeatedly during the sessions to enable session reception to MSs, which were switched off during the initial notification, were outside the service area or were changing cells. Even mobile stations in connected mode might not always be able to monitor the channel on which notifications are sent and might therefore miss notifications.

3.4.1 Notification on (P)BCCH

In this solution, proposed in [Nok03a], the notification message would be sent on *packet broadcast control channel (PBCCH)* as a new *packet system information (PSI)* message or a new information element of an existing message. Every mobile station has to read information broadcast on PBCCH if the channel is present in the cell (while the existence of BCCH is required for the support of GSM/GPRS, that of PBCCH is optional for GPRS). After receiving the notification message, MS would store it. A new notification message or change in contents of an existing one would be announced by PBCCH_CHANGE_MARK and PSI_CHANGE_FIELD information elements of PSI1. This is an existing mechanism, specified in [3GP02f], for announcing changes in system information, which lets MSs to read the system information only if there has been a change. If the notification would be sent in a new PSI message, only MBMS capable mobile stations would react to the changes of this PSI and reread it. The capacity required for the notification would depend on the number of simultaneously active sessions and on the repetition rate of the message.

3.4.2 Notification on (P)CCCH / (P)NCH

Another possible solution, described in [Sie03a], would be to send the notification messages on *packet notification channel (PNCH)*, which is a logical channel, part of *packet common control channel (PCCCH)* [3GP02i]. In the multiframe, PNCH blocks can occur on the position reserved for *packet access grant channel (PAGCH)* blocks. The MS will recognise to which logical channel the block belongs by reading the message type information element in the block header. Unfortunately this would require the MS to receive all the blocks reserved for PAGCH, which would result in high battery consumption. Therefore new parameters would be introduced into a PSI message describing the position of PNCH within the multiframe and its repetition rate. This will enable discontinuous reception and the MS will read the PNCH only when its blocks are actually broadcast. Further optimisation of power consumption lies in adding a small information field into every paging message sent on PPCH. Change in this field would indicate a change on PNCH and the MS would read it on the nearest occasion.

3.4.3 Notification on (P)PCH

Notification could also be provided by the existing paging mechanism. Solution based on PPCH was proposed in [Eri03]. Every paging message would contain a new information field that would indicate if a notification is being broadcast or not. The notification information would be sent in rest octets of a paging message sent in pre-defined paging group.

After the MS receives indication of notification in some paging message it would listen on this pre-defined paging group and read the notification. If the MS is interested in receiving the session identified by TMGI in the notification, it starts monitoring a special MBMS control channel. The location of

this channel is also given in the notification. This channel would convey any further information, such as whether counting will be used and its parameters and radio bearer set-up parameters for the channel, where the session will be broadcast.

3.4.4 Counting

Counting is a mechanism for discovering joined multicast group members in the cell that are in idle mode, thus the RAN does not know their location with cell accuracy. Counting can be performed before the radio bearer set-up to help RAN decide, whether to start the session transmission in a given cell or not, to fulfil requirement 5 in section 3.1. Activation and deactivation of counting is controlled via notification message separately for each cell of the service area. The BSC decides whether to perform counting in a cell based on how many connected mode listeners are there or how many idle mode listeners are expected to be there (e.g. no counting is performed in a hotspot area such as stadium). When the mobile station detects from the notification, that counting is going on, it shall respond to the RAN. There have been various suggestions how to implement the counting mechanism, some of which are briefly described below.

In one solution, proposed in [Qua03], the counting is indicated by an information element in the notification message with two values, meaning: “register” and “do not register / stop registering”. If the MS interested in the session detects the “register” value, it considers another parameter sent in the notification – “wait time” and determines a time instant (randomly) when it sends its response to the network. The response message would contain a user identifier and the identification of the multicast service (TMGI) the MS is interested in.

In another solution, explained in [Nok03b], the RAN would instruct at least one MS in each cell of the multicast service area to stay in connected mode. That way the MS would inform RAN when it changes cell. If the MS enters a cell within the service area but without the session being transmitted, it registers and the RAN starts the transmission. If the last connected mode MS leaves a cell, RAN instructs via notification idle mode mobile stations to register. If there is no response within certain time, RAN assumes there are no interested MSs and stops the session transmission in that cell.

3.5 Data Transfer

The session data will be transmitted on a PDTCH common to all interested listeners using existing modulation and coding schemes. This requires the BTS to transmit data with sufficient power for the signal to reach all the listeners. Moreover the BTS will a priori have no individual feedback from the receiving mobile stations, therefore power control will not be used and the transmitting power will be constant.

To keep the existing paging reception and measurement mechanisms unchanged, the PDTCH carrying session data should be allocated on the same and adjacent timeslots as PBCCH and PCCCH as proposed in [Nok03c].

The following subsections contain description of additional mechanisms, which are being studied as possible remedies for high RLC SDU error ratio of the proposed p-t-m links, whose performance is discussed in section 4.4.1.

3.5.1 Reed-Solomon Codes

Reed-Solomon (RS) codes are forward error correcting codes. They add redundant information to the data. If the data gets corrupted during the transmission, RS codes can automatically repair the errors up to certain limit. RS codes are non-binary linear codes especially suitable for correcting burst errors [REF Siemens_GP_031994_Outer_coding_for_MBMS \h Sie03b]. They are defined over Galois field 2^m by:

$$(n, k) = (2^m - 1, 2^m - 1 - r)$$

where

n is the number of symbols in the codeword (block length),

k is the number of systematic symbols in the codeword (i.e. the data symbols),

m is the number of bits per symbol and

r is the number of parity symbols.

A Reed-Solomon decoder can always fully recover an erroneous codeword, which contains s erroneous symbols in unknown locations in the codeword and t erroneous symbols in known locations (symbol erasures), when the following inequality holds:

$$s/2 + t \leq r$$

The codeword length n of the RS code can be adjusted to the application by shortening the code. For example a (255, 230) code can be shortened to a (200, 175) code while maintaining the same error-correcting ability [REF Siemens_GP_031994_Outer_coding_for_MBMS \h Sie03b]. The shortening is realised by adding d dummy symbols with value 0 before encoding at the transmitter side. After encoding, only $k - d$ systematic symbols and $n - k$ parity symbols are transmitted. At the receiver side, the d dummy symbols are inserted into the codeword before the decoding process.

By choosing $m = 8$, the symbols of the code correspond to one octet and by shortening the RS code it can be adapted to various block sizes. The outer coding using RS codes can be applied at RLC layer as proposed in [REF Siemens_GP_032100_Outer_coding_on_RLC \h Sie03c] or on the application layer with processing in BM-SC [REF Siemens_GP_031994_Outer_coding_for_MBMS \h Sie03b].

When outer coding is applied at RLC layer the RLC SDUs are segmented into blocks of k symbols and these segments are encoded. The resulting blocks of length n are interleaved among several RLC data blocks so that a loss of one RLC data block affects only one symbol in an encoded block (codeword). All symbol errors within codewords can be treated as symbol erasures, because the receiver can detect an erroneous RLC data block, it knows the location of the erroneous symbol and can pass this information to the RS decoder. Therefore up to r symbol errors can be corrected.

On the other hand, outer coding could also be used on the application layer, as proposed in [Sie03b]. In BM-SC the RS encoder takes k SDUs (systematic) and generates $n - k$ parity SDUs out of them by encoding the "columns" of symbols from the systematic SDUs (see Figure 3.1). The parity SDUs would be transmitted together with systematic ones, which would be corrected if an error occurs during the transmission. As there is no information about erroneous RLC data blocks on the application layer, those SDUs with bad checksum shall be treated as if all symbols within are bad (the worst case). The length of the SDUs determines the number of codewords (in columns) affected by one erroneous SDU. The RS decoder would be able to correct r symbol errors because it knows, which SDUs are received in error.

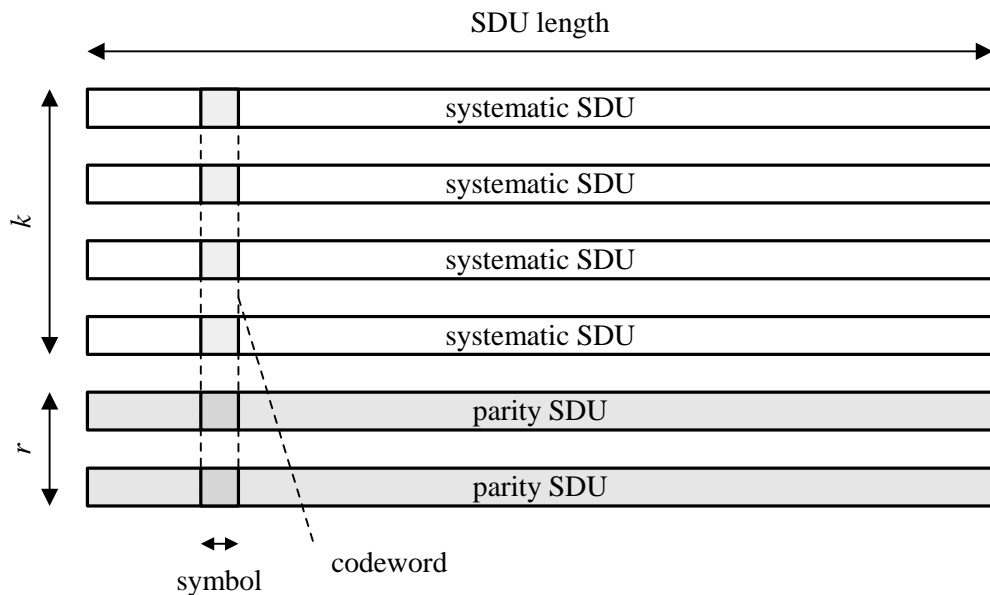


Figure 3.1: Outer coding on application layer (adapted from [Sie03b])

3.5.2 Feedback in Uplink

Another solution to lowering the SDU error ratio in p-t-m transmission was proposed in [Sie03d]. As opposed to RS codes, this solution is a backwards error correcting mechanism – it adds redundancy (here in the form of retransmissions) only after errors are detected and reported. A special common uplink channel would be reserved for feedback from MSs. If the MS does not receive an RLC data block transmitted at time t , it will send negative acknowledgement message (*NAK*) in the form of access burst at time $t + \Delta t$. The transmitter at network side would listen on the feedback channel for access bursts. If there is one detected at time $t + \Delta t$ the network knows there have been at least one MS, which did not receive an RLC data block sent at time t and may retransmit that block.

In this solution there will probably be problems when detecting *NAK*s sent from multiple MSs as these will collide. Therefore the network might measure the received power on the feedback channel and decide that a *NAK* has been sent if the received power is over certain threshold. However the high power might also be caused by interfering signals from other cells and thus might lead to useless re-transmissions and waste of radio resources.

4. EVALUATION OF MULTIMEDIA BROADCAST/MULTICAST SERVICE IN GERAN

In this chapter the performance of MBMS is evaluated. To show the possible gain which could be achieved using p-t-m in comparison to using p-t-p links to deliver the data, series of network simulations were carried out. As MBMS is not the only service in GERAN its impact on other services in the system has to be estimated as well. The simulations show how the different approaches (p-t-m and several variants of p-t-p) to session data delivery behave during busy hour time in the simulated cellular network.

First, the simulation tool is described. Then follow the general description of simulation model as well as the model covering essential features of MBMS such as point-to-multipoint transmission. The next section describes the simulation assumptions and scenarios used for the performance evaluation. In the last section, obtained simulation results are analysed and the achievable performance gain is illustrated in comparison with different p-t-p scenarios.

4.1 Description of the Simulator

There are two categories of GSM/GPRS simulators: static and dynamic.

In static simulators there is no relationship between the simulation steps. Static simulators take into consideration the positions of mobile stations and their mobility model, base transceiver stations, the transmission power, the interference in the system produced by the transmitters and propagation model of the environment. Based on that they compute received powers, carrier to interference ratios, bit error rates, block error rates and other statistics. But they do not model movement of mobile stations and thus they do not model procedures and traffic load connected with user mobility. Because the individual simulation steps are independent of each other, algorithms taking advantage of feedback cannot be simulated. The results of such simulations are too optimistic and can be used only as upper-bound estimations [Hal02].

On the other hand a dynamic simulator considers causality between simulated events. This enables the simulation of events, which are triggered by the movement of mobile stations, such as handovers and power control commands. The dependency between events enables to simulate the behaviour of feedback mechanisms, which are present in many algorithms. Dynamic simulators can be event-driven or time-driven. Event-driven one simulates the sequence of individual events as they appear in time. Each event might involve the simulation of tasks, which in turn can trigger other events with some delay in between. Such behaviour is typical for protocols. The time interval between individual events is usually variable.

Time-driven simulators have built-in internal clock that controls the whole simulation. Every clock tick determines one simulation step, during which new events may happen and the state of the simulated system may change. The length of simulation step is usually fixed. Time-driven simulators have less overhead than event-driven ones in case there is large number of events to simulate.

SMART is the GSM/GPRS network-level simulator used to obtain the results in this study. Network-level simulator simulates the flow of packets and messages over the network [Jer00]. SMART is a dynamic, time-driven simulator with fixed time steps of the duration of one TDMA frame (4.615 ms). Part of the simulator uses also event-driven approach to simulate for example handovers and power control. The simulator is written in C++, whose object orientation and high execution speed is well

suited for the complex task of simulation. SMART offers detailed modelling of user mobility and network layout. There are several network layouts available, for example:

- Micro cell – small cells scattered among buildings, which act as considerable obstacles to signal propagation. This scenario resembles typical cell layout in a city.
- Macro cell – hexagonal cell grid in open area.

Other layouts with different topology, antenna types and transceiver parameters can be defined via configuration files.

The simulator employs pseudo-random number generator to simulate randomly occurring events. This is important as it makes the simulation repeatable by using the same input parameters and seed settings for the generator. It also enables to run series of simulations with identical input parameters but the random seed, which can be used to average the results over several different sets of results to obtain higher confidence in the outcome.

The link level behaviour of the cellular system is not simulated in this network level simulator. Instead the bit error probability, dependent on radio link conditions and properties, is obtained via look-up table generated by separate link level simulators. This simplifies the design of the simulator and also helps to speed up the simulation.

4.2 Simulation Model Overview

When a new connection is started, the MS is placed at a random spot in the cellular network. The calls arrive to the network according to Poisson process. Every call belongs to certain service that is decided at connection start and the distribution of calls follows the service proportions given as simulation parameters. The simulated mobile stations move in random directions with constant speed. All transmitting entities in the network are potential sources of interfering signal for the others.

The simulation proceeds in steps, which consist of the following operations (based on [Hal02]):

- mobile stations move to new locations and calculate the level of received signal,
- every active link:
 - calculates carrier-over-interference ratio at location of its receiving terminal,
 - receives radio bursts and radio blocks,
- radio resource management algorithms are executed (channel scheduling, cell reselection, power control, etc.),
- new connections are started and finished connections are removed,
- the simulator updates general statistics.

The simulator was originally designed to support GSM/GPRS network hence allows for point-to-point connections only. This resulted in a model where both peer entities (network and mobile station) are united in one notion of “connection”. This model is enforced throughout the simulator structure and thus the new point-to-multipoint connection, fundamentally different from a p-t-p connection, requires major additions to the original simulator. In the implemented simulation model, a p-t-m connection is modelled by two kinds of entities:

- p-t-m source – is the transmitting part with limited receiving part i.e. the network and

- p-t-m listener – is the receiving part with limited transmitting part i.e. a mobile station. P-t-m source partly controls p-t-m listeners in that it activates and deactivates their receiving links according to block scheduling performed by packet control unit. Also when the source's TBF is released, all its listeners' TBFs are released too.

The simulation tool simulates only the two lowest layers: physical and RLC/MAC (see subsection 2.1.2.4). Data enters and leaves the simulation at the LLC/PDCP layer in the form of RLC service data units.

4.2.1 Point-to-multipoint Data Transmission Model

In the implemented simulation model there is exactly one p-t-m source per each BTS (the multicast service area covers the whole network). All p-t-m sources in the network broadcast the same session traffic belonging to one multicast group. This means that there is simulated only one multicast session in the network. This session lasts for the whole simulation. When a new p-t-m streaming “connection” is started (p-t-m listener) it attaches to a p-t-m source in the current cell and starts receiving session data. After it has received the amount of data specified by a parameter (480 kB in this case) it disconnects from the source and terminates.

If the listener's MS moves into another cell during the reception a cell reselection takes place. The p-t-m listener detaches from the p-t-m source in the old cell and after it has selected new BTS it attaches to the p-t-m source of the new cell. If there is some partially received RLC SDU in the listener's LLC/PDCP buffer it is lost. So are all subsequent RLC SDUs transmitted until the link is re-established in the new cell. No data flow synchronisation after cell reselection is performed.

Figure 4.1 shows more detailed structure of entities involved in simulating a p-t-m connection.

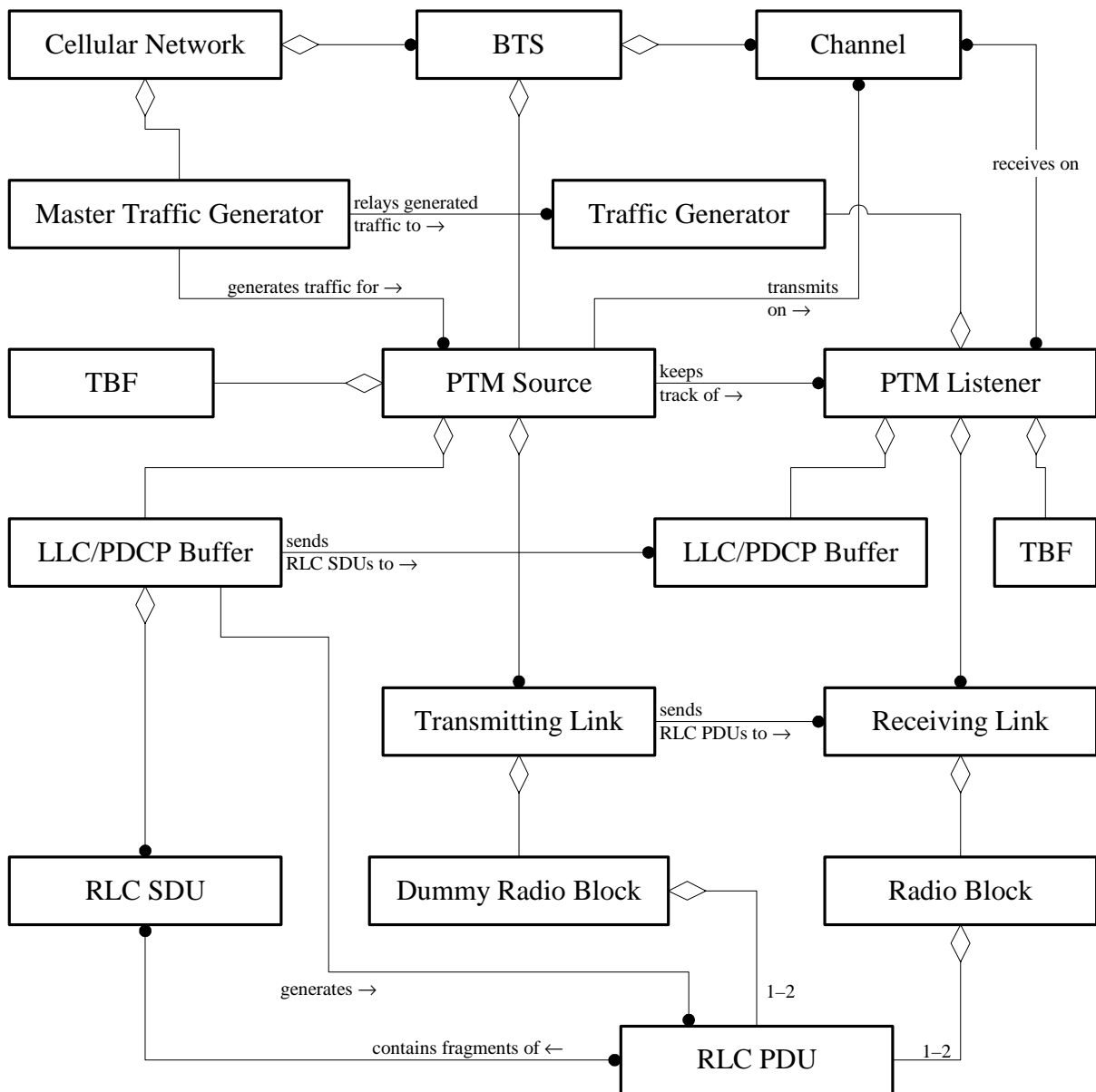


Figure 4.1: Simulation model structure

P-t-m source creates RLC SDUs according to the amount of traffic generated by the master traffic generator and stores them in its LLC/PDCP buffer. The LLC/PDCP buffer (in the model) contains partly RLC protocol functionality since it takes care of segmenting RLC SDUs into RLC PDUs and their storage for later assembly back into RLC SDUs. The master traffic generator is shared by all p-t-m sources in the network. When the source's transmitting link is allowed to send data it requests one or two RLC PDUs from the LLC/PDCP buffer depending on current modulation and coding scheme.

Transmitting link then sends the created RLC PDU to receiving links of all p-t-m listeners currently attached to the p-t-m source. In the receiving link, the RLC PDU is analysed and if it contains fragments of not-yet-received RLC SDU, then this SDU is transferred to listener's LLC/PDCP buffer. After analysing the RLC PDU the receiving link wraps it into a radio block. Then the reception of four

bursts comprising this radio block is modelled. After receiving each burst the experienced carrier-over-interference ratio at the listener's current location is mapped to bit error rate. When the whole radio block has been received the bit error rate from all bursts is averaged and the result is used to decide whether the radio block is correctly received or not. When all RLC PDUs from one RLC SDU are received (correctly or not) the RLC SDU is removed from the listener's LLC/PDCP buffer. At the same time, the source's transmitting link removes it from source's LLC/PDCP buffer.

Listener's LLC/PDCP buffer keeps track of how many bits remain to be received. Every received RLC PDU decreases this number by the amount of bits contained within until all session data is received. Then receiver's link terminates the temporary block flow and the connection representing p-t-m listener is finished.

4.3 Simulation Assumptions

Traffic load generated by signalling is not taken into account in the simulations because the corresponding traffic volume is negligible compared to the one generated by data connections. However the delays caused by signalling are simulated, for example delays in setup and tear down of layer 2 links (temporary block flows, TBFs).

When using p-t-p links the data flows are not synchronised, every data flow is independent of others. On the other hand the traffic delivered by p-t-m links originates in a single traffic generator. Yet the different load in each cell can cause the throughput to vary across cells and thus the data flows are synchronised within a cell but they are not synchronised across the whole network.

The simulation was aimed at performance comparison between p-t-m and various p-t-p links instead of detailed simulation of MBMS. Therefore the MBMS session is not simulated as such and the listeners can receive complete session data at arbitrary moments (unless the simulation ends before finishing the transmission). In scenario using p-t-m links (see subsection 4.3.1), the network broadcasts the session data from the beginning of the simulation to the end. During the simulation any newly joined listener will receive the complete session data regardless of the moment when they join.

4.3.1 Simulation Scenarios

The performance evaluation of MBMS is based on comparison between several scenarios:

1. PTM – session data is delivered using the new proposed p-t-m links,
2. PTP – session data is delivered via existing p-t-p links, using RLC unacknowledged mode,
3. PTP ACK – session data is delivered via existing p-t-p links, using RLC acknowledged mode,
4. PTP LA – session data is delivered via existing p-t-p links, using RLC unacknowledged mode and link adaptation,
5. PTP ACK LA – session data is delivered via existing p-t-p links, using RLC acknowledged mode and link adaptation.

All scenarios were simulated in two networks of identical layout (Figure 4.2) and traffic load parameters. Each scenario was simulated under four different loads: 20 000, 15 000, 10 000 and 5 000 users in the network. The simulated users were spread randomly according to a uniform distribution across the whole network and were moving with a pedestrian speed of 3 km/h in random directions.

The simulated cellular network used 900 MHz band and consisted of 75 BTS each having 3 transceivers. Three cells were grouped together to form a tri-sector site except for the border of the network where there are two 1-sector sites and two 2-sector sites. In the first network (normal network) the hexagons depicting the cell coverage had 500 m radius, while in the second network 200 m cell radius was used (hotspot network). The BTS used directional antennae with 65° beamwidth. Arrows in the figure show the direction of the antenna beam.

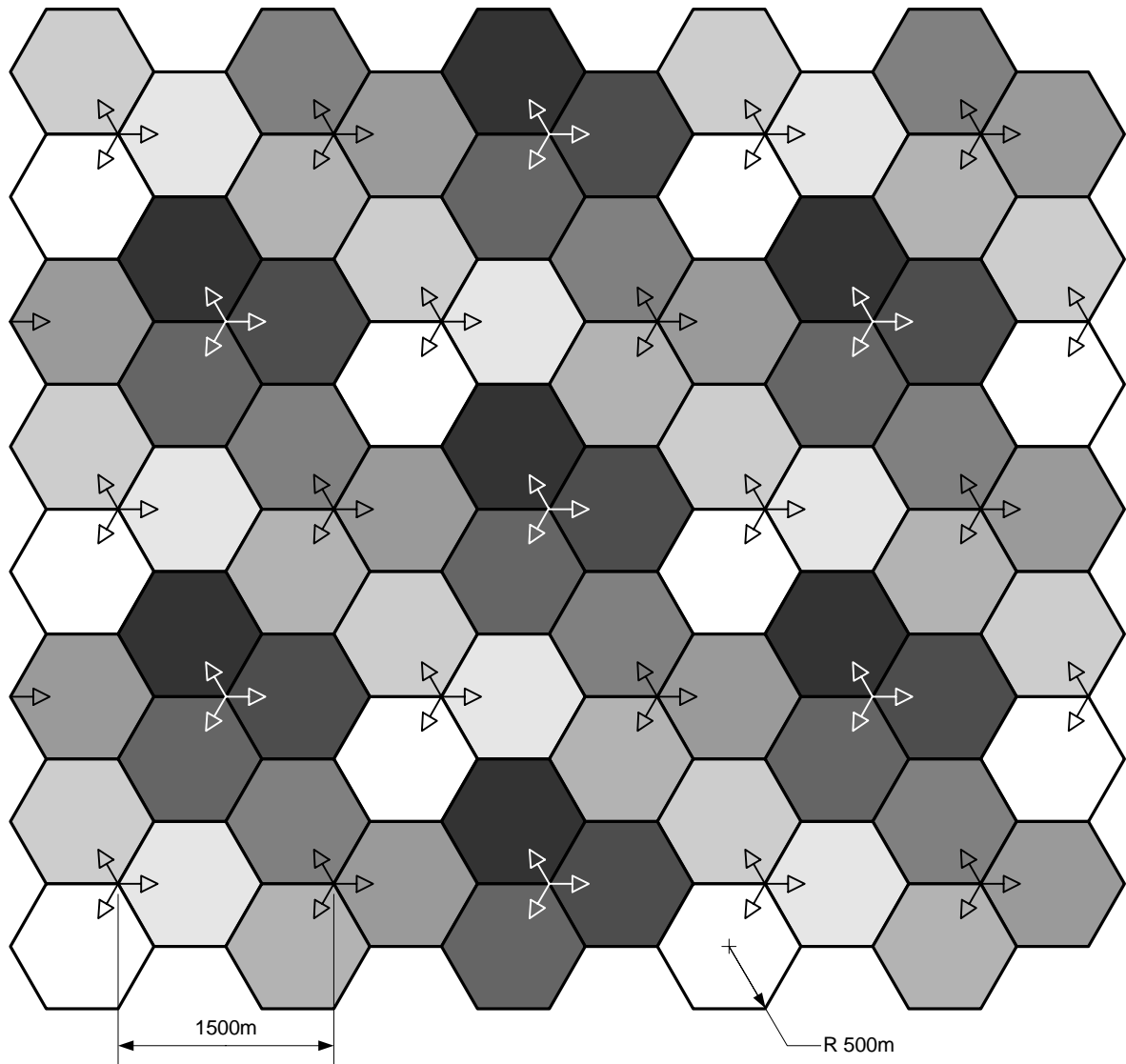


Figure 4.2: Layout of the normal network (3/9 frequency reuse, 5/15 for BCCH – not shown)

Frequency hopping was not used during the simulations. Frequency reuse was 15 different frequency sets every 5 sites for BCCH channel (5/15) and 9 frequency sets every 3 sites for the other channels (3/9). MBMS was simulated on transceivers, where the BCCH channel was located. Uplink traffic was not simulated.

In downlink, there were simulated two kinds of services: FTP file download and MBMS streaming. FTP file download was simulated to generate some interference in the system and also to see, how its

performance would be affected by the MBMS streaming. There were five different test scenarios described below. Every scenario consisted of background FTP file transfer always using p-t-p links (i.e. the FTP service was not using the MBMS bearer) and an MBMS streaming using p-t-m or p-t-p links. For MBMS streaming using p-t-p links, various combinations of the benefits of two-way communication were selected: RLC acknowledged mode with link adaptation, incremental redundancy and power control. Table 4.1 summarises the features used in each scenario.

The traffic proportions and characteristics of the simulated services were as follows:

- 75% FTP, 120 kB file per connection, EGPRS p-t-p links using link adaptation, power control, operating in RLC acknowledged mode with incremental redundancy, background traffic class, max. 3 timeslots in downlink,
- 25% streaming, 480 kB per connection, guaranteed throughput of 12 kbit/s, streaming traffic class, max. 1 timeslot in downlink. The traffic consisted of RLC SDUs of constant size 300 bytes emitted with constant bit rate of 32 kbit/s.

In PTM scenario EGPRS p-t-m links with MCS-5 [3GP02h] were used and because there is no uplink communication, feedback needed for power control cannot be sent and the BTS transmits with full power. There are also no retransmissions (RLC unacknowledged mode is used) and the link adaptation mechanism cannot be used as well due to lack of uplink feedback. MBMS broadcast mode was used (see section 2.3.3) thus the BTS were transmitting even when there were no joined listeners in the cell.

Table 4.1: Summary of scenario parameters

Scenario	RLC mode	Incremental redundancy	Link adaptation	Power control	MCS
PTM	unacknowledged	no	no	no	MCS-5
PTP	unacknowledged	no	no	yes	MCS-5
PTP ACK	acknowledged	yes	no	yes	MCS-5
PTP LA	unacknowledged	no	yes	yes	variable
PTP ACK LA	acknowledged	yes	yes	yes	variable

The call arrival rate was 5 calls/user/hour. Results were obtained through averaging 5 simulations. The lengths of the simulations under different loads were chosen so that the amount of simulated connections hence data collected for the results is about the same.

The simulations were intended to simulate cellular network during a busy hour, when the number of active connections is about constant. To achieve this the first 50 000 steps (TDMA frames) of the simulations were run without collecting any data thus letting the network to “warm up”. The number of steps required for a stable operation was determined experimentally.

The simulation parameters are summarised in Table 4.2 and Table 4.3.

Table 4.2: Summary of simulation parameters

Parameter	Value	Comment
Band	900 MHz	GSM
Number of BTSs	75	3 BTSs per site
Number of transceivers	3	per BTS
Cell radius	500 m / 200 m	normal / hotspot network
Site distance	1500 m / 600 m	normal / hotspot network
Frequency reuse	3/9, 5/15 (BCCH)	no frequency hopping
Call arrival rate	5 calls/user/hour	
Speed of MSs	3 km/h	
Proportion of FTP traffic	75 %	FTP traffic is only sent through p-t-p
Amount of FTP data	120 kB	
Traffic class for FTP	background	
Max. DL timeslots for FTP	3	
Proportion of streaming traffic	25 %	Streaming traffic sent through either p-t-m or p-t-p
Amount of streaming data	480 kB	
Traffic class for streaming	streaming	
Max. DL timeslots for streaming	1	
Guaranteed throughput ¹	12 kbit/s	used for scheduling
Streaming RLC SDU size ²	300 bytes	constant size

Table 4.3: Summary of simulated loads and simulation lengths

Number of users	Number of sim. steps ³ [TDMA frames]	Simulated time [mm:ss]
5 000	400 000	30:46
10 000	200 000	15:23
15 000	130 000	10:00
20 000	100 000	7:41

To see how the p-t-m links would perform in a hotspot area¹, simulations with all link types were carried out also in the network with small cells – the hotspot network. The cell radius was 200 m (and site

¹ This value includes also unsuccessfully transmitted bits as the scheduler has no feedback from the receivers.

² The value of 300 bytes was chosen to limit the amount of RLC PDUs into which an SDU is segmented, thus limiting the SDU error ratio.

³ Not counting the initial 50 000 steps for warm-up.

distance shrank to 600 m). In Figure 4.3 the size of the hotspot cells can be compared to the original cells. The 200 m radius was chosen because in the PTM scenario of the original network the MSs in the area covered by such cell would experience quite low RLC SDU error ratio.

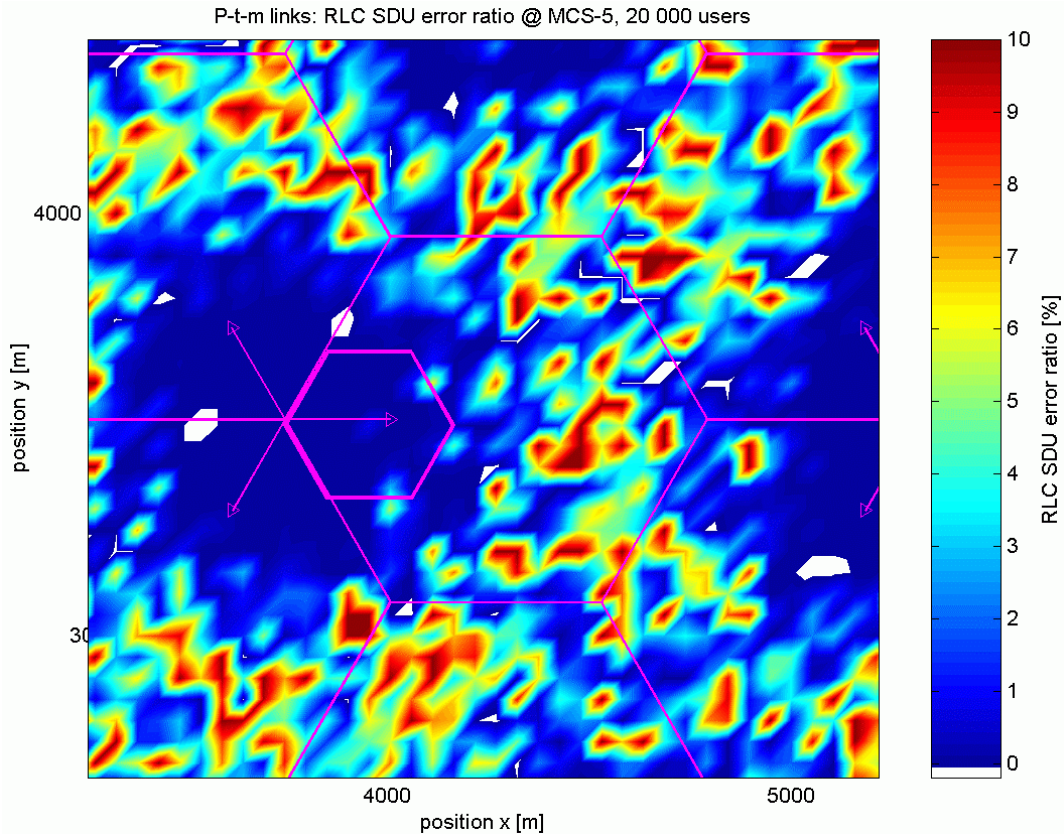


Figure 4.3: Detail of RLC SDU error ratio map shown in section 6.3. The small hexagon (drawn with thicker line) depicts the cell size in the hotspot network (200 m cell radius).

All simulation parameters such as BTS transmitting power and frequency reuse remained the same during simulations in the hotspot network.

4.4 Analysis of the Results

The simulation results are enclosed in Appendix 1 and analysed in this section. In section 6.1 there are connection throughputs (measured on RLC layer) and section 6.2 contains RLC SDU error ratios for the simulated scenarios. These parameters were chosen as performance indicators. As an additional criterion can be taken the network traffic load experienced in various scenarios, shown in Figure 4.4 and Figure 4.5. The network traffic load denotes the proportion of fully utilised GPRS channels to all existing GPRS channels in the network. For example 100% load would mean that all transceivers in the network are continually transmitting (in all timeslots).

The connection throughput is the total amount of bits in correctly received RLC/MAC data blocks divided by the period when the connection had a temporary block flow. The throughput is measured per timeslot, even for FTP connections, which were allowed to use up to three timeslots in downlink. The

¹ A geographical area in the cellular network with expected high number of users such as stadium.

RLC SDU error ratio is the number of erroneous RLC service data units divided by the number of all reconstructed RLC SDU delivered by RLC layer to upper layers.

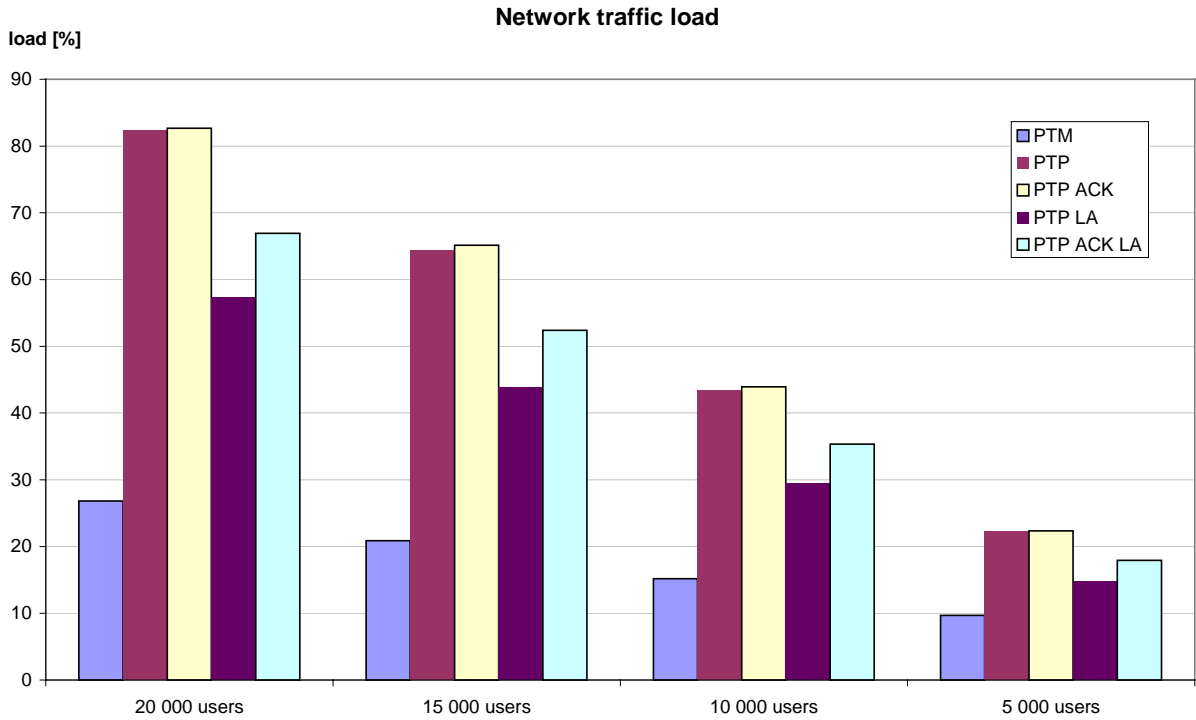


Figure 4.4: Normal network traffic load in different scenarios with different number of users

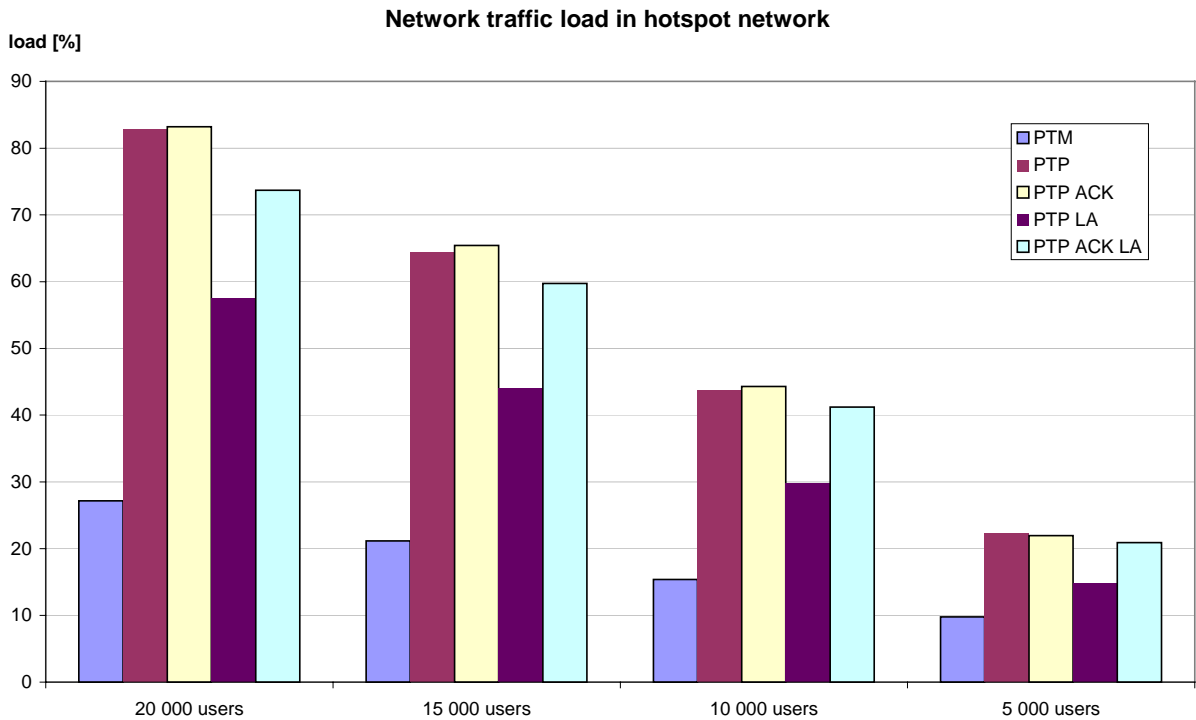


Figure 4.5: Hotspot network traffic load

As the results show, there are advantages as well as disadvantages in the usage of the proposed p-t-m links. Among the advantages are:

- The least impact on other users in the network, especially under heavier loads. The majority of background FTP connections in PTM scenario experience the highest throughput (see page 53).
- The same or higher throughput in comparison with other links with fixed modulation and coding scheme (neglecting high RLC SDU error ratio).
- Low sensitivity of performance on the number of users in the network, particularly when the majority of users uses p-t-m links, such as in hotspot areas.

Disadvantages of p-t-m links are:

- RLC layer cannot offer error free transmission when using p-t-m links.
- According to [3GP03c] the range of RLC SDU error ratio QoS attribute for the streaming class has upper limit 10 %. The simulation results show however that 99-percentile of RLC SDU error ratio (when whole coverage of the cell is considered) was around 20 %, which is therefore unacceptably high. Even in hotspot areas, where the cell radius is small, the RLC SDU error ratio remains over 10 % for 99 % of users.
- Only 10 % and 1 % RLC SDU error ratios (out of those values defined for this QoS attribute in [3GP03c]) can be sustained in part of the cell area. According to the results obtained for the PTM scenario shown in section 6.2, RLC SDU error ratio of 10 % or less is experienced on average by 94 % of the population of users and 1 % or less by about 58 % of users. The users are generated at random locations and move in random directions. From that the conclusion is drawn that the estimated cell coverage with 10 % RLC SDU error ratio is about 94 % of the cell area and with 1 % it is about 58 % (for this particular network layout). See also section 6.3 for a map showing RLC SDU error rate depending on the mobile station position in the network.

Achieved throughput gain in comparison to different p-t-p scenarios is summarised in Figure 4.6 and in Figure 4.7 (small network). The series correspond to the respective throughput percentile in results.

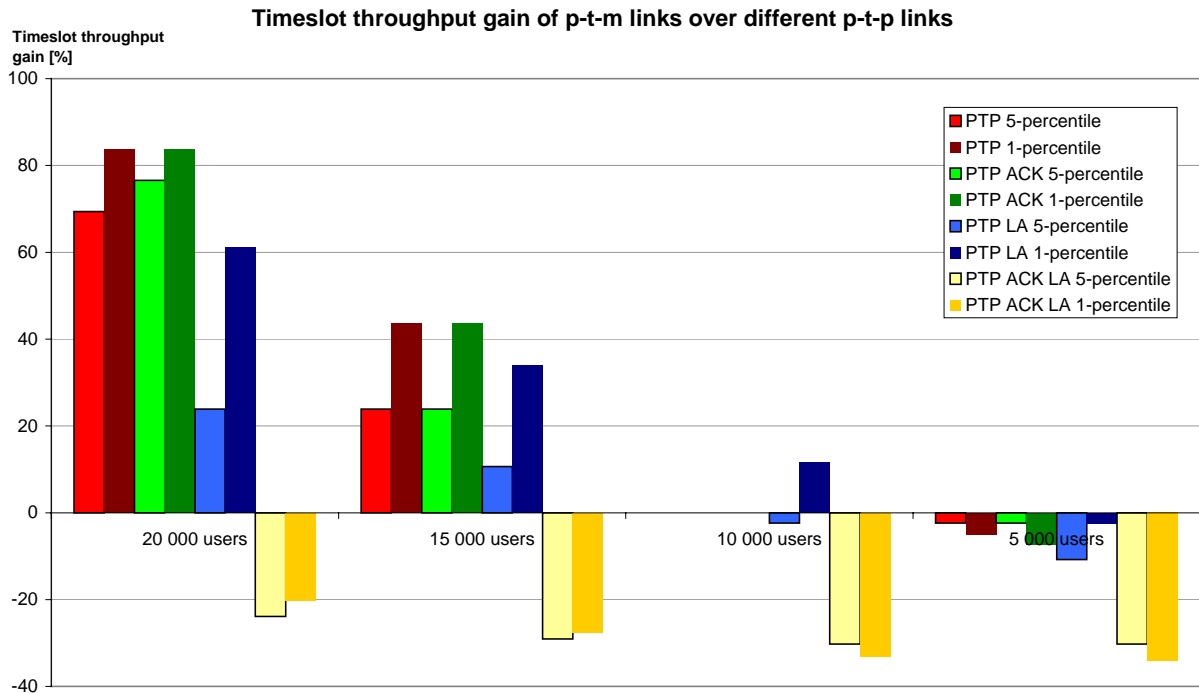


Figure 4.6: Achieved timeslot throughput gain in normal network

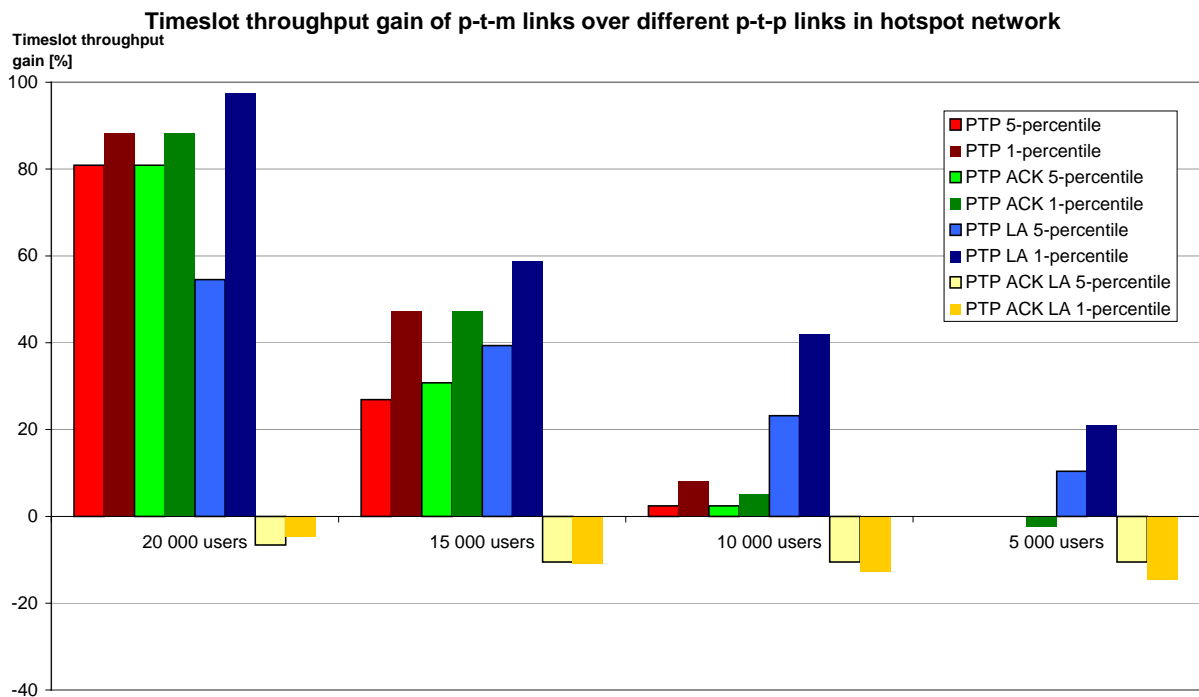


Figure 4.7: Achieved timeslot throughput gain in hotspot network

The throughput gain g_t was calculated as the difference between PTM throughput r_{PTM} and the other link type throughput r relative to r :

$$g_t = \frac{r_{PTM} - r}{r} \cdot 100$$

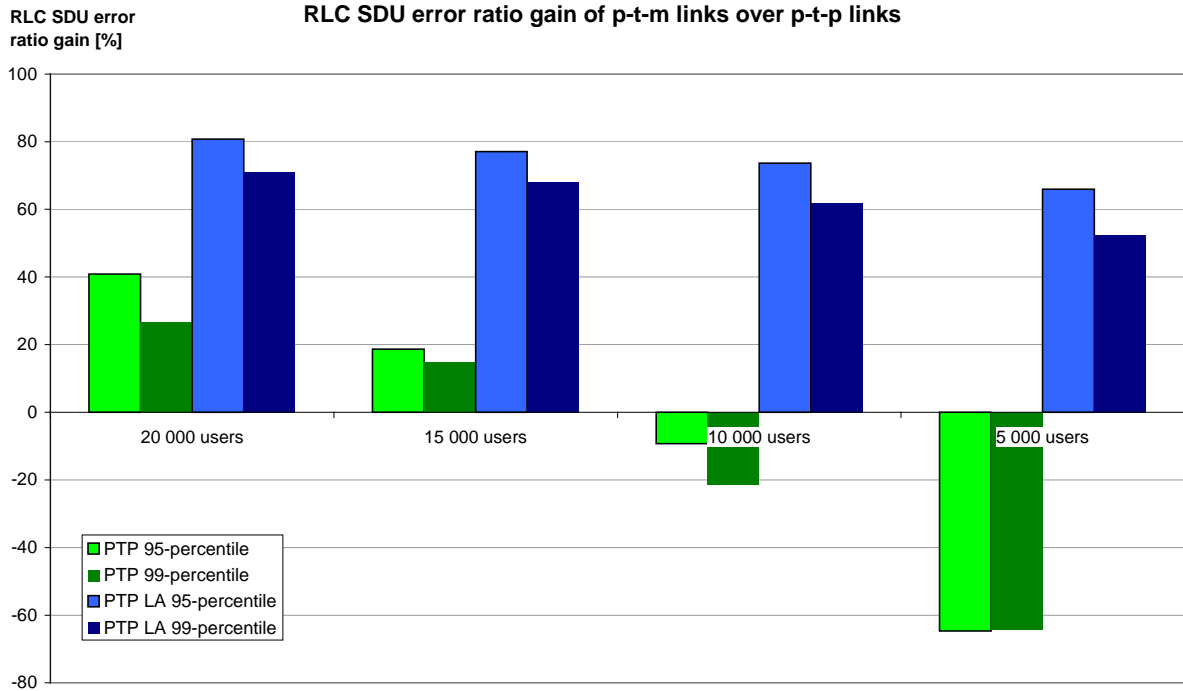


Figure 4.8: Achieved RLC SDU error ratio gain in normal network

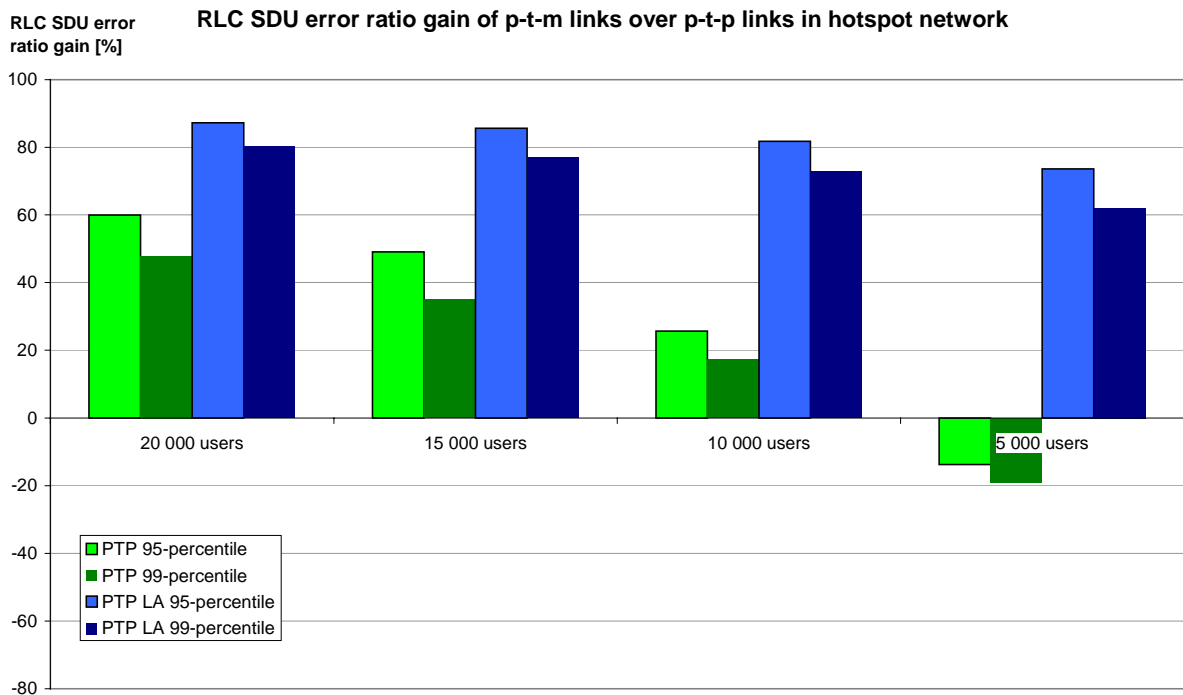


Figure 4.9: Achieved RLC SDU error ratio gain in hotspot network

Similarly, the achieved RLC SDU error ratio gain g_e is summarised in Figure 4.8 and in Figure 4.9 (small network). It was calculated as the difference between PTM RLC SDU error ratio q_{PTM} and the

other link type RLC SDU error ratio q relative to q :

$$g_e = \frac{q - q_{PTM}}{q} \cdot 100$$

In simulations in the hotspot network the SDU error ratio decreased to about 14% for 99% of users (as can be seen in the RLC SDU error ratio chart in section 6.2). This value is still over the required limit of 10% therefore the decrease of cell radius is not very effective means of lowering the RLC SDU error ratio. However, with decreasing cell radius the gains of p-t-m transmission increased both in throughput (Figure 4.7) and in RLC SDU error ratio (Figure 4.9).

4.4.1 Conclusions

The results show that in the chosen scenarios the usage of p-t-m links (PTM scenario) leads to less traffic load in the network, which leaves more radio resources for other types of traffic. Consequently the connections of other services experience higher throughput as can be seen on the example of FTP connections.

With regards to the performance of p-t-m links using MCS-5, the conclusion can be drawn that they are 20 to 34 percent¹ worse in throughput than p-t-p links using RLC acknowledged mode and link adaptation (PTP ACK LA scenario). P-t-m links also have about 21 percent RLC SDU error ratio, which almost does not change with increasing network load. As the map (in section 6.3) shows the RLC SDU error ratio is lower when the mobile station is closer to the base station.

The high RLC SDU error ratio experienced by p-t-m listeners is mainly caused by strong interfering signal from the first tier of interfering p-t-m sources. This conclusion can be drawn from the observed independence of RLC SDU error ratio on the number of users in the network. The number of p-t-m sources was constant in all load variants and they were transmitting with the same power.

Some complementary means to decrease the RLC SDU error ratio under acceptable limit are presented below:

- Using more robust modulation and coding scheme, such as CS-1 and MCS-1. Note that this would however increase the level of segmentation and decrease the maximum throughput that can be achieved per timeslot compared to MCS-5.
- Repeating the same radio blocks two or more times. As with the preceding solution, this considerably lowers the achievable throughput. However, the level of segmentation is kept the same as with a non-repetition scheme.
- Using outer coding to provide additional redundancy and increase of forward error correction capability. One suggested solution using Reed-Solomon codes, was described in section 3.5.1.
- Introducing feedback from p-t-m listeners. Proposal of such solution was briefly described in subsection 3.5.2.

¹ Taking 1-percentile values into consideration.

5. DISCUSSIONS

This study aimed at investigating and evaluating MBMS in GSM/EDGE radio access network. The investigations were based on existing, not yet finalised specifications produced by 3GPP and also on documents submitted to meetings of various workgroups of 3GPP. The requirements of MBMS and ideas about its functionality were the source for the simulation model used to evaluate MBMS in simulated environment. Existing GSM/EDGE network level simulator was extended to support p-t-m type of connections. This modified simulator was then used to produce series of results for five scenarios in two cellular networks. One with 500 m cell radius was modelling normal network while the other, with 200 m cell radius, was modelling cellular network in a hotspot area.

In each simulation scenario two services were simulated: FTP background file transfer and file streaming. The file streaming service was modelling MBMS broadcast session. The scenarios differed in type of radio links used to simulate the file streaming service. While the FTP file transfer service was always using p-t-p links, for streaming service the following types of links were used:

- PTM scenario: p-t-m links and MCS-5,
- PTP scenario: p-t-p links, unacknowledged RLC mode and MCS-5,
- PTP ACK scenario: p-t-p links, acknowledged RLC mode and MCS-5,
- PTP LA scenario: p-t-p links, unacknowledged RLC mode with link adaptation and
- PTP ACK LA scenario: p-t-p links, acknowledged RLC mode with link adaptation.

According to the results, the p-t-m links had the least impact on the background FTP file transfer service as expected. Also the distribution of throughput per connection was very good, given the fixed modulation and coding scheme. Unfortunately the RLC SDU error ratio was too high. Decreasing the cell radius in the hotspot network did not help much to reduce the error ratio. Therefore additional mechanisms such as repetition of RLC PDUs, outer coding using Reed-Solomon codes or sending negative acknowledgements in uplink need to be studied to see if with their inclusion the p-t-m links can fulfil the QoS requirements for streaming and background traffic class.

REFERENCES

- [Com00] COMER, D.: *Internetworking with TCP/IP, Volume 1: Principles, Protocols and Architecture*. 4th Edition, Upper Saddle River, New Jersey, Prentice-Hall, 2000, Chapter 17 (pp. 319–351).
- [Dee89] DEERING, S.: *Host Extensions for IP Multicasting*. RFC 1112, 1989, URL: <http://www.rfc-editor.org/rfc/rfc1112.txt> (April 2004).
- [Hal02] HALONEN, T., ROMERO, J., MELERO, J.: *GSM, GPRS and EDGE Performance*. Chichester, Great Britain, John Wiley & Sons, 2002
- [Hei00] HEINE, G.: *GPRS from A–Z*. London, Great Britain, Artech House, 2000
- [Jer00] JERUCHIM, M., BALABAN, P., SHANMUGAN, K.: *Simulation of Communication Systems*. Second edition. New York, U.S.A., Kluwer Academic/Plenum Publishers, 2000, Chapter 1
- [Kaa01] KAARANEN, H., AHTIAINEN, A., LAITINEN, L., NAGHIAN, S., NIEMI, V.: *UMTS Networks, Architecture, Mobility and Services*. Chichester, Great Britain, John Wiley & Sons, 2001
- [Mou92] MOULY, M., PAUTET, M-B.: *The GSM System for Mobile Communications*. Palaiseau, France, Cell & Sys, 1992
- [Séb02] SÉBIRE, B. ET AL.: *GSM/EDGE Radio Access Network: Evolution of GSM/EDGE Toward 3G Mobile Services*. In *Advances in 3G Enhanced Technologies for Wireless Communications*, editors: Jiangzhou Wang, Tung-Sang Ng. Artech House, 2002
- [Vai02] VAINIO, A.: *Broadcast and Multicast in UMTS Networks*. [Master’s Thesis] Helsinki University of Technology, Department of Computer Science and Engineering, Espoo, Finland, 2002

3GPP Specific References

- [3GP02a] 3GPP: *TS 22.060 General Packet Radio Service (GPRS); Service description, Stage 1 (Release 5)*. Version 5.2.0, 2002, URL: http://www.3gpp.org/ftp/Specs/archive/22_series/22.060/22060-520.zip (April 2004).
- [3GP02b] 3GPP: *TS 22.146 Multimedia Broadcast/Multicast Service; Stage 1 (Release 6)*. Version 6.1.0, 2002, URL: http://www.3gpp.org/ftp/Specs/archive/22_series/22.146/22146-610.zip (April 2004).
- [3GP02c] 3GPP: *TR 23.846 Multimedia Broadcast/Multicast Service; Architecture and Functional Description (Release 6)*. Version 2.0.0, 2002, URL: http://www.3gpp.org/ftp/Specs/archive/23_series/23.846/23846-200.zip (April 2004).

- [3GP02d] 3GPP: *TR 25.992 Multimedia Broadcast/Multicast Service (MBMS); UTRAN/GERAN Requirements*. Version 1.1.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/25_series/25.992/25992-110.zip (April 2004).
- [3GP02e] 3GPP: *TS 44.018 Mobile radio interface layer 3 specification; RRC protocol (Release 6)*. Version 6.0.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/44_series/44.018/44018-600.zip (April 2004).
- [3GP02f] 3GPP: *TS 44.060 General Packet Radio Service (GPRS); Mobile Station (MS) – Base Station System (BSS) interface; RLC/MAC protocol (Release 6)*. Version 6.0.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/44_series/44.060/44060-600.zip (April 2004).
- [3GP02g] 3GPP: *TS 23.060 General Packet Radio Service (GPRS); Service description, Stage 2 (Release 5)*. Version 5.4.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/23_series/23.060/23060-540.zip (April 2004).
- [3GP02h] 3GPP: *TS 43.064 General Packet Radio Service (GPRS); Overall description of the GPRS radio interface, Stage 2 (Release 5)*. Version 5.0.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/43_series/43.064/43064-500.zip (April 2004).
- [3GP02i] 3GPP: *TS 45.002 Multiplexing and Multiple Access on the Radio Path (Release 5)*. Version 5.7.0, 2002, URL:
http://www.3gpp.org/ftp/Specs/archive/45_series/45.002/45002-570.zip (April 2004).
- [3GP03a] 3GPP: *3GPP Home Page*. URL: <http://www.3gpp.org> (April 2004).
- [3GP03b] 3GPP: *TS 25.346 Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (Stage 2); (Release 6)*. Version 2.1.0, 2003, URL:
http://www.3gpp.org/ftp/Specs/archive/25_series/25.346/25346-210.zip (April 2004).
- [3GP03c] 3GPP: *TS 23.107 Quality of Service (QoS) Concept and Architecture; (Release 5)*. Version 5.10.0, 2003, URL:
http://www.3gpp.org/ftp/Specs/archive/23_series/23.107/23107-5a0.zip (April 2004).
- [3GP03d] 3GPP: *TR ab.cde DRAFT Feasibility Study on Single Antenna Interference Cancellation (SAIC) for GSM Networks; (Release 6)*. Version x.y.z, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_17_Budapest/Docs/GP-032675.zip (April 2004).

- [3GP03e] 3GPP: *TS 43.051 Radio Access Network; Overall description - Stage 2; (Release 5)*. Version 5.10.1, 2003, URL:
http://www.3gpp.org/ftp/Specs/archive/23_series/43.051/43051-5a0.zip (April 2004).
- [3GP03f] 3GPP: *TS 43.246 Multimedia Broadcast Multicast Service (MBMS) in the GERAN; Stage 2; (Release 6)*. Version 0.6.0, 2003, URL:
http://www.3gpp.org/ftp/Specs/archive/43_series/43.246/43246-060.zip (April 2004).
- [3GP03g] 3GPP: *TS 23.246 Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description (Release 6)*. Version 6.0.0, 2003, URL:
http://www.3gpp.org/ftp/Specs/archive/23_series/23.246/23246-600.zip (April 2004).
- [Eri03] ERICSSON: *Tdoc TSG GERAN GP-032024: Solutions for MBMS Notification*. In 3GPP TSG GERAN #16, New York, USA, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_16_NewYork/Docs/GP-032024.zip (April 2004).
- [Nok03a] NOKIA: *Tdoc TSG GERAN GP-030816: MBMS Notification*. In 3GPP TSG GERAN #14, Munich, Germany, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_14_Munich/Docs/GP-030816.zip (April 2004).
- [Nok03b] NOKIA: *Tdoc TSG GERAN GP-030817: Radio Resource Optimisation During the MBMS Session*. In 3GPP TSG GERAN #14, Munich, Germany, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_14_Munich/Docs/GP-030817.zip (April 2004).
- [Nok03c] NOKIA: *Tdoc TSG GERAN GP-030199: MBMS Data Transfer in GERAN*. In 3GPP TSG GERAN #13, San Antonio, USA, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_13_San_Antonio/Docs/GP-030199.zip (April 2004).
- [Qua03] QUALCOMM EUROPE: *Tdoc GERAN WG2 G2-030134: User counting in MBMS*. In 3GPP TSG GERAN2 #13bis, Winchester, United Kingdom, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/Wg2_Protocol_Aspects/GERAN2_13bis_Winchester/Docs/G2-030134.zip (April 2004).
- [Sie03a] SIEMENS: *Tdoc TSG GERAN GP-031392: Further considerations on the use of the PNCH for MBMS notification*. In 3GPP TSG GERAN #15, Fort Lauderdale, USA, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_15_Fort_Lauderdale/Docs/GP-031392.zip (April 2004).
- [Sie03b] SIEMENS: *Tdoc TSG GERAN GP-031994: Outer Coding for MBMS*. In 3GPP TSG GERAN #16, New York, USA, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_16_New_York/Docs/GP-031994.zip (April 2004).

- [Sie03c] SIEMENS: *Tdoc TSG GERAN GP-032100: Outer Coding on RLC Layer for MBMS over GERAN*. In 3GPP TSG GERAN #16, New York, USA, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_16_New_York/Docs/GP-032100.zip (April 2004).
- [Sie03d] SIEMENS: *Tdoc TSG GERAN GP-032604: P-t-m with Feedback for MBMS Delivery*. In 3GPP TSG GERAN #17, Budapest, Hungary, 2003, URL:
http://www.3gpp.org/ftp/tsg_geran/TSG_GERAN/GERAN_17_Budapest/Docs/GP-032604.zip (April 2004).

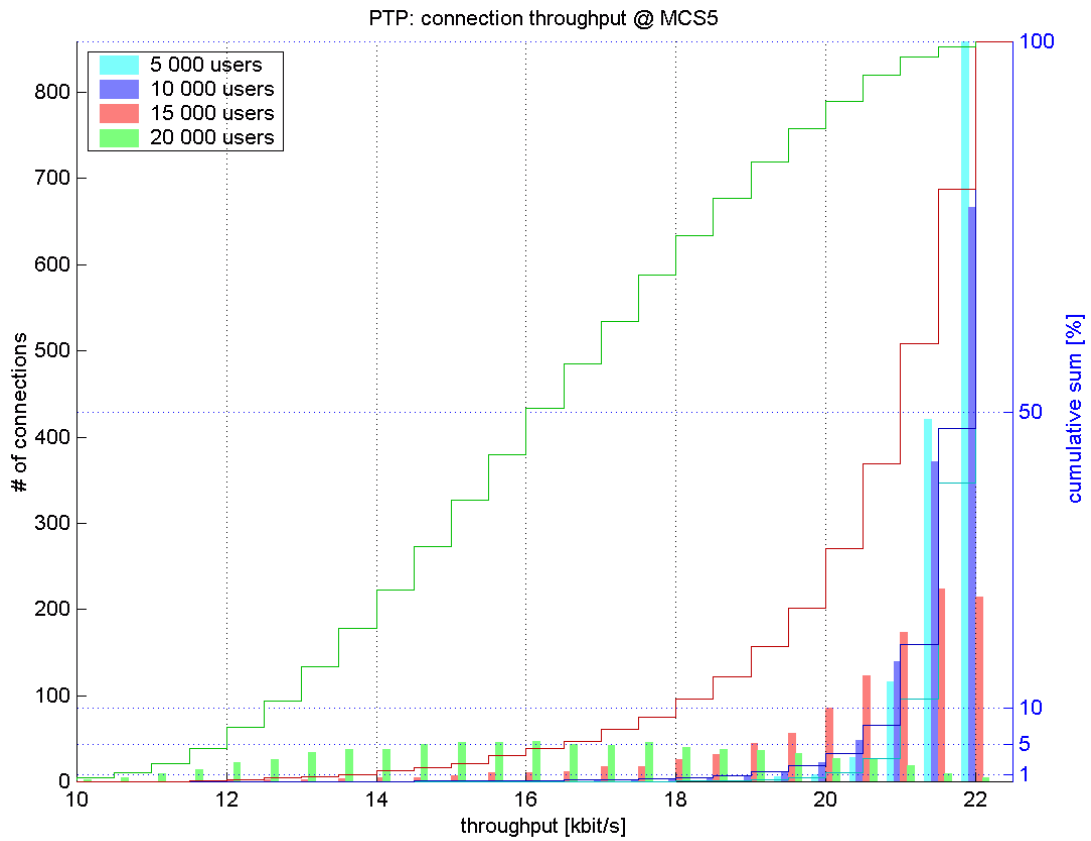
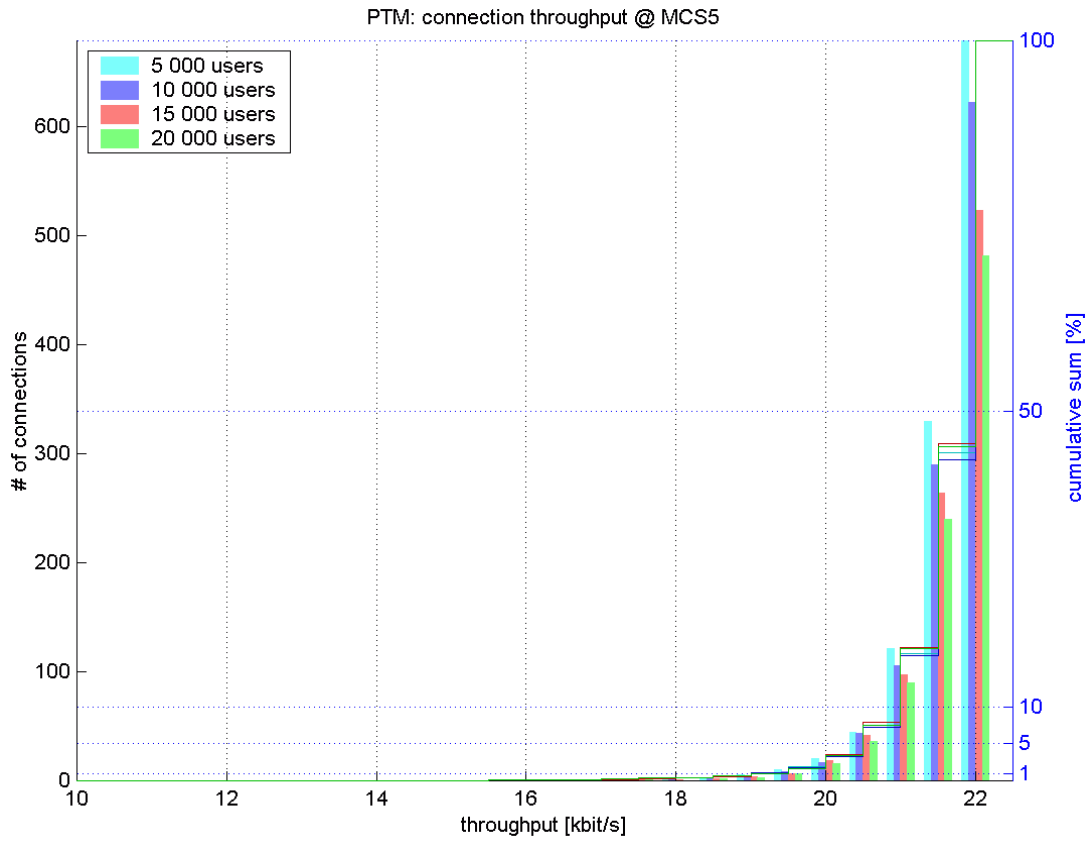
6. APPENDIX 1 – SIMULATION RESULTS

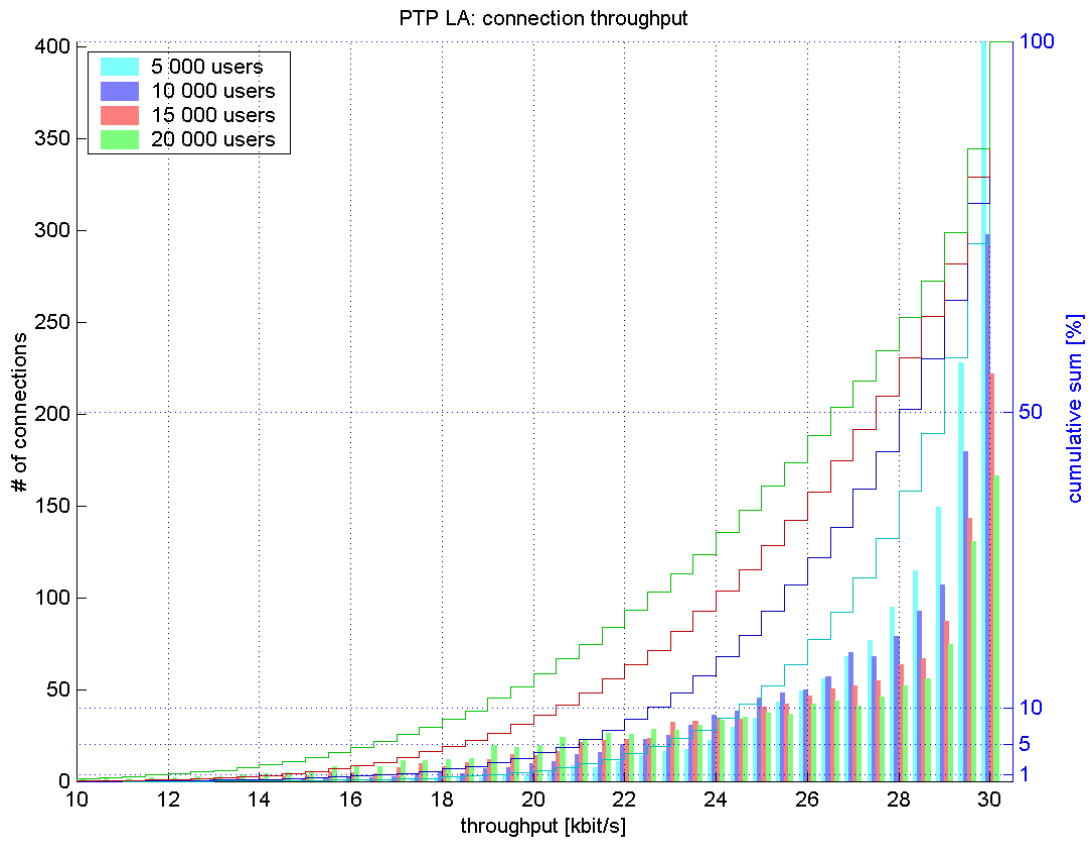
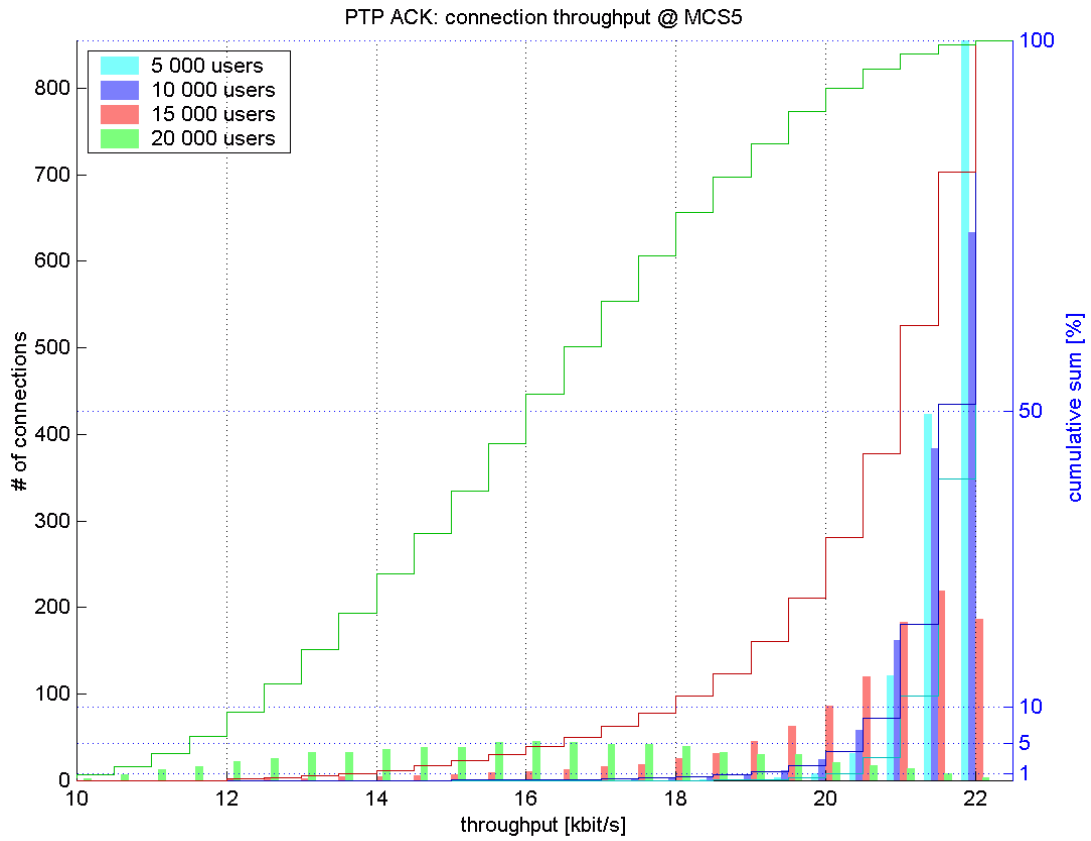
In this appendix the results of the simulations are collected and presented as charts. The first section deals with achieved connection throughput, while the second shows the RLC SDU error ratios.

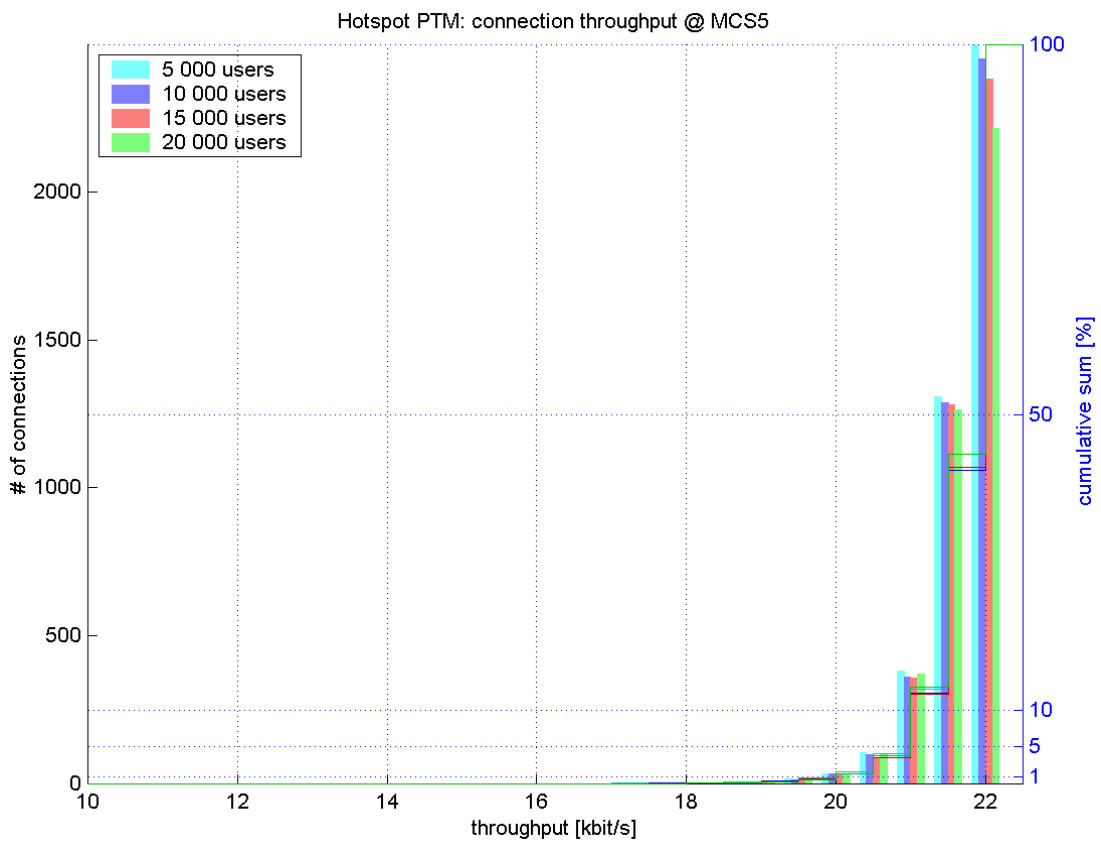
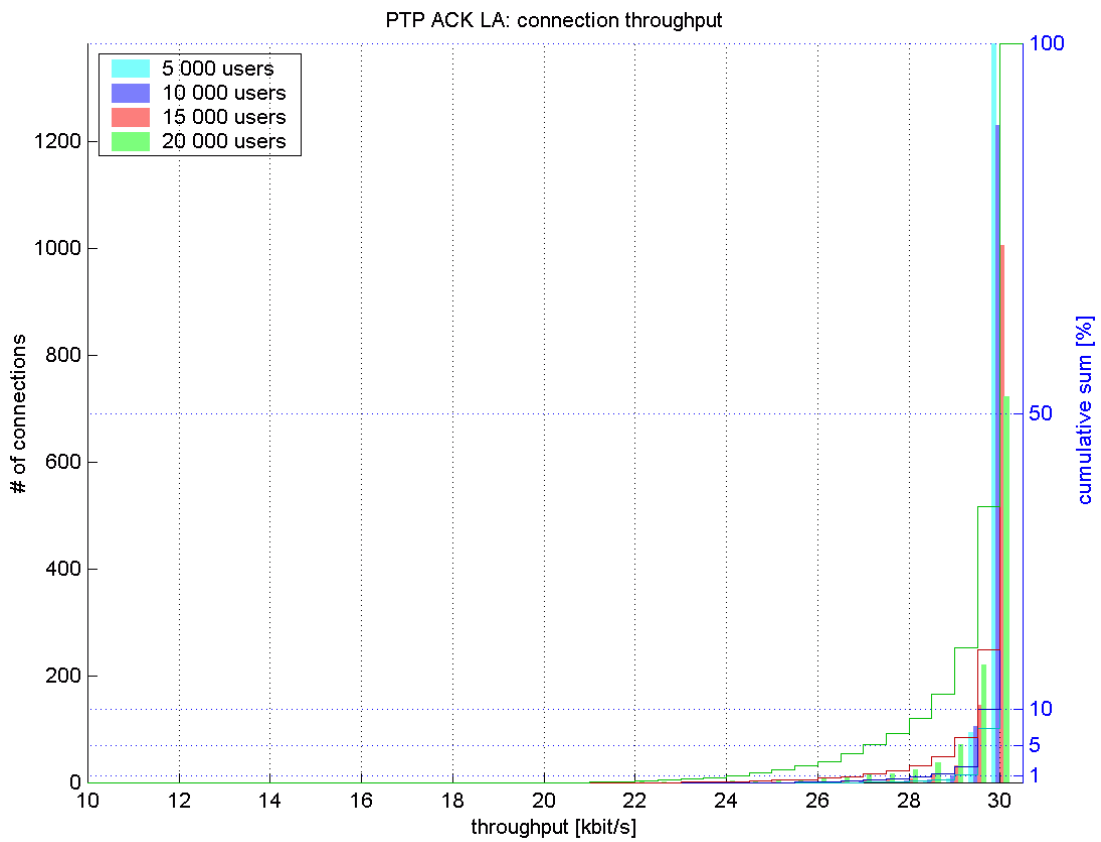
6.1 Connection Throughput

The charts in this section show the connection throughput for all scenarios under the four loads. The labels in legends show the number of users in thousands and the simulated scenario. The histograms show throughputs only for the streaming service. The impact on FTP service (background data calls) is shown on percentile charts, which follow the histograms.

Please note that in cases PTP LA and PTP ACK LA the x axis scale is up to 30 kbit/s.



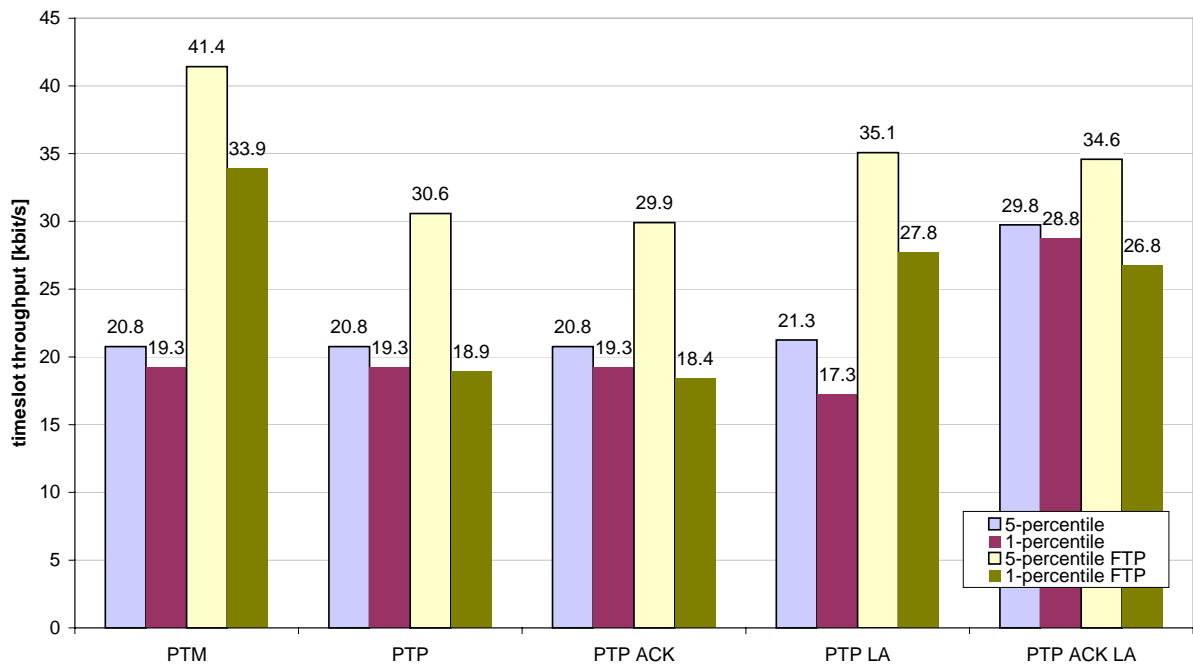




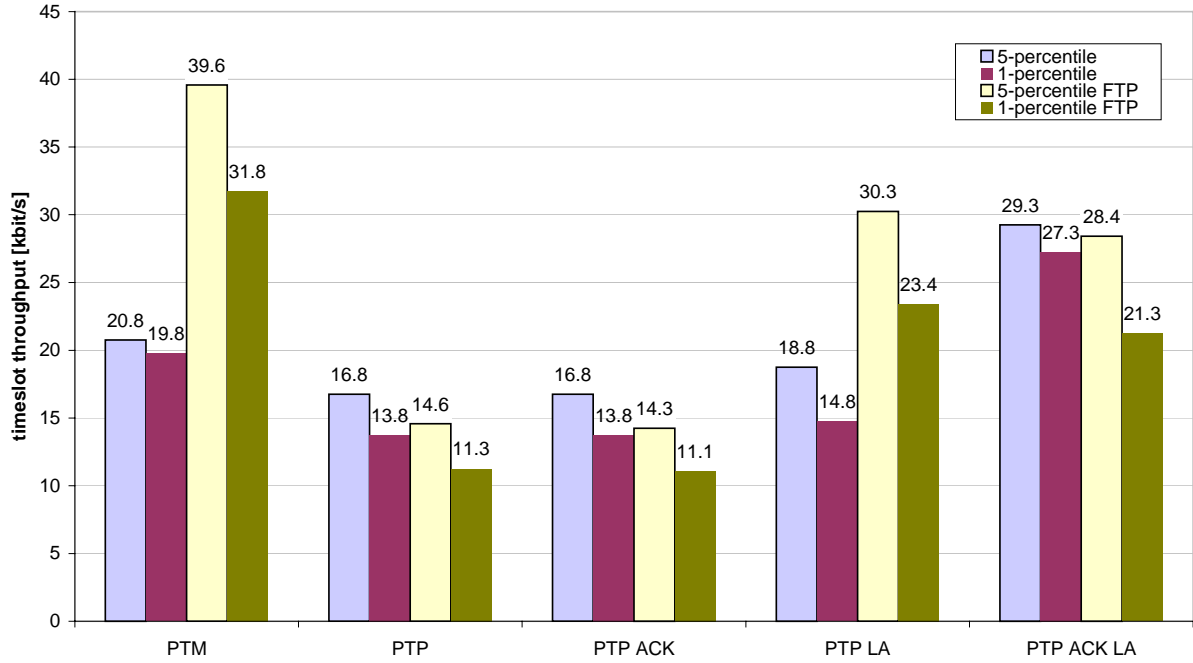
5 000 users: connection throughput per timeslot @ MCS5 (except LA cases and FTP)



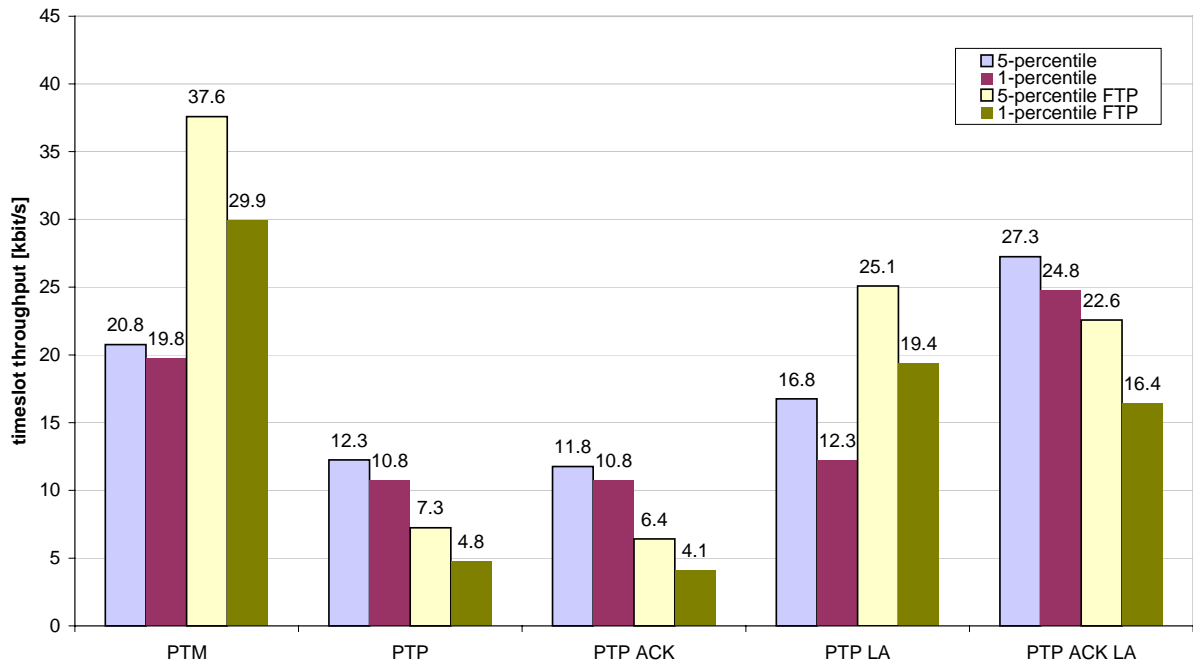
10 000 users: connection throughput per timeslot @ MCS5 (except LA cases and FTP)



15 000 users: connection throughput per timeslot @ MCS5 (except LA cases and FTP)

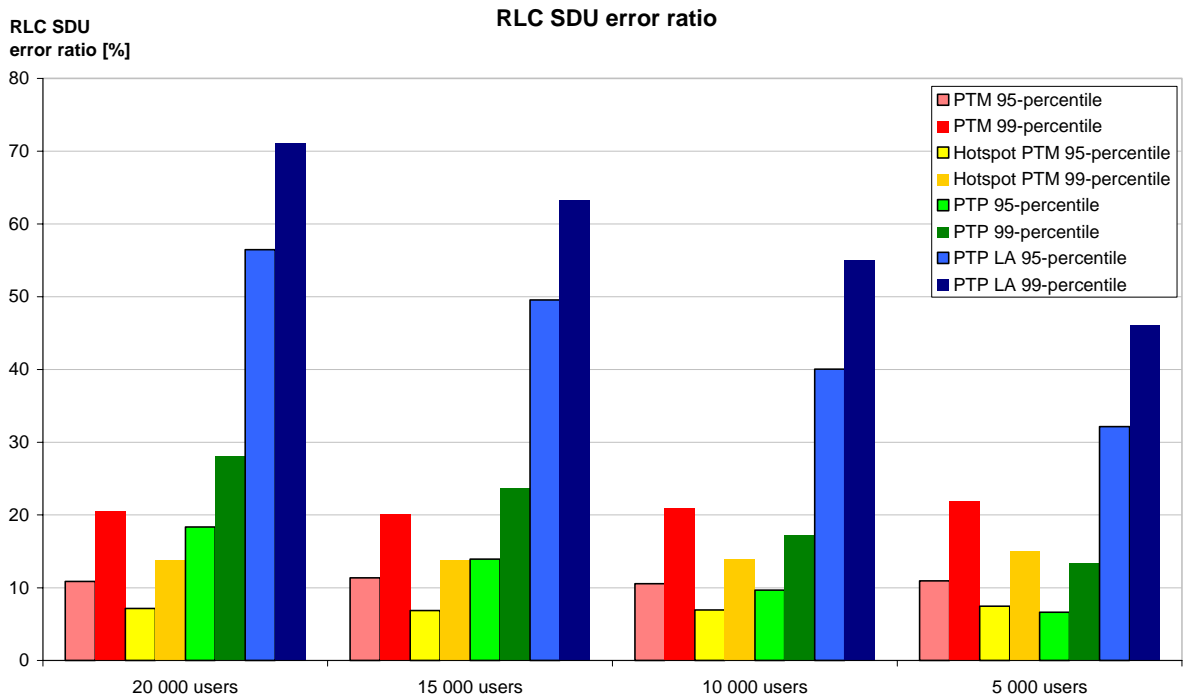


20 000 users: connection throughput per timeslot @ MCS5 (except LA cases and FTP)



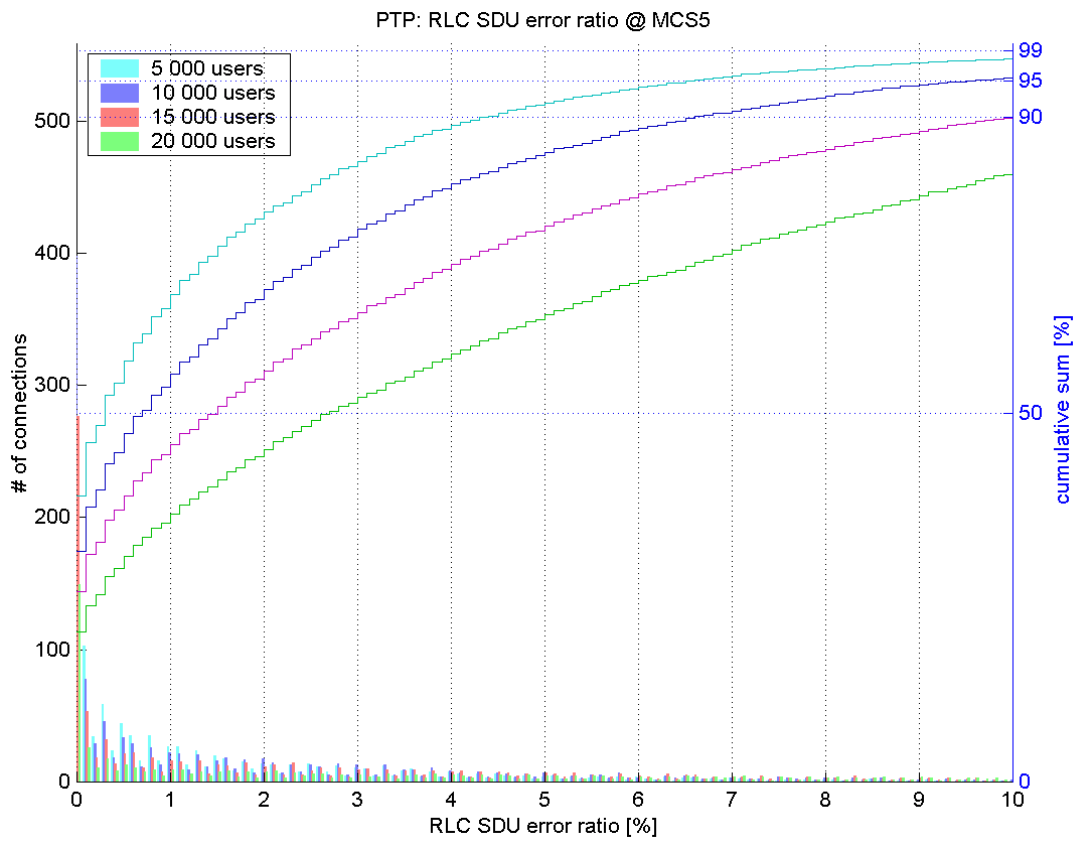
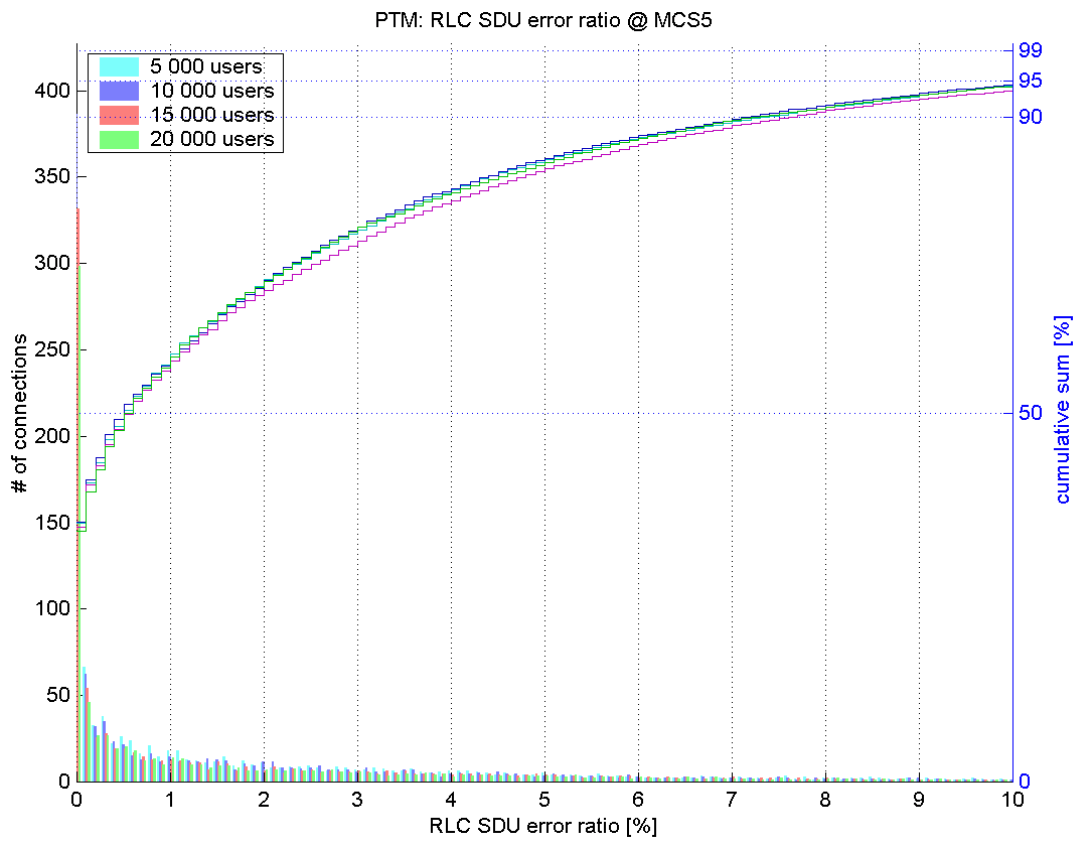
6.2 RLC SDU Error Ratio

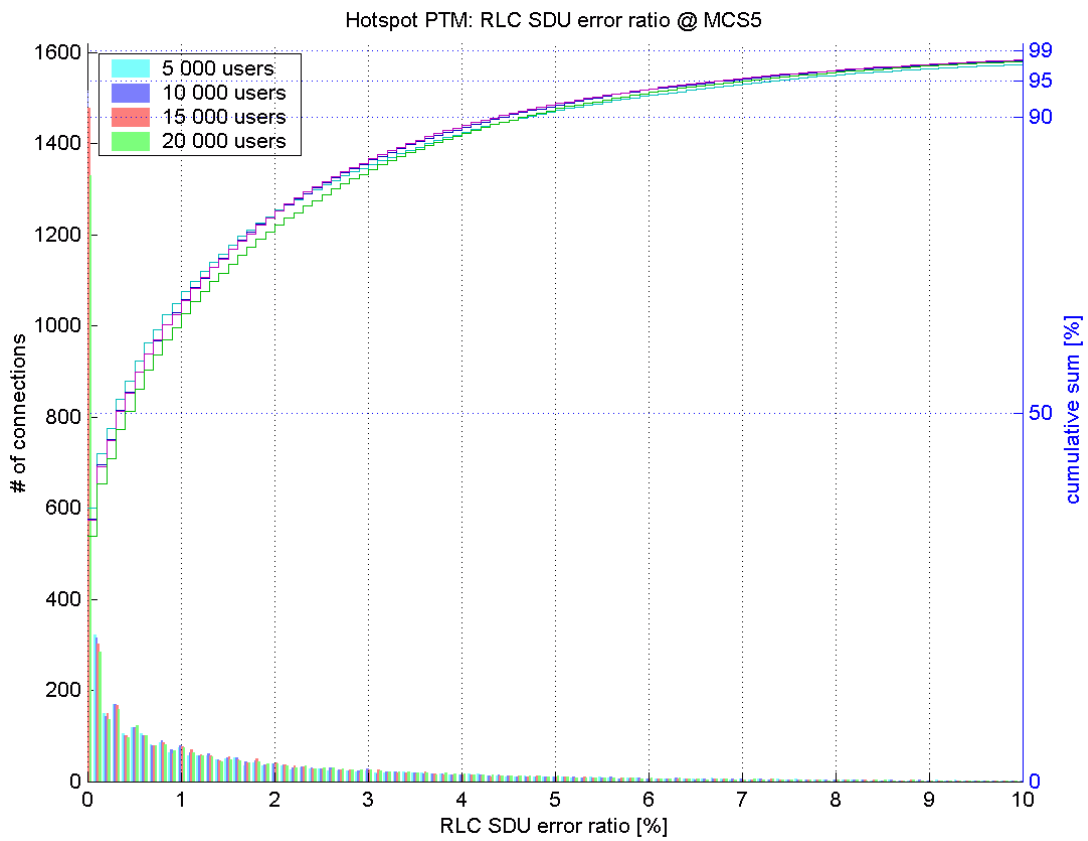
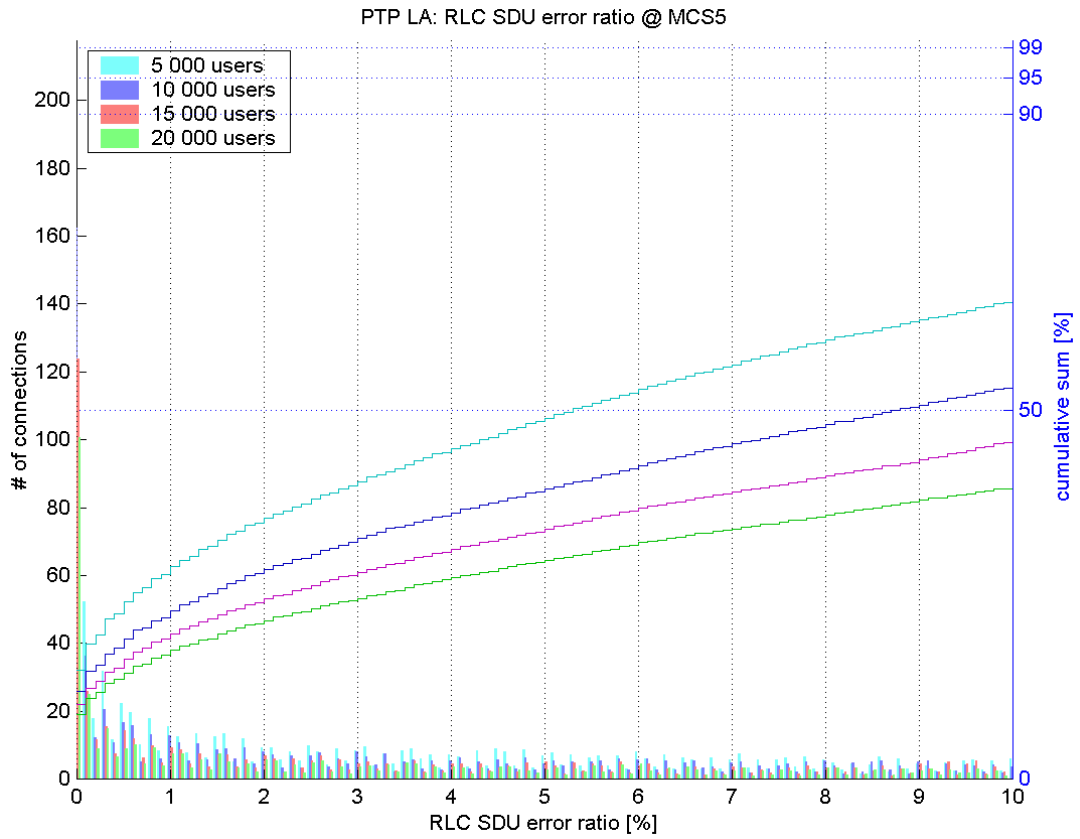
The RLC SDU error ratio is only shown for PTM, PTP and PTP LA scenarios because in PTP ACK and PTP ACK LA scenarios the error ratio is zero¹ due to use of RLC acknowledged mode (residual errors are not considered in the simulator).



Note that the PTP and PTP LA scenarios used unacknowledged RLC mode, therefore especially the p-t-p links in PTP LA scenario performed so badly.

¹ According to [Hal02] the residual radio RLC/MAC data block error rate in RLC acknowledged mode is $2 \cdot 10^{-4}$ (for EGPRS connections).





6.3 RLC SDU Error Ratio Map For PTM Scenario

The following maps show RLC SDU error ratio experienced by mobile stations in PTM scenario depending on their position in the cell. Only maps for 20000 users in the network are shown as examples because the situation with the other loads is similar. The arrows indicate the direction of the transceiver antennae. The scattered white spots are areas without collected values.

