Technische Universität Wien

Diplomarbeit

GPRS Performance Evaluation

ausgeführt am

Institut für
Nachrichtentechnik und Hochfrequenztechnik
der Technischen Universität Wien

unter Anleitung von

O.Univ.Prof. Dr. techn. Ernst Bonek

durch

Kurt Wagentristl
A-1100 Wien, Maria-Kuhn-Gasse 6/3/5

Wien, im April 2004 ______________________

Betreuer:

O.Univ.Prof. Dr. techn. Ernst Bonek

Zusammenfassung

GPRS, das *General Packet Radio Service*, ist dafür ausgelegt, Daten paketvermittelt über GSM Netze zu übertragen. Die Ressourcen der Funkschnittstelle werden dabei zwischen Sprachverkehr und Datenübertragung geteilt. Die theoretisch maximale Datenübertragungsrate für den Teilnehmer beträgt 171.2kbit/s, abhängig von der Anzahl der zugeteilten Übertragungskanäle (Zeitschlitze), den Interferenzbedingungen an der Funkschnittstelle und aber auch von der Fähigkeit der MS (Mobilstation), mehrere Übertragungskanäle gleichzeitig zur Datenübertragung zu verwenden.

Die GPRS Leistungswerte sind hauptsächlich von den Übertragungsprotokollen der Funkschnittstelle abhängig, nämlich *Radio Link Control* (RLC), *Logical Link Control* (LLC), und *Subnetwork Dependent Convergence Protocol* (SNDCP). Weiters hat das *Base Station System GPRS Protocol* (BSSGP), die höchste Protokollinstanz an der *Gb-Schnittstelle*, massgeblichen Einfluss auf die GPRS Leistungsfähigkeit. Die *Gb-Schnittstelle* verbindet den SGSN (*Serving GPRS Support Node*) mit dem BSC (*Base Station Controller*).

Im speziellen limitieren die hohen Ansprechzeiten an der Funkschnittestelle die Leistung interaktiver Anwendungen. Die starre Struktur des Zeitmultiplexverfahrens (*TDMA, Time Division Multiple Access*) bewirkt variable Wartezeiten, bis bestimmte Kanalzuteilungsnachrichten an der Funkschnittstelle übertragen werden können. Ausserdem schlägt zusätzlich die Übertragungszeit über die Funkschnittstelle aufgrund der begrenzten Bandbreite zu Buche. Zwar kann die Übertragungsbandbreite durch Zuteilung mehrerer TDMA-Zeitschlitze vervielfacht werden, allerdings ist diese Betriebsart durch die MS begrenzt. Die zur Zeit verfügbaren Mobiltelefone unterstützen vier Zeitschlitze in der Abwärtsstrecke (Netz zu MS) und drei Zeitschlitze in der Aufwärtsstrecke (MS zu Netz). Eine Erhöhung der Anzahl der Zeitschlitze pro MS ist in GPRS nicht möglich, dazu ist es notwendig *Enhanced GPRS (EGPRS)* zu implementieren. EGPRS ist die paketvermittelnde Komponente von *EDGE (Enhanced Data Rates for GSM Evolution)*. Die Bandbreite pro MS kann auch durch die hochdatenratigen Kanal-Codierschemata *CS-3* und *CS-4* erhöht werden. CS-3, CS-4 kann die GPRS Leistungswerte deutlich steigern, allerdings nur für stationäre GPRS-Anwender und nur wenn die Funkschnittstelle interferenzfrei ist. Für bewegte MS ist die Verwendbarkeit von CS-4 auf den Nahbereich der versorgenden Basisstation beschränkt, wo die Sig-

nalstärke entsprechend hoch ist und Sichtverbindung zwischen MS und Basisstation besteht.
Mobile Anwender müssen aber noch weitere Nachteile in Kauf nehmen. Ein Wechsel der versorgenden Zelle und die Änderung des Zustellpfades der Datenpakete geht manchmal mit Datenverlust einher. Wenn der Zellwechsel innerhalb einer BSS stattfindet, können Datenpakete aus der ursprünglichen Zelle in die neue Zelle ohne Datenverlust transferiert werden. Wenn hingegen die neue Zelle zu einer anderen BSS gehört, ist Datenverlust unvermeidbar. Ein weiteres Protokoll, dass die GPRS Leistungsmerkmale bestimmt, ist BSSGP, das die Flusskontrolle an der Gb-Schnittstelle bereitstellt. Die Flusskontrolle regelt den Datenfluss zwischen SGSN und PCU in der Abwärtsstrecke. Der aktuelle Pufferstand in der PCU wird aus einer begrenzten Anzahl aus Nachrichtenpaketen von der PCU berechnet. Pufferstandberechnungen, die vom Istwert abweichen, können einerseits zu Pufferüberlauf in der PCU führen, oder zu reduziertem Datendurchsatz, wenn nicht genügend Daten von der SGSN bereitgestellt werden.
Die anderen GPRS Protokollinstanzen, wie *GTP (GPRS Tunneling Protocol)* aus dem GPRS *Backbone Netzwerk*, ermöglichen keine Leistungssteigerung, abgesehen von ausreichender Dimensionierung.
Das Verhalten von Internet-Anwendungen und deren Übertragungsprotokolle, wie zum Beispiel *TCP/IP* (Transport Control Protocol, Internet Protocol) oder *UDP/IP* (User Data Protocol), zeigen sehr unterschiedliches Verhalten in GPRS-Netzen. Das meistverbreitete Transportprotokoll, TCP, muss in möglichst aktueller Version implementiert sein, um hohe Durchsatzraten trotz hoher Latenzzeiten und Datenverluste im Funkfeld gewährleisten zu können. Verbesserte TCP Implementierungen können TCP Leistungsmerkmale steigern, eine generelle Verbesserung von Internetanwendungen über GPRS ist aber nur möglich, wenn *Performance Enhancing Proxies* (PEP) zum Einsatz kommen. Die üblichen Konzepte von PEPs optimieren die Internetprotokolle für den Einsatz in Mobilfunknetzen, implementieren Daten- oder Protokollkompression, oder implementieren Verbesserungen auf Applikationsebene. Dazu zählen Zwischenspeichern von Webseiten mit häufigen Zugriffen oder Veränderung des Inhalts der Webseite, wie z.B. Entfernen von Skriptelementen. Der Einsatz eines leistungssteigernden Proxies ist essentiell für GPRS Netze, will man die beschränkte Bandbreite und hohe Latenzzeiten umgehen.
Moderne GPSR Netze bieten maximale Datenübertragungsraten von 81.6kbit/s in der Abwärtsstrecke und 61.2kbit/s in der Aufwärtsstrecke, was zwar weit unter den die theoretischen 171.2kbit/s ist, allerdings deutlich über den Werten von ISDN Basisanschlüssen liegt. Im Allgemeinen sind auch die GPRS Leistungmerkmale besser als bei analoger, drahtgebundener Datenübertragung.

Abstract

GPRS, the General Packet Radio Service, is designed to transmit packet data via GSM networks by sharing the same radio interface resources. The maximum theoretical data rate per mobile client is 171.2kbit/s, depending on the number of channels assigned by the network, the interference conditions in the radio path but also on the mobile station's (MS) multislot capabilities.

I demonstrate that GPRS performance is predominantly determined by the protocols acting over the radio link, namely *Radio Link Control* (RLC), *Logical Link Control* (LLC), and *Subnetwork Dependent Convergence Protocol* (SNDCP) protocol. Additionally, the *Base Station System GPRS Protocol* (BSSGP) protocol, which acts on the *Gb-interface*, influences GPRS performance capabilities.

Especially the high response times due to channel access delays on the radio interface are limiting the performance of interactive applications. The rigid *Time Division Multiple Access* (TDMA) frame structure introduces variable waiting times until channel set-up messages can be transmitted over the radio link. The transmission time over the bandwidth-limited radio link increases the delay on a GPRS network. The radio bandwidth can be increased by assigning multiple timeslots to a user, but multislot operation is limited by the GPRS handsets. Currently available handsets support four downlink timeslots and three uplink timeslots. Further increase of multislot operation is only possible by implementing *Enhanced GPRS* (EGPRS), the packet switched component of *Enhanced Data Rates for GSM Evolution* (EDGE). The bandwidth per user can also be increased by the higher coding schemes *CS-3* and *CS-4*. High utilisation of CS-3 and CS-4 is improving the GPRS performance to a great extent, but only on condition of a low interfered radio environment. In addition, CS-4 can only be used when the mobile client is static; as soon as the MS sets into motion the transmission errors of CS-4 coded radio blocks are increasing. Even at moderate MS speed, the usage of CS-4 is restricted to the inner zone of a cell, with good coverage level and particular line-of-sight to the base station.

A disadvantage of GPRS data transmission is the limited support of user mobility. Changing the serving cell always results in an intermittent connection and requires a change in the routing of *packets data units* (PDU) from one base station to another, what inherently involves loss of data packets. Forwarding data packets from the original cell to the new serving cell is possible in recent GPRS implementa-

tions, when both cells belong to the same BSS. Cell reselections between cells of different BSSs are inevitably leading to loss of PDUs.
The Gb-interface, which connects the SGSN and the BSC, is also determining the end user performance perception. The flow control of the BSSGP layer has to manage the data flow between (SGSN) and (PCU) in downlink direction only by estimating the actual PCU buffer level based on limited feedback from the PCU. A difference between estimated and actual bucket level can either overload the PCU buffer or reduce the throughput by not supplying the PCU with a sufficient amount of packet data units.
Other GPRS protocol functions like *GPRS Tunnelling Protocol* (GTP), which acts on the *GPRS backbone network*, are not offering room for performance improvement beyond proper dimensioning.
The behaviour of Internet applications and their associated protocol stack such as *TCP/IP* (Transport Control Protocol, Internet Protocol) or *UDP/IP* (User Data Protocol) perform very differently when transmitted over a GPRS network. The most important network-layer transport protocol, TCP, has to be implemented in up-to-date versions to cope with the high latency and packet losses in the wireless environment. Advanced protocol implementations such as *TCP Vegas* or new features such as explicit congestion notification are able to improve the TCP performance over low bandwidth and high delay GPRS network links. However, a general improvement for the majority of applications used in GPRS networks can only be achieved by using *Performance Enhancing Proxies* (PEP). The most common concepts of PEPs are optimising the Internet protocols for use in wireless networks or to shadow servers in the Internet using legacy TCP implementations. Compression of data and protocol headers is also used to save transmission time. Enhancements on application layer are caching frequently accessed web pages, modifying web page content by removing tags and elements like applets and script elements. Implementing a performance enhancing proxy is essential for GPRS network links to overcome the bandwidth limitation and high system latency.
Current GPRS networks are offering maximum downlink bandwidth of 81.6kbit/s and 61.2kbit/s in the uplink rather than the theoretical maximum of 171.2kbit/s. However, the GPRS downlink bandwidth at four timeslots and CS-4 is fairly above the bandwidth offered by basic ISDN dial-up connections. Generally spoken is the performance perception over GPRS in the majority of mature GSM networks improved compared to typical analog, wired-link dial-up connections.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, I give a comprehensive introduction of GPRS, the General Packet Radio Service. A review of GPRS architecture and protocol stack is presented in section 1.1, followed by a short introduction of GPRS mobility management, which shows the procedures to establish the logical connection between a mobile station (MS) and the GPRS network.

1.1 GPRS Overview

GPRS is a packet-switching extension of GSM with the aim to enable high-speed data transmission over GSM radio channels.
The technical specification (TS) of GPRS, as it was defined by ETSI (European Telecommunication Standards Institute)[1], has been introduced in the third phase (Phase 2+) of the proceeding development of GSM specifications[2]. The first specification phase, Phase 1, was covering the basic GSM service description without including data transmission capabilities. In the second phase, Phase 2, data transmission based on RLP (Radio Link Protocol) has been included in early 1992, see [GSM04.22][3]. This circuit-switched method for data transmission and telematics according to the RLP protocol offers data transmission using a full duplex connection with 9.6kbit/s or an improved user data rate of 14.4kbit/s with less signalling overhead and a different channel coding, TCH/F14.4, see [GSM04.22].
The shortcomings of the RLP based data transmission led to the definition of

[1] Link to European Telecommunication Standards Institute: http://www.ETSI.org/

[2] The technical specifications for GSM comprise 3 major development stages, Phase 1 specifies the basic GSM functions. In the second phase, Phase 2, enhancements in GSM services like data transmission have been introduced and finally in the third phase, Phase 2+, services like HSCSD, GPRS or EDGE have been included in the specification suite.

[3] The version numbers of specification documents quoted in this thesis refer to the final publication date of the particular feature. The development of all the different versions of the GSM specification documents can be found in ftp://www.3GPP.org/Specs/archive, the most recent version of specifications can be found in ftp://www.3GPP.org/Specs/latest

HSCSD (High Speed Circuit Switched Data) in Phase 2+. HSCSD theoretically allows simultaneous usage of up to 8 TCH (Traffic Channels) and has the advantage that no changes in the existing GSM infrastructure are required, but it still sticks to the circuit-switched technique. Circuit-switched data transmission techniques dedicate the resources for the entire duration of the connection for data transmission in both directions, uplink and downlink, resulting in a waste of radio resources when the data traffic is of bursty nature, i.e. when data traffic consists of a sequence of data packets following silent intervals without data transmission. These disadvantages have been passed by defining a packet-switched service (GPRS) in the Phase 2+ specifications.

GPRS to be seen as the superior method for data transmission compared to the circuit-switched HSCSD. It offers several advantages due to the packet-switching technique. Packet switching accommodates the bursty nature of typical data traffic and enables dynamic radio resource sharing between voice and data traffic. Uplink and downlink channels can be set up separately to efficiently support asymmetric data traffic. GPRS introduces four different channel coding schemes providing different levels of data protection, enabling higher transmission speeds to capitalise on better channel conditions. Consequently, a selective ARQ (Automatic Repeat ReQuest) is implemented on the radio link layer for backward error correction and to cope with changing channel quality. Apart from the enhanced channel coding GPRS uses the same physical layer parameters as the circuit-switched voice traffic does. In particular the 200 kHz channel spacing, the TDMA (Time Division Multiple Access) frame structure with 8 timeslots per frame and 4.615ms frame duration are common.

In contrary to HSCSD, GPRS requires additional network entities to be merged with existing GSM network infrastructure. Additionally an increase of the BSS backhaul bandwidth is required when increased data rates using higher coding schemes (see chapter 6) have to be provided.

The final enhancement of data transmission in 2G (Second Generation) mobile communication networks is represented by EDGE (Enhanced Data rates for GSM Evolution)[4]. EDGE comprises a circuit-switched component, ECS (Enhanced Circuit Switched) and EGPRS (Enhanced GPRS), the packet-switched component. EGPRS is a further development of GPRS, mainly improving the radio path to enable higher data transmission speeds. EGPRS introduces an 8-PSK modulation scheme for higher data rates and improved channel coding and link layer functionality to facilitate a maximal gross data rate per carrier of 473.6kbit/s [EGPRS01]. The specifications of channel coding for EGPRS can be found in [3GPP-TS05.03]. EGPRS requires far reaching BSS hardware changes, especially in the BTS (Base

[4] The Phase 2+ specifications followed 4 major releases, Release 96 and Release 97 covering the basic HSCSD and GPRS specification, which are enhanced in Release 98. EDGE has been specified in progress of the final set of GSM specifications in Release 1999, additionally is Release 99 preparing the ground for moving from GSM to next generation mobile networks, i.e. UMTS (Universal Mobile Telecommunication System)

Transceiver Station) radio part and the BSS backhaul bandwidth.
The mandatory changes in GSM system architecture to support packet data and the new entities introduced are described in the following.

1.2 GPRS Architecture

The GPRS architecture is designed to provide an interconnection between a packet-switched, IP based network and the existing circuit-switched GSM network structure, in particular the BSS, with the aim to improve data transmission capabilities of the existing GSM architecture [GSM03.02]. Figure 1.1 depicts the new data transmission functional units introduced in the existing GSM architecture.

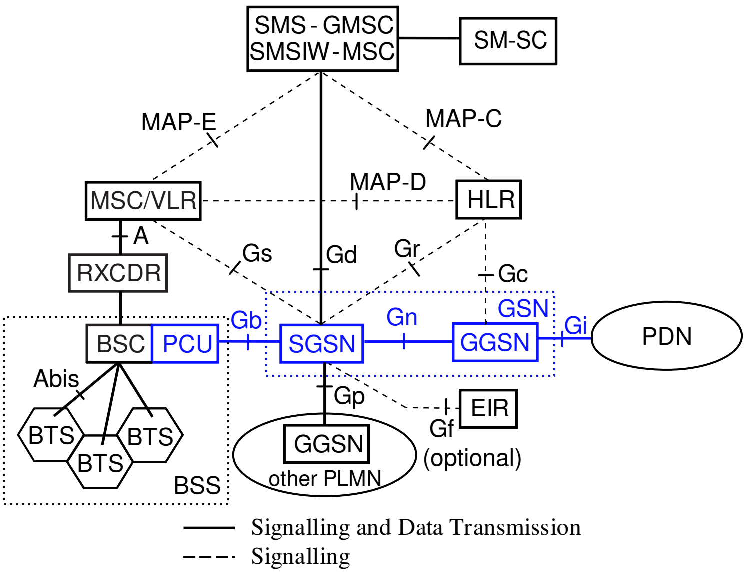


Figure 1.1: GSM and GPRS network architecture

1.2.1 PCU - Packet Control Unit

The BSS is extended by adding a PCU to provide the radio link between the GPRS network infrastructure and mobile subscribers. The PCU is closely integrated into the BSS, it fulfils radio functions such as selection of the proper channel coding scheme, power control, congestion control, broadcast of control information messages, or scheduling of data packets. The existing BSS infrastructure has also to be adapted for GPRS functionality. A typical BSS configuration roughly comprises about 100 BTS or 300 cells, strongly depending on the network architecture.

1.2.2 GSN - GPRS Support Node

The combination of SGSN (Serving GPRS Support Node) and GGSN (Gateway GPRS Support Node) builds the GSN (GPRS Support Node), which is the main element in the GPRS infrastructure. It is a broadband, packet switching backbone, which provides mobility management functions with the GPRS registered mobile stations and delivery of data packets to mobile stations on one hand, and it provides connection and inter-working with external data networks on the other hand.

1.2.3 SGSN - Serving GPRS Support Node

The SGSN is the network entity, which provides the transition between wireless radio environment and the IP based external networks, or strictly spoken the GPRS backbone network. The SGSN detects the mobile's position, i.e. the mobile stations routing area and the serving cell. It is responsible for network access control by making use of the enhanced HLR (Home Location Register), which includes GPRS subscription information, and it negotiates the QoS (Quality of Service) profile. The SGSN is also provides GPRS billing interface functions. The SGSN also performs authentication and ciphering based on the same authentication algorithms used in GSM, but uses a modified A5 encryption algorithm, which is optimised for packet data transmission. In contrast to GSM is data encryption in GPRS spanning MS, BSS, Gb-interface and the SGSN. The Gb-interface is included in the secured path to ensure data security over a leased frame relay network. A typical GPRS SGSN is able to serve about 20 BSS, or roughly 6000 cell sites.

1.2.4 GGSN - Gateway GPRS Support Node

The GGSN provides the connection between backbone network and external data networks, like the Internet. It de-encapsulates and forwards external data network packets to the appropriate SGSN or MS respectively. This forwarding of GPRS user data packets or GPRS-PDUs (Packet Data Units) across the GPRS backbone network (Gn-interface) is also referred to as *tunnelling*. One GGSN typically serves 10 SGSN, indicating that the mobility functions in the GGSN are quite reduced compared to the SGSN; the GGSN mainly fulfils routing functions. The GGSN is also involved in the billing of data traffic.

1.3 GPRS Protocol Stack for Data Transmission

The GPRS protocol stack for data transmission according to [3GPP-TS03.60] is shown below. The GPRS protocol stack is mainly designed to transport PDP (Packet Data Protocols) based on the Internet Protocol suite, i.e. applications using TCP/IP or UDP/IP as transport and network protocols. Release 1997 of

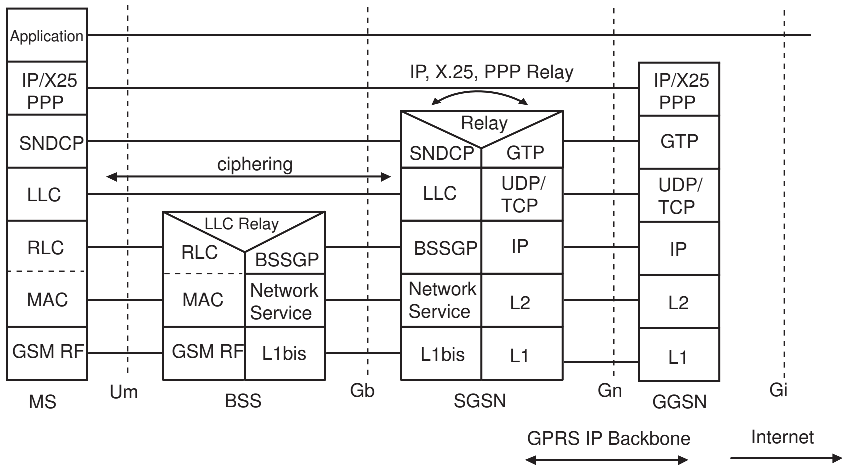


Figure 1.2: GPRS data transmission plane

the GSM specifications also propose an interface to the X.25 network protocol, but actually X.25 is of minor importance compared to widespread IP based PDPs. Furthermore, the support of PPP (Point to Point Protocol) was included in the Release 1998 TS, see [3GPP-TS03.60], version 7.4.0. PPP would offer the advantage of having a single PPP connection, e.g. between a computer connected to the MS using a dial-up connection, across the GPRS core network, to a host in an external PDN.
The highest protocol layers below the application protocols are GTP (GPRS Tunnelling Protocol) on the side towards the PDN and SNDCP (Subnetwork Dependent Convergence Protocol) on the radio side. The following description of data transmission protocols follows the data flow in downlink direction, from an external PDN to the MS.
Once a logical connection between MS and SGSN and between SGSN and GGSN is established, the MS is assigned an IP address and data packets can be transmitted via the GPRS network. This procedure is known as PDP context activation, see [3GPP-TS23.060].

1.3.1 GTP - GPRS Tunnelling Protocol

GTP enhances the static routing functions of IP to transmit data to and from mobile hosts. Network PDUs arriving at the GGSN via the Gi-interface are encapsulated within GTP-packets and transferred to the serving SGSN. Different network protocols can be encapsulated within GTP. GTP tunnels N-PDUs (Network-Packet Data Units) over the GPRS backbone network either to a SGSN within the GSN or to a SGSN in another GSN. GTP is terminated in the GGSN and in the SGSN; from the end-to-end point of view GTP is not visible. Signalling functions of GTP are required for PDP context activation.

1.3.2 GPRS Backbone Network Protocols

The transmission of GTP-PDUs across the backbone network is done via TCP/IP or UDP/IP protocols. The TS [3GPP-TS29.060] recommend the use of a connection-oriented path on the backbone network when the T-PDUs (T-PDU is the payload that is tunnelled in the GTP tunnel) are based on connection-oriented protocols, such as the X.25. More common is a connection-less path using UDP/IP when the T-PDUs are based on connection-less protocols, such as IP.
Backbone network protocols and GTP have little influence on the GPRS end-to-end performance as long as no N-PDUs are discarded on congestion.

1.3.3 SNDCP - Subnetwork Dependent Convergence Protocol

SNDCP is the utmost protocol on the wireless radio part of the GPRS network. The SGSN receives T-PDUs on the Gn-interface, encapsulated in the GTP. After removing the GTP header, the original network protocol PDU (IP, X.25 or PPP) is passed to the SNDCP layer.
The purpose of SNDCP is to provide protocol transparency for a variety of network protocols (IP, X.25, or PPP). Therefore, introducing new network layer protocols to be transferred over GPRS is possible without any changes to the underlying GPRS link protocols.
Another link layer function of SNDCP is to improve efficiency by supporting TCP/IP header and V.42bis data compression algorithms.
SNDCP also provides the session management including multiplexing of N-PDUs and mobility management, to keep track of the mobile stations position within the serving area of the SGSN.

1.3.4 BSSGP - Base Station System GPRS Protocol

BSSGP is the highest protocol layer acting on the Gb-interface. It provides a connection-less link between the SGSN and the BSS to transfer data unconfirmed between the SGSN and the BSS. It performs routing functions to transmit data to the serving BSS and radio related quality of service. BSSGP has to perform control of the flow of data between the SGSN and the BSS, it handles paging requests from the SGSN to the BSS and it supports multiple layer 2 links between the SGSN and the BSS. BSSGP uses frame relay as underlying network service protocols.
Flow control on the Gb-interface is a major issue in GPRS networks, leading to improvements incorporated into the flow control algorithm of the Release 99 TS.

1.3.5 LLC - Logical Link Control

The LLC provides a highly reliable logical connection between the SGSN and the MS. In addition, the LLC has been designed to be independent of the underlying radio interface protocols, which is important for a future incorporation of EGPRS. LLC includes functions for provision of multiple logical link connections discriminated between by a different SAPI (Service Access Point Identifier). Additionally LLC comprises sequence control, error detection and recovery and ciphering. LLC can be operated in acknowledged or unacknowledged mode. The main LLC parameters influencing GPRS performance are acknowledged vs. unacknowledged operation; especially the interaction between LLC flow control and TCP flow control are of major interest.

1.3.6 RLC - Radio Link Control

The RLC layer has to provide methods for uplink and downlink channel access, resolve simultaneous assignment requests from different MS within the same cell and control the assignment of multiple timeslots to a MS. RLC blocks can be coded in four different channel coding schemes, offering a wide range of data protection from a secure channel coding with FEC (forward error correction) at low user data rates to a coding scheme without any channel coding protection, but therefore with very high user data rates. Data transmission over the error-prone radio link makes the implementation of BEC (Backward Error Correction) mandatory. This is provisioned by a selective retransmission algorithm. Blocks transmitted in error are requested to be retransmitted within an acknowledgement message. RLC can also be operated in unacknowledged mode; however, unacknowledged RLC mode is not used in GPRS. A detailed description of the RLC layer is provided in chapter 6.

1.3.7 GPRS MS - Mobile Station

The MS has to support almost all GPRS features, while most of the GPRS features are optional on the network side. The requirements on protocol functions for the MS are more severe compared to the requirements the network has to fulfil. This increases the demands on the MS software. Only new features according to Release 1999 or Release 4 of the technical specifications can be negotiated between MS and network. Two examples of features which are mandatory for the MS but optional for the network are shown below:

- Layer 3 signalling can be done using GSM signalling channels on the BCCH (Broadcast Control Channel), or alternatively a GPRS specific signalling protocol on the PBCCH (Packet Broadcast Control Channel). Both control channel structures have to be supported by the MS, PBCCH is optional for the network.

- It is mandatory for the MS to support all four coding schemes, whereas the BSS just has to support CS-1. As stated in [3GPP-TS03.64]: *"All coding schemes (CS-1 to CS-4) are mandatory for MSs supporting GPRS. CS-1 is mandatory for a network supporting GPRS."*

1.4 GPRS Protocol Overhead

The effect of GPRS protocol overhead on user throughput on different protocol layers and for a single timeslot GPRS connection is depicted in table 1.1. The data

Protocol Level	CS-4	CS-3	CS-2	CS-1
Nominal RLC/MAC Data Rate [kbit/s]	21.40	15.60	13.40	9.05
RLC/MAC Payload Data Rate [kbit/s]	19.55	13.80	11.50	7.67
LLC Payload Data Rate [kbit/s]	19.43	13.72	11.43	7.62
IP Payload Data Rate [kbit/s]	19.02	13.43	11.19	7.46
TCP Payload Data Rate [kbit/s]	18.77	13.25	11.04	7.36
Number of RLC Frames	30	43	51	77
Number of LLC Frames	3	3	3	3
DL RLC/MAC Efficiency [%]	0.96	0.96	0.96	0.96
UL RLC/MAC Efficiency [%]	0.92	0.92	0.92	0.92

Table 1.1: GPRS data rates on the different protocol levels

rates stated below have to be considered as guide numbers only, the actual data rate may change in dependence of TCP/IP packet size or LLC frame size. Table 1.1 assumes a TCP connection with 1460 byte segment size, SNDCP compression is disabled and LLC operates in unacknowledged mode, maximal LLC frame size negotiated between SGSN and MS is assumed to be 500 byte, as many of the available GPRS MS support LLC sizes in the range of 500 byte. The rows stating efficiency on RLC layer consider the required signalling on RLC layer. These values differ depending on the used coding scheme.
It can be seen that CS-4 includes about 10% overhead on the RLC layer, whereas CS-1 almost 15% overhead on RLC level includes, pointing out the high amount of bits used for BCS (Block Check Sequence) error detection at CS-1. The user data rates on TCP layer are at CS-4 12% and at CS-1 19% below the nominal data rate.

1.5 GPRS Signalling Plane

The GPRS signalling functions control network access and support of user mobility by establishing routing paths to keep track of the MS position. Figure 1.3 shows the signalling plane according to [3GPP-TS03.60]. The GMM/SM (GPRS

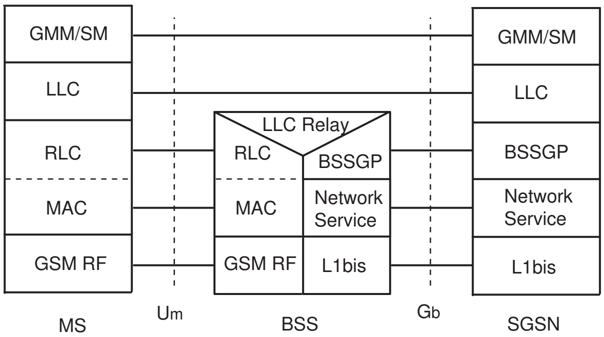


Figure 1.3: Signalling plane MS - SGSN

Mobility Management and Session Management) resides above the LLC layer and interacts with SNDCP. Mobility management and access control is performed between the MS and the SGSN.
The signalling functions itself are in general of minor importance from the performance evaluation point of view, but during the signalling procedures performance affecting parameters are negotiated. Additionally, user mobility is an important issue for GPRS performance. It will be shown in chapter 7.5 that user mobility considerably influences the application throughput, depending on the networks behaviour on cell reselection.

1.6 GPRS Mobility Management

Mobility Management (MM) includes three different states denoted by MM context at the SGSN and the MS. The MS states (Idle, Standby and Ready) are related to the actual data transmission between SGSN and MS.

- **Idle state**
  A MS in idle state is not able to access the GPRS network. Both, MS and SGSN do not have any mobility context and data transmission over the GPRS network is not possible. The MS in idle state is just reachable for circuit-switched voice connections.

- **Standby state**
  The MS and the SGSN have a mobility context including the MSs Routing Area (RA), but not the cell in which the MS actually is located. In this state, the MS will not send any cell update information but will only send RA updates, periodically or when the MS moves to a cell within a new RA.

- **Ready state**
  The MS and the SGSN have a mobility context, which includes information

in what particular cell the MS is currently located. In this state, every time the MS moves to a new cell, it has to send a cell update message to the SGSN. The TS [3GPP-TS03.60] states: *"A cell update is any correctly received and valid LLC-PDU carried inside a BSSGP PDU containing a new identifier of the cell"*.
The ready state requires a high amount signalling data exchange between MS and SGSN, and so, a high amount of radio interface resources is used for signalling. To avoid this high usage of signalling, the MS will go to standby after a certain time[5] the MS does not send any data.

The state transitions and the events causing state transitions (e.g. mobility management messages) are illustrated in [3GPP-TS03.60]. The current mobility state only depends on the state of uplink data transmission. The MS is forced to be ready when data is transmitted; the MS goes into standby mode after the transmission finished and the ready timer expired. Obviously, it is possible for the MS to be in standby state with an active PDP context established and an application active. The following example shall illustrate the transition from ready and standby state.
In both, standby mode and ready mode the GPRS MS is reachable for data services and is able to set up a data call. Figure 1.4 shows the results from transmitting

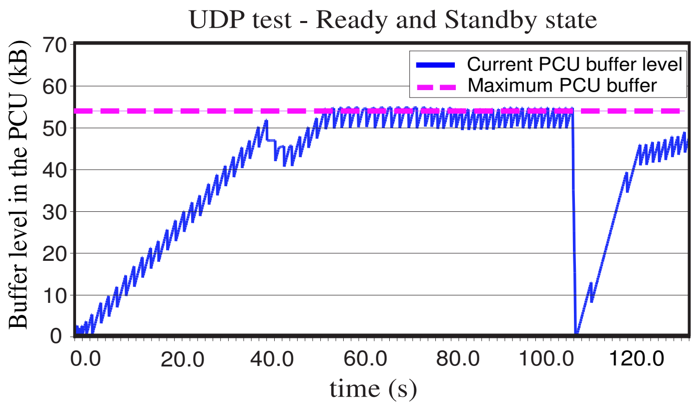


Figure 1.4: Test of mobility management functions

a UDP data stream from an external server to the MS without any transmission from the MS in uplink direction. The graph shows the current level of buffered LLC frames in the PCU. After 40 seconds, the PCU buffer gets full and the PCU indicates the SGSN a flow control situation. The SGSN throttles the data rate and the PCU buffer stays at a maximum level. Approximately 120s without any uplink activity (any kind of LLC-PDU sent to the SGSN) the ready timer expires

[5] The Ready timer typically expires after thirty seconds to one minute

and the MS goes into standby state. The SGSN has to stop forwarding the UDP packets to the MS since it is not aware of the serving cell of the MS in standby state, the PCU clears the buffer assigned to the MS in standby state. The SGSN is still receiving data from the sending peer. Following, the SGSN needs to page the MS to deliver these data packets. The paging forces the MS to ready again and the data delivery continues. The standby period can be seen in the interrupted data transmission.

1.7 Considerations on the "Always On"

The 'always on' situation means that the MS is attached to the GPRS network and has a PDP Context activated. In this situation, the MS is assigned an IP address and can receive or send data through the GPRS transport pipe. The MS is also reachable for push services, where the request for application data exchange is initiated by an entity in the external packet data network. The PDP context can be activated from the MS or from the network side, i.e. the GGSN. Currently, due to the limited IPv4 address range, the network requested PDP context activation is not deployed.
It has to be considered that billing commences once a PDP context is successfully activated, but the signalling messages are not charged, though. It is just LLC-BSSGP protocol messages and will generate a Mobility Management CDR (Charging Data Record), called M-CDR. For data, which means packets over SNDCP protocol in the SGSN, and packets over GTP on the GGSN, the system will send other types of CDRs, called SGSN CDR (S-CDR) and GGSN CDR (G-CDR), which include fields related to the volume of bytes in the uplink and downlink direction, see [3GPP-TS32.015].

Chapter 2

Analysis of GPRS Protocol Functions

In this chapter, I identify the performance relevant GPRS network entities and their associated protocol functions. Initially, I discuss the key performance metrics required to assess a GPRS network and I show examples of typical throughput and latency results from life GPRS networks. The main parameters determining GPRS performance are summarised at the end of this chapter.

2.1 GPRS Key Performance Metrics

The end-to-end performance of a GPRS network link requires two fundamental performance parameters: the *latency* or *r*ound trip delay for a data packet to be transmitted from one end to the other and the *throughput*, i.e. the end-to-end data transmission rate to transfer data packets from one end to the other. Beside these two key parameters the *variability of latency*, the *throughput-variability* and the *packet loss rate* are important and common metrics for performance evaluation of IP based, packet-switched networks.
Typical examples for throughput and latency of a GPRS network are given next to illustrate the performance to be expected for existing GPRS networks.

2.1.1 Round Trip Delay Examples

Table 2.1 shows ICMP ping results for 10 byte and 1500 byte payload. The measurements were done in a life network. Actually, the round trip delay of an ICMP ping includes uplink latency and downlink latency (in that order). A detailed analysis of a 10-byte ping, measuring the timestamps of LLC frames passing the Gb-interface and the MS, decoding the radio interface signalling messages and measuring timestamps of ICMP packets on the client computer indicates that the

Ping payload size (byte)	10	1500
Pings sent	1000	1000
Echoes received	1000	1000
Pings lost	0	0
Reply time minimum (ms)	508	2098
Reply time maximum (ms)	883	2414
Reply time average (ms)	590	2103

Table 2.1: Example of ping results for 32 byte and 1500 byte payload

main contribution to the round trip delay is accumulated in the BSS, the radio path and the connection of MS and client PC, whereas the delay within the GPRS core network and the PDN connecting the pinged host is very small. Measurement set-up and results of ping analysis are shown below.

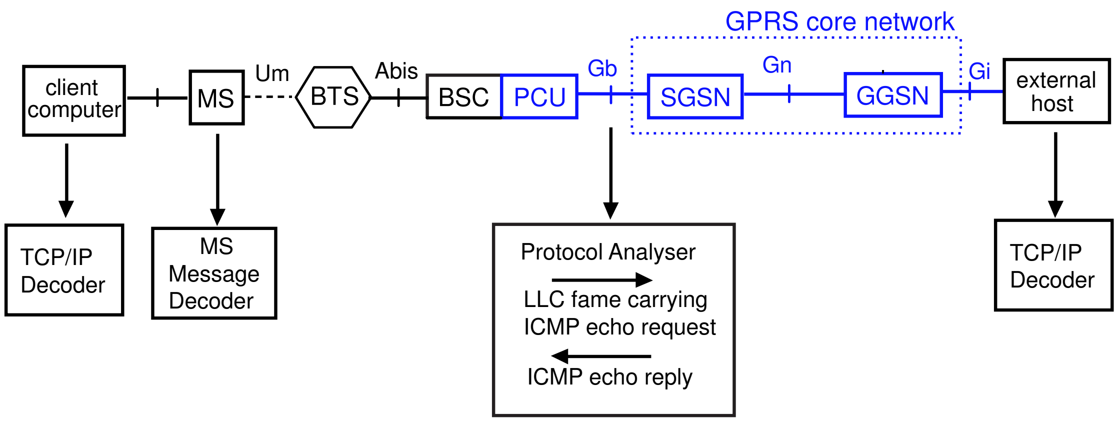
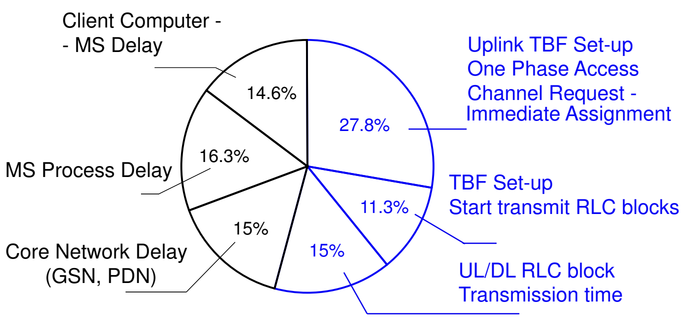


Figure 2.1: Measurement set-up for ping analysis



	Time (ms)	Percent (%)
Channel Request - - Immediate Assignment	148	27.8
Start transmit RLC blocks	60	11.3
UL/DL RLC transmission	80	15
Core Network Delay	80	15
MS Process Delay	87	16.3
Delay Client Computer - MS	78	14.6
Total Round Trip Delay	533	100

Figure 2.2: Analysis of a 10-byte ping

Only 15% of the round trip delay is accumulated in the core network and external

PDN. The ICMP packet suffers a relatively short delay of 80ms from the Gb-interface, through the GPRS core network, to the pinged host and back. The major contribution to the delay, 288ms or 54.1%, is accumulated in the BSS, in particular the signalling and transmission on the radio path. In detail, the BSS delay includes a period of 208ms or 39.1% just for radio interface signalling[1] and only 15% of the ping reply delay comes from the actual data transmission on the radio link. 165ms or 30.9% of the round trip delay is piled up in the client computer, the transmission between MS and client and process delay of the MS. This indicates that not only the network but also the MS and the dial-up connection, as well as the client PC noticeable affect the end-to-end performance perception.

2.1.2 Downlink Throughput Examples

Most of the GPRS networks currently deploy coding scheme CS-1 and CS-2 only, the majority of the available MS support three or at most four downlink timeslots. This leads to a maximum gross data rate on RLC layer of 53.6kbit/s and therefore, the application layer throughput will be approximately 44kbit/s. Networks supporting all four coding schemes provide 85.6kbit/s or 74kbit/s net data rate accordingly. Table 2.2 gives typical FTP throughput results from life networks.

Multislot class	Number of Rx slots	Throughput CS-1,2 (kbit/s)	Throughput CS-3,4 (kbit/s)
4	3	32.5	55.2
8	4	44.0	74.0

Table 2.2: Example of FTP throughput results from a life network

The examples above show throughputs slightly above 10kbit/s per timeslot when the network supports CS-1 and CS-2. The throughput increases by about 60% when the network supports CS-3,4. The end-to-end throughput results per timeslot are growing disproportionately the more timeslots are allocated. I analyse this effect in section 5.1.

2.2 GPRS Performance Key Parameters

In this section, I discuss the connection between GPRS protocols and GPRS performance. As shown previously, the radio link is the main instance determining GPRS performance.

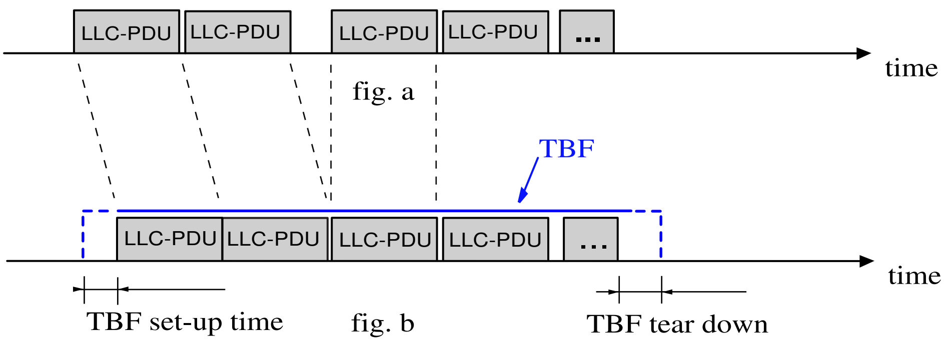
[1] TBF set-up, see chapter 3

2.2.1 Radio Interface Parameters

The parameters having effect on the radio performance are first the radio channel set-up and second the available bandwidth characterised by the number of timeslots assigned to the MS. Finally, the efficiency of the RLC layer, mainly determined by coding scheme usage and BLER (Block Error Rate), are key parameters of GPRS performance.

2.2.1.1 TBF - Temporary Block Flow Set-up

Once the application provides the upper layer protocols in the MS with N-PDUs, the MS requests an uplink channel to transmit its data packets; the PCU grants network access in reply. This temporary connection is named TBF. On the network side, the arrival of LLC-PDUs in the PCU triggers the establishment of a downlink channel. The PCU pages the MS if it currently is in standby mode and sets up the downlink TBF. In the best case, the connection is as long available as data arrives in the PCU to be transmitted to the MS.
The set-up times increase the systems response time and latency. High latent links potentially reduce the upper layer protocols performance. In particular, acknowledgement driven protocols such as TCP suffer from high latent links. The effects of TBF set-up times are shown in figure 2.3. It can be seen that the initial TBF set-up delays the transmission on the radio interface. Hence, it reduces the effective bandwidth of the radio interface.



The TBF set-up procedure delays the transmission on the radio interface. Fig. a. shows the arrival of LLC-PDUs in the PCU, fig. b is showing the delayed transmission after setting up the TBF. Packets arriving at the PCU with a TBF already established can be delivered immediately, apart from the processing delay.

Figure 2.3: Establishment of a TBF (Temporary Block Flow)

2.2.1.2 Timeslot Allocation

A TBF can comprise multiple timeslots of a TDMA frame, depending on the available timeslots and the MS multislot capabilities. Therefore, the radio throughput increases compared to a single timeslot connection. The GPRS timeslot allocation can change dynamically depending on the demand for voice and data traffic in the cell. A comprehensive modelling of dynamic resource allocation in GPRS is discussed in [OEZC01].

2.2.1.3 TBF Maintenance

After setting up a TBF using one or multiple timeslots, GPRS performance is predominantly determined by the quality of the radio environment. The network assigns either a strong coding scheme with high protection against transmission errors and reduced data throughput, or a weak coding scheme with low error protection overhead but high data throughput. The coding scheme selection depends on the radio link quality and the RLC block error rate. Simulation results according to [ETSI-TR101.115] identify the main parameters influencing coding scheme usage and BLER:

- Co-channel interference (C/Ic) due to tight frequency reuse
- Environmental conditions, e.g. build-up of land
- Mobile station speed
- Frequency hopping or non-hopping

A further discussion of the RF channel properties in given in chapter 6.

2.2.2 LLC and SNDCP Layer Parameters

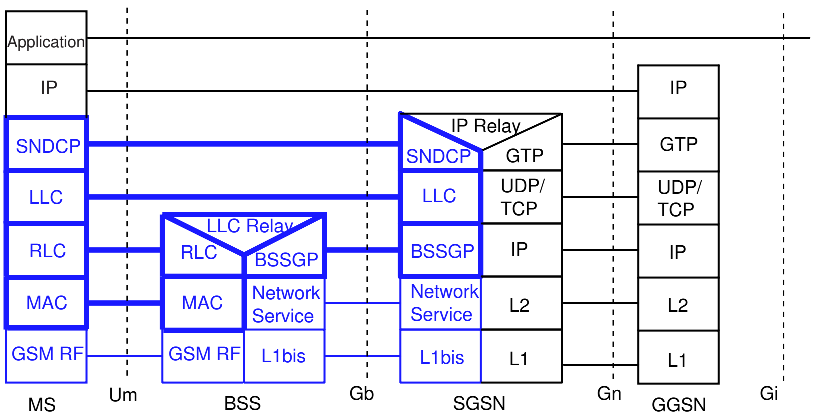
The LLC layer can be operated in acknowledged mode or unacknowledged mode. The LLC acknowledged mode can improve the performance when loss of LLC-PDUs on the Gb-interface has to be considered, e.g. due to limited capacity of the frame relay network, or when frequent cell reselects between two BSS require flush of LLC-PDUs in the source PCU, see [3GPP-TS08.18]. On the other hand, destructive interactions between the LLC retransmission mechanism and the TCP retransmission based on the acknowledged operation are a known issue.
The SNDCP layer offers data and TCP header compression. Similar to the LLC acknowledged mode the efficiency of compression is a discussed issue. Further results can be found in chapter 7.

2.2.3 TCP/IP Parameters

The GPRS system performance does not just depend on the GPRS functionality itself but also on the applications using the GPRS transport pipe. The majority of typical applications used in a GPRS network are build on the TCP/IP protocol, but TCP is known to be sensitive to high latency and low bandwidth links such as GPRS. TCP can only effectively utilise a high latency and low bandwidth links if it sends a high amount of data into the system (determined by the TCP window size, see chapter 8.1), subsequently high amount of buffer resources are required within the GPRS network to avoid packet loss on buffer overflow.

2.3 Performance Relevant GPRS Protocols

As shown previously, the analysis of GPRS protocol performance implications requires focus attention on the protocols managing the radio link, see figure 2.4, below. It can also be seen that the protocol implementation in the MS highly influences GPRS performance, e.g. the multislot capability is mentioned as most obvious example.



The emphasized, blue marked protocol functions decisively influence data transmission over GPRS.

Figure 2.4: Protocol layers determining GPRS performance

2.4 Classification of GPRS Network Links

Considering a MS supporting at most four timeslots, the maximum GPRS bandwidth will approximately be 50 to 80kbit/s, depending on the coding scheme. The round trip delay typically ranges from one to two seconds, depending on timeslot allocation, radio quality and N-PDU size. The resulting bandwidth delay product (BDP) typically covers a range of 4 to 20kB. Based on these key metrics and according to [RFC2757] GPRS network links can be classified as LTN links (long, thin network).

LTN links potentially reduce the performance of acknowledged driven protocols such as TCP (Transport Control Protocol). A TCP sender adapts the use of available bandwidth based on feedback from the receiver. Thus, the high latency of GPRS networks reduces the dynamic of TCPs adaptation algorithms. A detailed description of performance implications for TCP-based applications over GPRS are discussed in chapter 7.

Chapter 3

Radio Link Set-up and Latency

The transmission of RLC blocks on the radio path is done by setting up a TBF, a Temporary Block Flow. This TBF set-up as well as the TBF tear down procedures are required for every burst of application layer PDUs transmitted over the radio link.
The TBF set-up time within the fixed frame structure of the TDMA domain is a decisive parameter for evaluation of GPRS performance. The TBF set-up times contribute substantially to the high latency of GPRS networks, as shown in figure 2.2, previously.
The TBF set-up procedure uses the control channels on the BCCH/CCCH[1] as circuit-switched traffic does. GPRS also defines optional PBCCH signalling channels for independent signalling for circuit-switched voice traffic and packet-switched data traffic, but PBCCH is still not implemented in GPRS networks. The processes are only slightly different for TBF set-up when PBCCH and PCCCH control channels exists and furthermore, PBCCH will most probably not be implemented in the near future. Hence, I only discuss CCCH signalling.
In this chapter, I provide a brief overview of the signalling on the radio link and the RLC layer, followed by the analysis of TBF set-up times for various MS transmission states.

3.1 Description of RLC/MAC Functions

The control channels for signalling are using the 51-multiframe structure for CCCH, which control the MS in idle mode, and a 52-multiframe structure for the GPRS packet transmission mode. These control channel structures are summarised briefly, as far as relevant for TBF set-up and latency analysis.

[1] Broadcast Control Channel, Common Control Channel

3.1.1 GPRS Multiframe Structure

GPRS uses a new 52 TDMA frame structure for transmission of RLC/MAC blocks, different from the existing 51- and 26-multiframe structures for circuit-switched GSM. The 52-multiframe does not have a rigid structure; essentially the different channels are identified by the message type. The multiframe can carry control channels or data channels. Each RLC block in the GPRS 52-multiframe consists of four consecutive bursts. This new frame structure leads to a RLC/MAC block period of 20ms, see figure 3.1. The frame structure is applied during GPRS data transmission, when the MS is in packet transfer mode.
One TDMA frame lasts: $576.9\mu s * 8 = 4.615ms$
One RLC block spans 4 TDMA frames: $4.615\mu s * 4 = 18.460ms$
The entire duration of a 52-multframe comprises 12 RLC blocks and 4 idle blocks:
$18.460ms * 12 + 18.460ms = 240ms$
Thus it follows: 1 RLC block period $= \frac{240ms}{12} = 20ms$

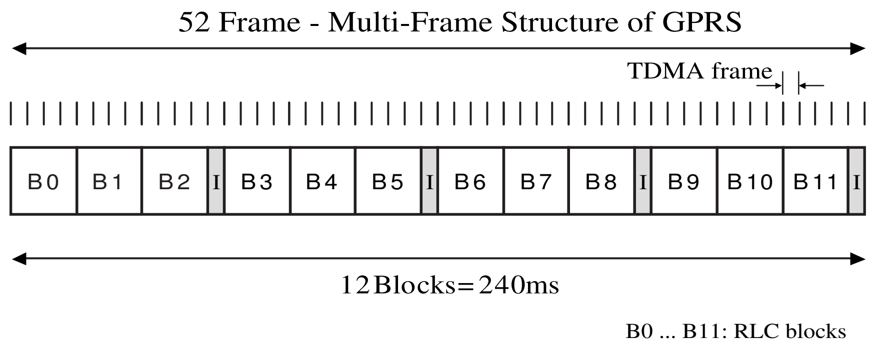


Figure 3.1: GPRS 52-multiframe structure

3.1.2 Multiframe Structure for Common Control Channels

The CCCH frame structure consists of 51 consecutive frames on the BCCH timeslot 0, for downlink only (see figure 3.2). One multiframe contains the four BCCH blocks, broadcasting the system information (SI) messages such as SI-13[2]. Then five frequency correction channels (FCCH) and five synchronisation channels (SCH) follow. Finally, nine paging groups, each comprising four TDMA frames complete the multiframe. Paging messages sent in a particular paging group are required to locate the MSs position in the cellular network. An important configuration parameter defined by the TS is BS_PA_MFRMS. Paging can only occur within the MSs paging subgroup, which appears once in every BS_PA_MFRMS multiframe on

[2] SI-13 basically informs the MS that GPRS is enabled in the cell

the CCCH. This is important for downlink TBF set-up in standby mode, where the MS needs to be paged before a DL TBF can be established. The parameter BS_AG_BLKS_RES determines the sharing of paging channels (PCH) and access grant channels (AGCH)[3]. The AGCH carries the IMMEDIATE ASSIGNMENT (IA) message, which is used for uplink TBF set-up. The impact of the CCCH structure

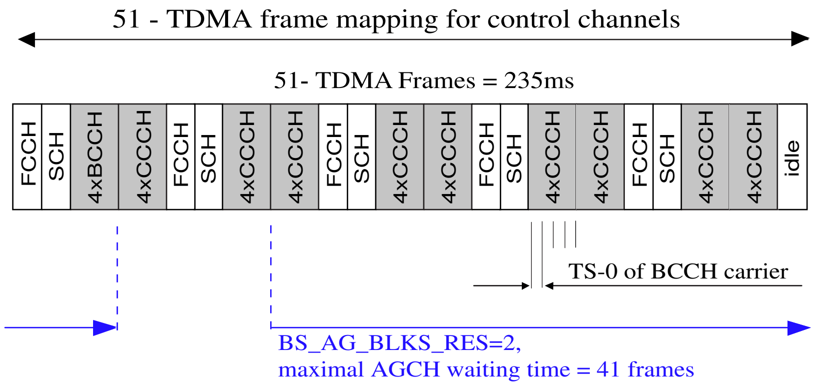


Figure 3.2: Frame mapping for control channels

is on TBF set-up procedures can be gathered from the example included in figure 3.2. Setting BS_AG_BLKS_RES = 2 will reserve paging groups B0 and B1 for AGCH, seven groups remain available for paging[4]. The maximum waiting time in case the PCU is not able to schedule the transmission of an IA message in block B1 is 189ms or 41 TDMA frames. The dimensioning of AGCH and PCH to achieve the best compromise between fast GPRS and GSM signalling and paging load has to consider circuit-switched and packet-switched traffic. The interconnection between GSM and GPRS signalling channels prevents a reduction of TBF set-up times by increasing the number of access grant blocks.

3.2 Radio Link Set-up

The following sections now describe contributions of radio link set-up to the GPRS network latency as well as the expected minimum, maximum and average times to set up a TBF on the radio link. I analyse the two basic methods of uplink TBF set-up, the *one-phase* and the *two-phase* TBF set-up. TBF set-up in downlink direction has to consider the mobility context and the DRX mode[5] to provide timings for TBF set-up. Finally, the signalling procedures to establish concurrent

[3] Refer to [3GPP-TS05.02], clause 6.5. and 6.5.6. Some access grant blocks may also be used for the notification channel (NCH), which is not considered here

[4] see [3GPP-TS05.02] clause 7, table 5, Mapping of logical channels onto physical channels

[5] see TS 100 531, GSM 03.13 version 6.0.0 Release 1997

TBFs are analysed.
The TBF set-up times include estimations on BSS response delays as well as MS response delays. I assume the latency in the BSS is 60ms. This value is assumed to be constant and equal for uplink and downlink latency. The MS process delay on one-phase access is assumed to be 80ms. These estimations may change for different BSS or MS types. Current BSS implementations further reduced the latency; a requirement to support multiple timeslot assignments per MS. However, the assumptions drawn here do not affect the generic evidence of TBF set-up delays stated in the following.

3.2.1 Mobile Originated Data Transfer

A GPRS capable MS scans the SI-13 message on the BCCH to get knowledge of GPRS availability in the serving cell. An example of a SI-13 message is shown in appendix A, table a-1. If a SI-13 message is present, the MS can request radio resources for uplink data transfer. The two major procedures (two-phase access and one-phase access) are discussed in the following.

3.2.1.1 Two-phase Uplink TBF Set-up

Two-phase access is initiated by a random access CHANNEL REQUEST (CR) on the RACH channel of the CCCH (BCCH carrier, timeslot 0). The PCU responds to this request with an IA message on the AGCH, which contains details of the PDTCH assignment together with timing advance information derived from the random access burst. This is the first phase of uplink TBF set-up. The signalling messages to invoke the two-phase TBF set-up procedure are shown in figure 3.3, below. The IMMEDIATE ASSIGNMENT message assigns a single uplink block to

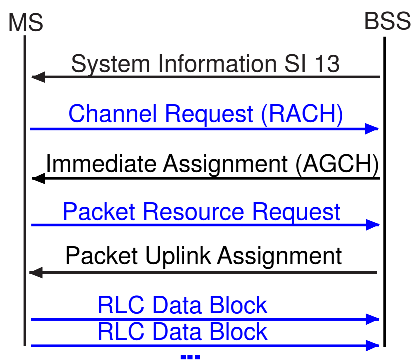


Figure 3.3: Two-phase uplink access

the MS. The MS switches to the PDTCH referenced in this message and sends a PACKET RESOURCE REQUEST (PRR). The PCU responds to the PRR with a

PACKET UPLINK ASSIGNMENT (PUA) message. This second phase of the uplink TBF set-up is finally exactly negotiating the uplink TBF. The set-up is considered to be finished when the first RLC data block is transmitted on the radio interface. An example of a two-phase access is depicted in table 3.1[6]. The first access phase

TDMA FN	Delay (ms)	Delta (ms)	Link	Message Type
775996			UL	Channel Request
776005	41.54	41.54	DL	Immediate Assignment
776025	133.84	92.30	UL	Packet Resource Request
776061	299.98	166.14	DL	Packet Uplink Assignment
776074	359.98	60.00	DL	Packet Downlink Dummy Control Block
776077	373.83	13.85	UL	Uplink Data Block
776078		4.62	DL	Packet Downlink Dummy Control Block
776081		13.85	UL	Uplink Data Block

Table 3.1: Example of two-phase access

from sending the CR until receiving the IA takes 41.54ms. The physical mapping of access grant channels on the 51-multiframe structure is specified by the parameter BS_AG_BLKS_RES.
Setting BS_AG_BLKS_RES = 3 will reserve paging groups B0, B1 and B2 for AGCH, while 6 groups remain available for paging. Thus, the maximum waiting time to transmit the IA message is 166ms or 36 TDMA frames. The minimum waiting time is assumed to be 5ms or 1 TDMA frame.
The duration of the second set-up phase is determined by processing times in the BSS and it can be delayed further by the TBF_STARTING_TIME parameter sent in the IA message. This timer delays the transmission of the PRR to ease re-synchronisation when the allocated PDTCHs are on a different frequency band as the BCCH carrier. The transmission delay of the PRR through the BTS to the PCU and process time in the PCU has to be considered as well as the delay and processing time to transmit the PUA. Below, a generic example of the timing of two-phase TBF set-up is shown. Comparing these theoretical values with the measured set-up time from table 3.1 shows slightly higher delay measured until the PUA is received by the MS, which presumably is explained by a queuing delay on the assigned PDCH. The transmission of the first RLC data block mainly depends on the MS process delay, but it has also to be considered that the MS has to decode the USF bits to enable uplink data transmission. When no downlink TBF is established the network has to send PACKET DOWNLINK DUMMY CONTROL BLOCKs to assign uplink timeslots.

[6] The AFN shown in the tables below denote the absolute frame number. According to [3GPP-TS05.10], the TDMA frame number - FN is defined to be in the range: TDMA FN : [0 to $(26 * 51 * 2048 - 1) = 2715647$]. The complete message content of the decoded assignment messages can be found in the appendix A.

Message	from - to	Um Tx Delay (ms)	BSS Delay min/max (ms)	Total Time min/max (ms)
RACH	MS - BTS	4.6	5	9.6
Imm. Assignment	PCU - MS	18.4	5/166	23.4/184.4
Packet Resource Request	MS - PCU	18.4	60	78.4
Packet Uplink Assignment	PCU - MS	18.4	60	78.4
First RLC Block	MS - PCU	20	60	80
Total Delay				
Minimum				**269.8**
Average				**350.3**
Maximum				**430.8**

Table 3.2: Estimated two-phase access timing

3.2.1.2 One-phase Uplink TBF Set-up

One-phase uplink TBF set-up is also initiated by a random access CHANNEL REQUEST on the RACH channel of the CCCH. The PCU responds to this request with an IA message on the AGCH of the CCCH with all details of the PDTCH assignment, such as Uplink State Flag (USF), channel description or coding scheme usage together with timing advance information based on the random access. The MS switches to this PDTCH, and after a certain mobile reaction time to the uplink assignment, the MS sends an uplink RLC data block with the contention resolution TLLI. The PCU sends a PACKET UPLINK ACK/NACK (PUAN), which contains TLLI and TFI in response to the first data block to the MS to complete the contention resolution procedure. The flow of TBF set-up procedure using one-phase access is depicted in figure 3.4, below.

The establishment of an uplink TBF in a single phase obviously is faster compared

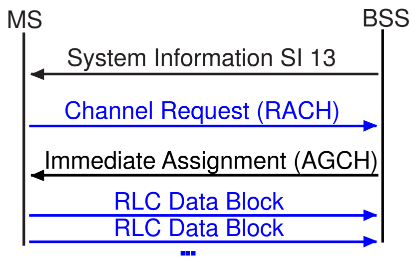


Figure 3.4: One-phase uplink access

to the two-phase TBF set-up. The example in table 3.3 shows a TBF established after 189.22ms. Expected timings for one-phase access associated with each stage of the message flow is shown in table 3.4. The TDMA frame structure again

TDMA FN	Delay (ms)	Delta (ms)	Link	Message Type
2505708			UL	Channel Request
2505726	83.07	83.07	DL	Immediate Assignment
2505749	189.85	106.15	UL	Uplink Data Block

Table 3.3: Example of one-phase access

introduces a maximum IA waiting time of 166ms. One-phase access typically causes a higher delay in the MS until the first data block can be transmitted and it has to be considered that the PCU cannot initiate the activation of USF before receiving the IA.

Message	from - to	Um Tx Delay (ms)	BSS Delay min/max (ms)	Total Time min/max (ms)
RACH	MS - BTS	4.6	5	9.6
Imm. Assignment	BTS - MS	18.4	5/166	23.4/184.4
MS Process Delay	MS - PCU			80
First RLC Block	MS - PCU	20		20
Total Delay				
Minimum				**133.0**
Average				**213.5**
Maximum				**294.0**

Table 3.4: Estimated one-phase access timing

3.2.1.3 Comparison of Two-phase and One-phase Uplink TBF Set-up

Comparing the average access times for one and two-phase access shows clearly faster one-phase TBF set-up. The reason for defining the two-phase access though is to indicate the MS capability of using multiple uplink timeslots while one-phase access using the CCCH inherently requests one uplink timeslot only (see 3.5, channel request message format). In the PRR, sent in the second phase of the two-phase access, information about the MS multislot class is included, see appendix A, table a-5 and the network will try to assign multiple uplink timeslots according to the MS multislot class. The MS distinguishes between one-phase and two-phase access based on the amount of data to be transmitted in the uplink. The format of the channel request to initiate either two-phase access or one-phase access is described in table 3.5, the complete description of the CR message is defined in [3GPP-TS04.18], Table 9.1.8.1, CHANNEL REQUEST message content. Assignment of multiple uplink timeslots is based on the amount of data scheduled for uplink transmission. Hence, the application behaviour influences the TBF set-up. Once the application changes the traffic bias, it is the networks responsibility

Channel Request Message Format	
MS codes according to establishment cause:	
bits 8 ... 1	
011110xx 01111x0x 01111xx0	One-phase packet access with request for single timeslot uplink transmission; one PDCH is needed.
01110xxx	Single block packet access; one block period on a PDCH is needed for two-phase packet access or other RR signalling purpose.

Table 3.5: Channel request message content

to decide which timeslot configuration allows optimal data throughput, especially for multislot classes which do not support simultaneous transmit and receive on all available PDTCHs[7]. The algorithms in the BSS, measuring the traffic volume in both link directions and determining proper timeslot assignment are not specified by the TS, so different BSS vendors will have proprietary implementations.

3.2.2 Mobile Terminated Data Transfer

After analysing the uplink channel access, the effects of the different MM context states on timing of downlink TBF set-up are analysed now in the following section.

3.2.2.1 Downlink TBF Set-up in Standby State

Establishing a downlink TBF when the MS is in standby state requires that the MS is paged initially to locate its position, i.e. the serving cell. This is achieved by broadcasting a PACKET PAGING REQUEST message on the paging channel of the paging sub-group the MS belongs to. Following this, the MS is required to set up an uplink TBF and send an arbitrary LLC frame (dummy frame) to the SGSN. The required uplink TBF set-up basically follows the procedures described in clause 3.2.1. As soon as the SGSN updates the cell location, the downlink TBF can be initiated. By this stage, the MS will be in ready state and hence listen to the CCCH. The establishment of the downlink TBF itself is done by sending an IMMEDIATE ASIGNMENT including a PACKET DOWNLINK ASSIGNMENT (PDA). The MS responds to the PDA with a PACKET CONTROL ACKNOWLEDGEMENT (PCA) message in the frame number indicated in the RRBP-field (Relative Reserved Block Period) of the PDA. However, it is likely that the timing advance information about that MS is out of date. Therefore, the PCA message, which in fact consists of four RACH access bursts, is used to update the timing

[7] E.g. MS class 6 supports three Rx timeslots and two Tx timeslots, but only four timeslots can be simultaneously active. Using two Tx timeslots reduces the maximum downlink allocation to two Rx, see appendix D

TDMA FN	Delay (ms)	Delta (ms)	Link	Message Type
1142002			UL	Channel Request
1142011	41.535	41.535	DL	Immediate Assignment
1142027	115.375	73.84	UL	Packet Resource Request
1142067	299.975	184.6	DL	Packet Uplink Assignment
1142080	359.97	59.995	DL	Packet Downlink Dummy Control Block
1142083	373.815	13.845	UL	Uplink Data Block
1142084		4.615	DL	Packet Downlink Dummy Control Block
1142088		18.46	UL	Uplink Data Block
			...	

Table 3.6: Example of paging for downlink TBF set-up

advance information. And finally, the network can start sending downlink RLC blocks, three blocks[8] after the PCA sent by the MS. Therefore, downlink TBF starts on average 120ms after the PACKET DOWNLINK ASSIGNMENT is sent (6 block periods or 120ms, RRBP = 3 + reaction time = 3 blocks).
Timing advance information gained from the initial random access should still be valid for the downlink TBF. If not, the PCU sends a PACKET TIMING ADVANCE/POWER CONTROL (PTA/PC) in the block indicated in the TBF starting time. An example of the second phase of downlink TBF establishment is illustrated in table 3.7. After successful TBF establishment the downlink RLC blocks are transmitted to the target MS. The relative timestamps reflect the 4-4-5-structure of RLC blocks and idle frames (see figure: 3.1). The following table

TDMA FN	Delay (ms)	Delta (ms)	Link	Message Type
1144245			DL	Immediate Assignment
1144268	106.15	106.15	DL	Packet Downlink Assignment
1144280	161.53	55.38	UL	Packet Control ACK (access burst)
1144281	166.15	4.62	UL	Packet Control ACK (access burst)
1144282	171.12	4.62	UL	Packet Control ACK (access burst)
1144283	221.89	4.62	UL	Packet Control ACK (access burst)
1144294	272.66	50.77	DL	Downlink Data Block
1144299		23.08	DL	Downlink Data Block
1144303		18.46	DL	Downlink Data Block
1144307		18.46	DL	Downlink Data Block
1144312		23.08	DL	Downlink Data Block
				...

Table 3.7: Example of completing DL TBF set-up when MS was standby

shows the message flow and associated timing for downlink TBF set-up for a MS

[8] Refer to [3GPP-TS05.10], clause 6.11.1

in standby state. The estimations are again based on a paging configuration with BS_PA_MFRMS = 3. This corresponds to a maximum delay of 5 multiframes or 1175ms, the minimum delay is also assumed to be 5ms.

Message	from - to	Um Tx Delay (ms)	BSS Delay min/max (ms)	Total Time min/max (ms)
Packet Paging Request	PCU - MS	18.4	5/1175	23.4/1193.4
Channel Request (RACH)	MS-PCU	4.6	5	9.6
IA	PCU - MS	18.4	5/166	28.4/189.4
PRR	MS -PCU	18.4	60	78.4
PUA	PCU - MS	18.4	60	78.4
LLC frame-paging response-	MS - SGSN	20	60	80
SGSN Process Delay				100
IA	PCU - MS	18.4	5/1175	23.4/1193.4
PDA	PCU - MS	18.4	60	78.4
Packet Control ACK (RRBP=3)	MS - PCU	18.4	60	78.4
transmit LLC frames, first RLC block	PCU - MS	20	60	80
Total Delay				
Minimum				**658.4**
Average				**1908.9**
Maximum				**3159.3**

Table 3.8: Estimated DL TBF set-up time for MS in standby state

3.2.2.2 Downlink TBF Set-up in Ready State

The downlink TBF in ready state will be the most common downlink assignment situation. The SGSN knows the serving cell for the MS and instructs the PCU to initiate a TBF establishment in that cell. The downlink set-up process follows the procedure shown for an MS in standby state except the initial paging procedure, which is not required when the serving cell is known.

3.2.2.3 Downlink TBF Set-up During Non-DRX Mode

When a non-DRX[9] enabled MS goes from packet transfer mode to packet idle mode, it monitors all CCCH blocks until the non-DRX timer expires. After expiry of the DRX-timer, the MS enters DRX mode and monitors every BA_MS_MFRMS paging subgroup (see section 3.1.2) only. The DRX mode increases the battery lifetime on expense of higher delay for the paging procedure.
During the non-DRX period, the network can schedule an IA on any of the CCCH blocks belonging to that paging group. This leads to fast downlink TBF set-up.

[9] DRX - Discontinuous Reception

Message	from - to	Um Tx Delay	BSS Delay min/max (ms)	Total Time min/max (ms)
Immediate Assignment	PCU - MS	18.4	5/1175	23.4/1193.4
PDA	PCU - MS	18.4	60	88.4
Packet Control ACK (RRBP=3)	MS - PCU	18.4	60	78.4
First Downlink Block	PCU - MS	20	60	80
...				
Total Delay				
Minimum				**260.2**
Average				**845.2**
Maximum				**1430.2**

Table 3.9: Estimated DL TBF set-up time for MS in ready state

Message	from - to	Um Tx Delay	BSS Delay min/max (ms)	Total Time min/max (ms)
Immediate Assignment on any CCCH block	PCU - MS	18.4	5/166	23.4/184.4
PDA	PCU - MS	18.4	60	78.4
Packet Control ACK RRBP = 3	MS - PCU	18.4	60	78.4
Total Delay				
Minimum				**185.2**
Average				**265.7**
Maximum				**346.2**

Table 3.10: Estimated DL TBF set-up during non-DRX

3.2.3 Access Times at Concurrent TBFs

An established TBF can be used to set-up a TBF in the opposite link direction (concurrent TBF) by sending a TBF set-up message on the PACCH of the already established TBF. Signalling messages on the CCCH are not required for concurrent TBF set-up, what reduces the TBF establishment times.

3.2.3.1 Downlink TBF Set-up During Uplink TBF

The PCU establishes the downlink TBF by sending a PACKET DOWNLINK ASSIGNMENT on the PACCH of the uplink TBF addressed directly to the target MS. This procedure results in very low downlink TBF establishment times, since no IMMEDIATE ASSIGNMENT on the CCCH is required to get the MS listen to the PDTCH.
Examples of downlink TBF establishments via an active uplink TBF are given next. The TS allow two different methods of concurrent TBF establishment. In the first case the downlink TBF is initiated by sending a PACKET DOWNLINK AS-

Message	from - to	Um Tx Delay	BSS Delay min/max (ms)	Total Time min/max (ms)
PDA	PCU - MS	18.4	60	78.4
Packet Control ACK				
RRBP = 3	MS - PCU	18.4	60	78.4
Downlink RLC block	PCU - MS	20		20
Total Delay Average				**176.8**

Table 3.11: Estimated DL TBF set-up time during concurrent uplink TBF

SIGNMENT. The second scenario for setting up a concurrent downlink TBF uses the PACKET TIMESLOT RECONFIGURE (PTR) sent on a PACCH. The network sends a PTR when it reassigns the uplink timeslot allocation in addition to the new downlink TBF establishment.

TDMA FN	Delta (ms)	Link	Message Type
2581110		UL	Uplink Data Block
2581114	18.46	UL	Uplink Data Block
2581118	18.46	UL	Uplink Data Block
2581120	9.23	DL	Packet Downlink Assignment
2581127	32.305	UL	Packet Control Acknowledgement (access burst)
2581128	4.615	UL	Packet Control Acknowledgement (access burst)
2581129	4.615	UL	Packet Control Acknowledgement (access burst)
2581130	4.615	UL	Packet Control Acknowledgement (access burst)
2581131	4.615	UL	Uplink Data Block
2581136	23.075	UL	Uplink Data Block

Table 3.12: Example of DL TBF establishment during concurrent uplink TBF

TDMA FN	Delta (ms)	Link	Message Type
2587861		UL	Uplink Data Block
2587865	18.46	UL	Uplink Data Block
2587871	27.69	DL	Packet Timeslot Reconfigure
2587874	13.85	UL	Uplink Data Block
2587876	9.23	DL	Packet Uplink ACK/NACK
2587878	9.23	UL	Packet Control Acknowledgement (access burst)
2587879	4.615	UL	Packet Control Acknowledgement (access burst)
2587880	4.615	UL	Packet Control Acknowledgement (access burst)
2587881	4.615	UL	Packet Control Acknowledgement (access burst)
2587897	73.48	DL	Downlink Data Block
2587902	0.00	DL	Downlink Data Block

Table 3.13: Example of DL TBF at concurrent uplink TBF

Conspicuous in both examples above is the transmission of uplink data blocks without downlink blocks carrying the USF bits. This is an indication for fixed uplink allocation. In the beginning of a TBF, the PCU assigns an allocation bitmap, which tells the MS the blocks it is allowed to transmit its data. So, no USF in a downlink RLC block is required to enable the uplink transmission. I mention fixed uplink allocation for the sake of completeness, but actually, no BSS infrastructure supplier supports fixed allocation any more. Furthermore, fixed allocation was removed in Release 5 of the 3GPP specifications [3GPP-TS44.060-540].

3.2.3.2 Uplink TBF Set-up During Downlink TBF

The MS requests an uplink TBF within a PACKET DOWNLINK ACKNOWLEDGMENT (PDAN) sent to the network. The PCU starts establishing the UL TBF by sending a PACKET UPLINK ASSIGNMENT on the PDCH on which the DL TBF is established. That is, without requiring the MS to perform a one or two-phase uplink access on the RACH, leading to very low TBF establishment times. The channel request packed into the PDAN is illustrated in appendix C, table a-9.

Message	from - to	Um Tx Delay	BSS Delay min/max (ms)	Total Time min/max ms)
PDAN	MS - PCU	18.4	60	78.4
PUA	PCU - MS	18.4	60	78.4
Packet Control ACK RRBP = 3	MS - PCU	18.4	60	78.4
Total Delay Average				**235.2**

Table 3.14: Estimated UL TBF set-up time during concurrent downlink TBF

3.3 Conclusion from TBF Set-up Times

Uplink TBF establishment can either be done using one-phase access or two-phase access. The uplink TBF establishment procedure is requested by the MS, depending on the MS multislot class and the amount of data to be transmitted in the uplink. One-phase access always requests one uplink PDTCH only. Requesting more than one timeslot in the uplink requires a two-phase access procedure. Two-phase access takes considerably longer compared to the one-phase access. The additional delay has to be justified by reduced transmission time on multiple timeslots.
Downlink TBF set-up delay is strongly depending on the mobility context or DRX mode. Actually, most of the currently used applications over GPRS are MS originated. Therefore, the MS will always be in ready mode when a downlink TBF is set up by the network. Downlink TBF set-up in standby state happens rarely for

common applications. New applications like *Push to Talk* will introduce network initiated GPRS sessions and increase the number of downlink TBF set-up for MS in standby state. Reducing the high downlink TBF set-up times by keeping the MSs in ready state will result in massive signalling overhead, reducing the available GPRS bandwidth.
TBF establishment using concurrent TBFs generally is a very fast method. Applications using the acknowledgement driven TCP/IP protocols will predominantly use TBF set-up via the concurrent TBF. Especially uplink TBF set-up is improved since due to timing restrictions many multislot classes cannot listen to all available Rx timeslots of a downlink TBF and send a RACH on the CCCH simultaneously (see chapter 4). TBF set-up times are determined by the TDMA frame structure. The network as well as the MS is only allowed to send particular signalling messages at specified point in time, according to the CCCH frame structure. The rigid frame structure also introduces high variability of TBF set-up times. Improving the TBF set-up delay is not possible without changing the CCCH channel structure, what in turn affects circuit-switched call set-up.

Chapter 4

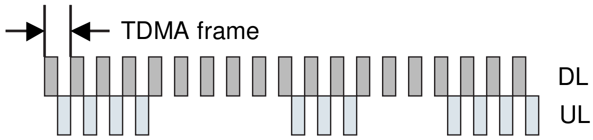
GPRS Multislot Usage

In this chapter, I discuss performance implications of data transmission on multiple timeslots. Beside practical limitations such as mobile battery consumption and MS hardware expenses timing restrictions in the MS limit the multislot usage.

4.1 Timing Restrictions of MS Multislot Operation

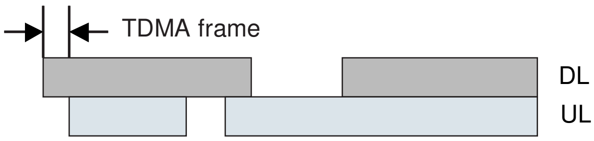
During packet transfer mode the MS has to listen to multiple timeslots in either downlink or uplink and measure the received RF signal level on the BCCH carriers of the serving cell and the neighbouring cells. The synchronisation after switching between transmit- and receive-state and the period to perform signal-strength-measurements are the parameters, which limit the MS multislot capability. While channel measurements are not critical in packet idle mode a strict timing is required during packet transfer mode to switch between Tx and Rx and perform the measurements. The strictest rules apply during concurrent TBFs when all channels supported by the MS multislot class are allocated to TBFs. At least 5 received signal level measurement samples are required for a valid channel measurement, see [3GPP-TS05.02]. 29 MS multislot classes have been defined in the technical specification [3GPP-TS05.02] in order to organise the timeslot assignment for multislot capable mobile stations. A complete list of MS multislot classes is shown in appendix D. The MS multislot classes are grouped into three categories.

- **Semi-duplex classes, MS class 1 through 12:**
  The multislot classes 1 to 12 belong to the group of type 1 terminals. These classes are following the multiplexing strategy of GSM voice mobile stations and they are not able to transmit and receive simultaneously. All available GPRS MS classes are type 1, nowadays are class 8 and 10 most common. The semi-duplex characteristic is illustrated below.



12 Semi-duplex MS Multislot Classes

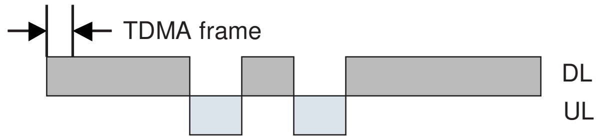
- **Full-duplex classes, MS class 13 through 18:**
The specification defines 6 full-duplex multislot classes which support simultaneous operation of up to 8 timeslots in uplink and downlink. Tx and Rx are not required to be coordinated. This requires separated transmit and receive front ends in the MS (type 2 terminals). Currently full duplex MS are commercially not available and it appears that due to the improved hardware requirements a common spreading of type 2 MS is not very likely to happen for GPRS. However, type 2 terminals supporting EGPRS could be deployed, depending on market requirements and the success of deploying EGPRS.



6 Full-duplex MS Multislot Classes

- **Half-duplex classes, class 19 through 29:**
Half-duplex MS are type 1 MSs, which are not able to transmit and receive simultaneously but support operation of up to 8 timeslots either in downlink or uplink direction. These types of multislot MS were intended to transmit or receive continuously for long intervals. Therefore, interactive applications will not benefit from half-duplex MS. The advantages are lower power consumption and lower price compared to full-duplex MS and they offer particularly more throughput than semi-duplex MS. Half-duplex operation requires use of fixed uplink allocation [3GPP-TS04.60]. Fixed allocation has been introduced mainly with the purpose of extending the uplink capability of the MS. Fixed allocation controls the contention resolution for different MS accessing the uplink by assigning an uplink allocation map during the uplink TBF establishment. The MS is not required to monitor the downlink RLC blocks for USF bits. Thus, the MS can send or receive on up to 8 timeslots. That would not be possible using dynamic allocation. Fixed allocation has been removed from the 3GPP technical specifications as already shown in 3.2.3. Hence, class 19 to class 29 MS are effectively obsolete.

I assume that other multislot configurations than class 1 to 12 will not be available for GPRS since multislot operation in general causes high heat dissipation and high



11 Half-duplex MS Multislot Classes

battery consumption while the MS is in packet transfer mode. Furthermore, the sufficient provision of GPRS resources is difficult in the shared voice and GPRS infrastructure. Finally multislot operation is challenging the BSS in terms of delay requirements, as to be seen in section 4.2, below.

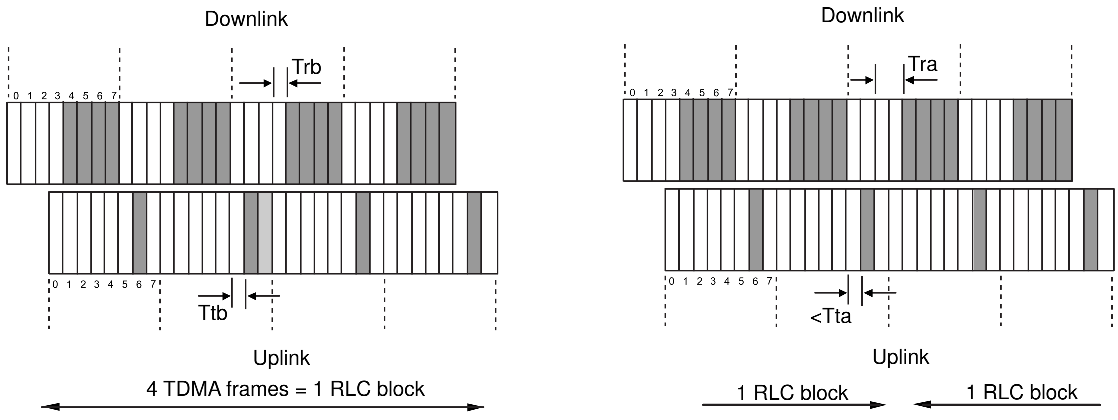
4.1.1 Mobile Station Timing for Four Downlink PDCHs Allocated

The maximum multislot assignment of type 1 MS is limited to four PDCH. The multislot classes 8, 10, 11 and 12 are able to receive data on four timeslots while concurrent transmission is limited to one PD channel. The requirement to re-synchronise between Tx and Rx and to perform measurements results in specific timeslot assignment patterns. The timing requirements are fulfilled when either uplink timeslot 6 or 7 is used to transmit during concurrent TBFs, as to be seen in figure 4.1-a.). All four timeslots (TN 4,5,6 and 7) are allocated in the downlink and no signal level measurement is done. The preferred uplink timeslot assignment is shown in figure 4.1-b.). Only this allocation pattern allows to perform a signal strength measurement when the MS moves from transmit to receive state, (T_{ra}=2).

4.1.2 MS Timing for Multiple Uplink PDCHs Allocated

The timing requirements for uplink data transmission depend on the medium access mode. In standard *dynamic allocation mode*, the MS is required to decode the USF bits - included in the MAC header of downlink RLC/MAC blocks - for every uplink timeslot, the MS intends to transmit data. Using standard dynamic allocation will restrict type 1 multislot capable MS to two uplink timeslots only, in order to provide sufficient time decode the USF bits in the corresponding downlink timeslots and switch to transmit state. Possible channel combinations for a class 12 MS at dynamic allocation are shown in figure 4.2, figure a.). At class 12 the maximum number of timeslots simultaneously active is 5, subsequently are at most three downlink PDCHs active when two Tx channels are used. Again, T_{ra}=2 applies.

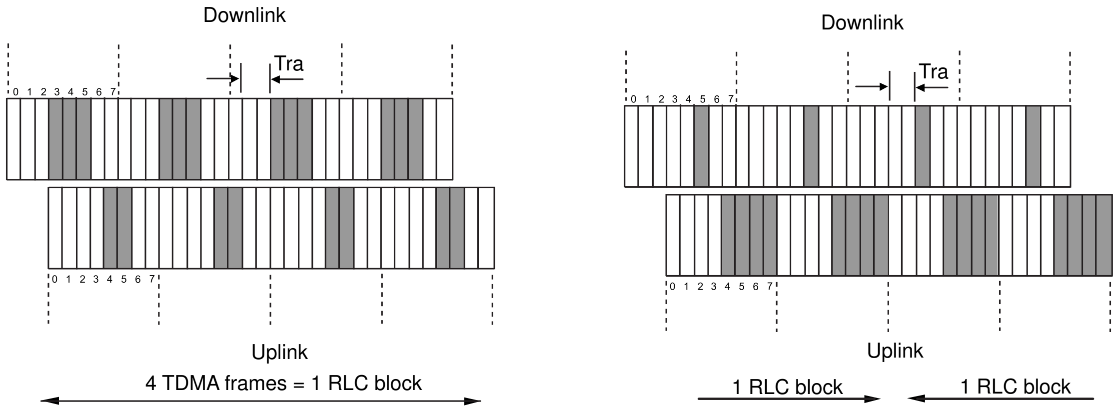
This limitation was intended to be addressed in the *fixed allocation mode*. Possible channel combinations for a class 12 MS in fixed allocation are shown in figure 4.2, figure b.). As already stated in 3.2.3 has fixed allocation been removed from the



a. b.

T_{ti} = Time to get ready to transmit
T_{ri} = Time to get ready to receive
Index i = a, A measurement report is required to be made by the MS
Index i = b, No measurement report has to be made by the MS

Figure 4.1: Timeslot allocation for a class 8 MS



a. b.

Figure 4.2: Dynamic vs. fixed allocation for a class 12 MS

3GPP TS. The only possibility to transmit on more than two PDCHs is to use *extended dynamic allocation*. An MS operated in extended dynamic allocation has to detect an assigned USF value on the lowest numbered downlink PDCH assigned, and that allows the MS to transmit on that PDCH and all higher numbered, assigned PDCHs [3GPP-TS05.08].
Type 1 MS in enhanced dynamic allocation mode are theoretically able to utilise

four uplink PDCHs. Unfortunately, the definition of multislot class 12 timing requirements does not allow to utilise all four uplink channels when signal strength measurements have to be done simultaneously. The specification [3GPP-TS05.02], clause 6.4.2.2, "Multislot configurations for packet-switched connections" defines that the MS in extended dynamic allocation has to perform signal strength measurements during T_{ra}, when the MS moves from transmit to receive state. Looking at 4.3, figure a.) will show the minimum period of T_{ra} cannot be meet at four Tx channels, only at three uplink PDCH allocated is T_{ra}=2 fulfilled, see 4.3, figure b.). I could not verify a proposed solution for this issue since enhanced dynamic

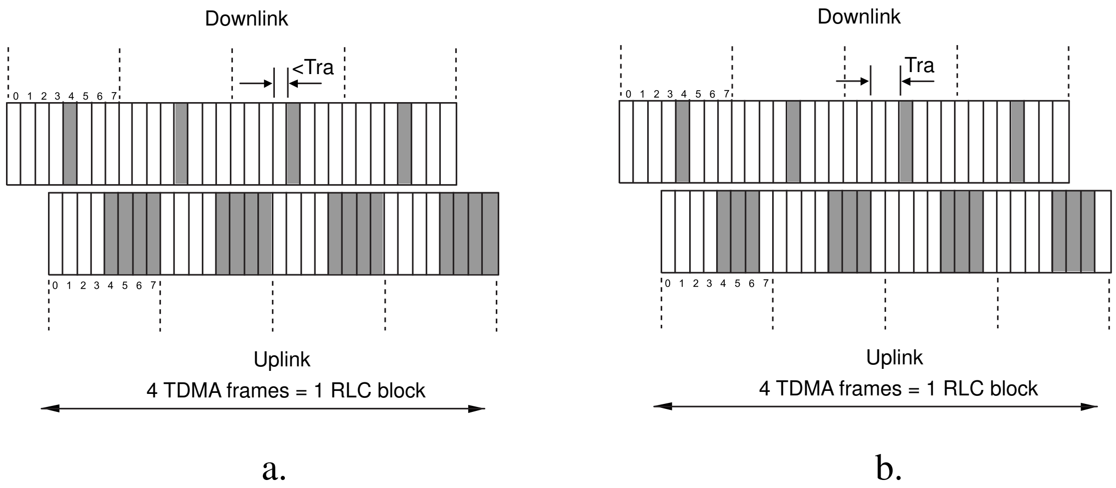


Figure 4.3: Timeslot allocation for a class 12 MS

allocation is not deployed yet. It appears that a network supporting extended dynamic allocation will have to limit uplink multislot assignment of type 1 MS to three Tx channels.

4.2 RLC Acknowledged Mode and Multislot Operation

The limitations in multislot usage discussed above are mainly a result of MS timing restrictions. An additional requirement for multislot operation is low latency within the BSS.

In acknowledged mode[1], the RLC layer ensures in-order delivery of RLC/MAC blocks to the LLC layer. It uses a selective ARQ and a sliding window with a window size of 64, see section 6.1.2. The transmitting side receives positive acknowledgements for correctly received RLC/MAC blocks, erroneously received blocks cause a negative acknowledge to be sent back to the transmitting entity. Once block errors happen, the transmission of new RLC blocks will be stuck when one block remains unacknowledged and 63 new blocks have already been sent. Thus, the RLC layer is *stalled*. Following this, the sender has to try to deliver the unacknowledged block (or multiple unacknowledged blocks) and receive an acknowledgement before the transmission of new blocks can resume. During the period of time when the RLC layer inevitably stalls, the data throughput will come to naught.

Considering a data transmission on four timeslots, a RLC stall will happen when within 16 block periods ($16 * 4 = 64$) no positive acknowledgement for the retransmission is received by the sending entity. The following example illustrates a data transmission at four timeslots and an acknowledgement interval of four block periods. A successful retransmission within 16 block periods requires a round trip

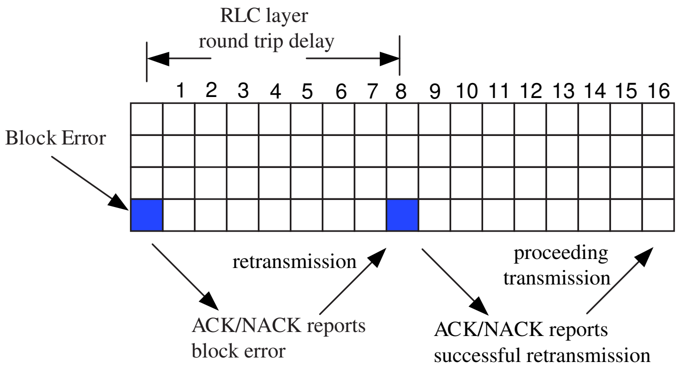


Figure 4.4: ACK/NACK interval vs. RLC window stall

delay smaller than 8 block periods. The round trip delay comprises the delay in the BTS, BSC and PCU, the delay on the links within the BSS, the transmission time on the radio link and finally the period from receiving an erroneous block until sending the ACK/NACK. From figure 4.4 can be concluded that the maximal

[1] RLC acknowledged mode is commonly used, even for streaming applications

interval for sending ACK/NACK messages is of high importance to avoid RLC window stalls. With an RLC window size of 64 and four PDCHs a single block error can be retransmitted without provoking a RLC stall. When the transmission fails a second time, the RLC window is stalled for the duration of one BSS round trip delay. The interaction of RLC acknowledged mode and multislot usage have also been identified by [TABO2002]. The simulations show decreasing MAC buffer occupancy[2] with increasing number of timeslots used.
Using multiple PDCHs requires low acknowledgement intervals to avoid RLC layer stalls and ineffective utilisation of the RLC layer bandwidth. On the other hand has to be considered that ACK/NACK control messages reduce the effective bandwidth of the opposite link direction, since the block period cannot be used for data transmission.

4.3 Resource Sharing for Multiple Multislot MS

Assigning multiple receive or transmit timeslots to a mobile station has to be done within the MS timing restrictions to offer the MS enough idle timeslots to perform signal level measurements and synchronisation after switching between send and receive state. As already shown in section 4.1, for every type 1 terminal just limited timeslot allocation schemes fulfil the timing restrictions. The graphs in figure 4.5 illustrate preferred timeslot allocations for different common multislot classes. The class 1 MS, supporting a single timeslot in both directions has to be

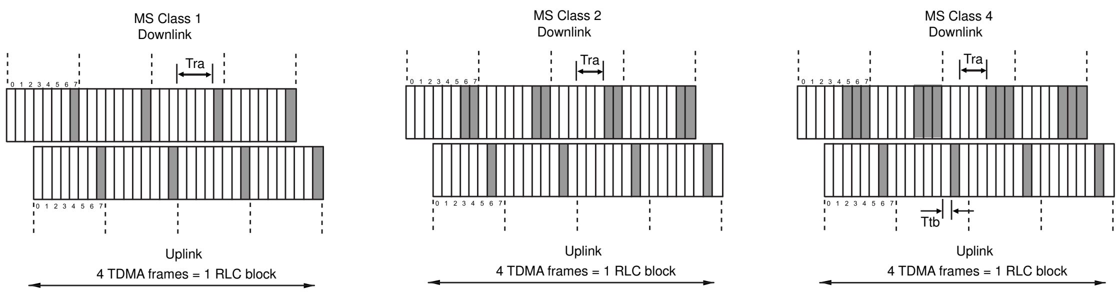


Figure 4.5: MS timeslot allocations

assigned the same timeslot number for uplink and downlink when using dynamic allocation, class 2 MS can use both of the uplink timeslots equally, whereas the class 4 MS is just able to use one timeslot assignment pattern. Class 8 MS also have to use one specific allocation scheme to allow signal level measuring when the maximum number of supported timeslots are assigned, see figure 4.1 or refer to figure 4.3 for a class 12 MS. These restrictions in allocation schemes are limiting

[2] I.e. MAC queue length is not reaching the full RLC window size and the radio link is not fully utilised

the utilisation of a GPRS cell serving multiple mobile stations. Once a GPRS cell with four PDCH, serving a single class 8 MS is receiving a request for resources from a second class 8 MS, the network either decides to reconfigure the first class 8 MS to use two PDCH only and assign the two freed PDCHs to the second MS, or it leaves all four PDCH assigned to the first MS and shares the available PDCHs with the second MS. Both possibilities are depicted in figure 4.6. Assigning a

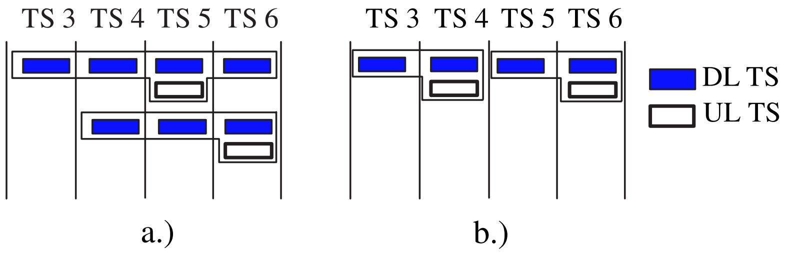


Figure 4.6: Examples of timeslot sharing schemes

timeslot allocation according to case a.) with interleaving downlink TBFs reduces one of the class 8 MS to use three timeslots at most when the uplink TBFs are not forced to share the same timeslot. Considering continuous data transfer in both directions the allocation of case a.) offers 2.5 downlink timeslots to one MS and 1.5 timeslots to the other MS, respectively, when assuming equal resource sharing on RLC/MAC layer.
The timeslot allocation algorithm in the network has to consider the multislot capability of all MS having active TBFs in the cell, but also the duration of a TBF. Reconfiguring an MS occupying all available PDCHs in the cell when another MS is requesting PDTCH can be a waste of resources when the second MS is just doing a routing area update requiring a short downlink assignment duration. Additionally has to be considered that the multislot capability is not known by the network initially, when the MS is requesting a one-phase TBF set-up, see 3.2.1.2. The complexity of the channel assignment in the network will be considerably higher with raising GPRS data traffic or when implementing quality of service.

Another aspect to be considered when assigning timeslots to multiple MS has already been described in section 5.2. An MS with high uplink traffic, e.g. when sending an MMS (Multimedia Messaging Service) or email, will be able to increase the uplink throughput when a lower number of downlink PDCHs is assigned. This effect is even more distinct when one of the following multislot classes is used. These multislot classes have an interdependent number of uplink and downlink timeslots supported, e.g. a class 12 MS can be operated as a class 8 MS with four downlink and one uplink timeslot, or it can be three uplink timeslots assigned but just two downlink timeslots.
Proper assignment of timeslots according to the traffic volume sent in each link

direction requires a traffic-balancing algorithm in the network in order to choose the optimal timeslot allocation for concurrent TBFs.

MS multislot class	Maximum number of timeslots Rx	Tx	sum
3	2	2	3
6	3	2	4
7	3	3	4
10	4	2	5
11	4	3	5
12	4	4	5

Table 4.1: Type 1 multislot classes with interdependent TS allocation

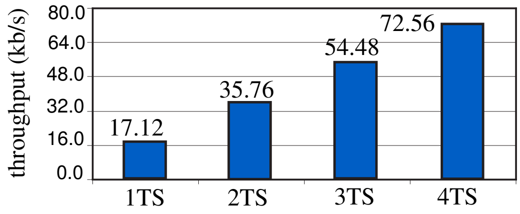
Chapter 5

Throughput on Multiple Timeslots

Throughput measurement results using FTP show a disproportionate growth in the application layer throughput when multiple timeslots are used, i.e. the throughput per timeslot increases when using multiple timeslots. An opposed effect of multislot usage can be seen for uplink FTP throughput. Uplink throughput decreases with higher number of downlink PDCHs assigned, even when only one PDCH is used in the uplink.
In this chapter the interaction of downlink throughput of multislot usage followed by an analysis of the dependency of uplink throughput and multislot usage are discussed.

5.1 Downlink Throughput as Function of Multislot Usage

The effect of higher throughput per timeslot is illustrated in figure 5.1, showing results for FTP downlink transfer of a 400kB file. The throughput measurements were done with exactly the same set-up, a clean radio environment. A class 8 MS was used and on the network side the number of GPRS timeslots was increased from one to four. The network supports coding scheme CS-4. The FTP client reports download times and throughput results as shown in figure 5.1. Following effect can be observed. The throughput when using four PDCH is about 0.96kbit/s per TS higher compared to the single timeslot case. In the following, I compare the single and the four-timeslot case to determine the performance gain of multislot usage as suggested by the results above. The reason for the performance gain and the interaction of the FTP application and the RLC layer is analysed. The TCP throughput graphs shown in figure 5.2 again illustrates this effect. The instantaneous throughput after TCP slow start is about 73.6kbit/s for the four-timeslot download, whereas the single timeslot case gets up to 17.6kbit/s.



file size 400kB, CS-4 used				
	1 TS	2 TS	3 TS	4 TS
duration (s)	191.81	91.66	60.18	45.15
throughput (kb/s)	17.12	35.76	54.48	72.56

Figure 5.1: Throughput for different multislot classes

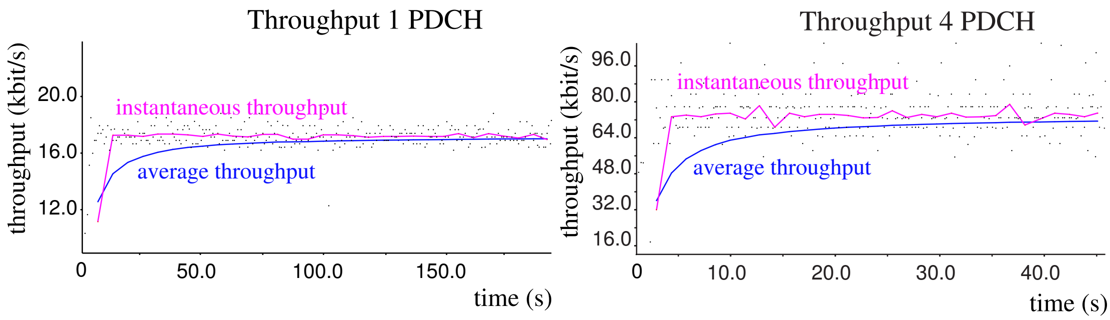


Figure 5.2: TCP throughput, one PDCH vs. four PDCH

Firstly, the way the FTP client measures the throughput has to be verified. Looking at the single timeslot transfer shown in table 5.1 it can be seen that the throughput calculation attained by the FTP client is based on the duration from sending the FTP message that the binary data connection is established, until sending the FTP protocol message for binary transfer complete. Considering the entire duration of the FTP session, the throughput results look like shown below. Including the TCP and FTP signalling at the beginning and the end of the file transfer, the download at one PDCH lasts just slightly more than four times the download duration at four PDCH.

file size 400kB, CS-4 used				
	1 TS	2 TS	3 TS	4 TS
total FTP session (s)	195.23	94.31	63.12	48.39
throughput (kbit/s)	16.8	34.72	51.92	67.68

5.1.1 Latency Difference for FTP Signalling

Using four PDCH the signalling on application layer is slightly faster than using one PDCH. The TCP trace for the 1 PDCH transmission in table 5.1 shows a

Frame Number	Total-Time (s)	Link	Bytes	Protocol	Protocol Summary
1	0	DL	78	FTP	Port=1083 145,151,1,49,4,73
2	1.192	UL	84	FTP	200 Port command successful.
3	1.202	DL	69	FTP	Port=1083 RETR 400kB
4	1.813	UL	66	TCP	D=1097 S=20 SYN SEQ=97348427 LEN=0
5	1.823	DL	66	TCP	D=20 S=1097 SYN ACK=97348427 SEQ=265006742 LEN=0
6	1.883	UL	54	TCP	D=1083 S=21 ACK=4246958620
7	2.423	UL	54	TCP	D=1097 S=20 ACK=265006743
8	2.574	UL	131	FTP	150 Binary data connection for 400kB (409600 bytes).
9	2.734	DL	54	TCP	D=21 S=1083 ACK=816589850
10	3.866	UL	1460	FTP	Port=1097 FTP DATA
				...	
574	194.381	DL	946	FTP	Port=1097 FTP DATA
575	194.391	UL	54	TCP	D=20 S=1097 ACK=973893872
576	194.391	UL	54	TCP	D=20 S=1097 FIN ACK=97389387 SEQ=265006743 LEN=0
577	194.421	DL	85	FTP	226 Transfer complete.
578	194.431	UL	54	TCP	D=21 S=1083 ACK=816589881
579	195.233	DL	54	TCP	D=1097 S=20 ACK=265006744

Table 5.1: FTP signalling and data transfer at 1 PDCH

signalling overhead of 3.866 seconds before and 0.842 seconds after the pure data transmission. Whereas from the TCP trace for the 4 PDCH transmission shown in table 5.2, a signalling overhead of 2.834 seconds before and 0.851 seconds after data transmission can be measured. The initial set-up of the FTP session and the TCP handshake includes uplink and downlink traffic, in this case, the increased downlink bandwidth improves the signalling delay and the FTP on 4 PDCH saves 1.03 seconds. The closing of the FTP session mainly consists of sending an ACK and the FIN ACK in uplink direction, answered by an ACK in downlink direction. So, increased downlink bandwidth is not significantly contributing to a reduced signalling delay, in either both cases are similar delays of about 0.8 seconds seen. The improvement in downlink direction is compensated by the variation of UL TBF set-up delay.

The pure data transmission from receiving the first TCP data packet to the last TCP packet without TCP handshake signalling lasts 190.515 seconds at one PDCH and 44.855 seconds for the four PDCH download. Still, a throughput improvement at 4 PDCH is seen; 11.1 seconds are saved compared to the single timeslot download.

Five main effects contributing to this difference are identified in the following.

Frame Number	Total-Time (s)	Link	Bytes	Protocol	Protocol Summary
1	0.000	UL	78	FTP	Port=1083 123,123,123,12,4,60
2	1.051	DL	84	FTP	200 PORT command successful
3	1.061	UL	69	FTP	Port=1083 RETR 400kB
4	1.712	DL	66	TCP	D=1084 S=20 SYN SEQ=8286205 LEN=0
5	1.722	UL	66	TCP	D=20 S=1084 SYN ACK=828620 SEQ=4270996746 LEN=0
6	1.772	DL	54	TCP	D=1083 S=21 ACK=424695850
7	2.333	DL	54	TCP	D=1084 S=20 ACK=427099674
8	2.534	DL	131	FTP	150 Binary data connection for 400kB (409600 bytes)
9	2.684	UL	54	TCP	D=21 S=1083 ACK=816589436
10	2.834	DL	1460	FTP	Port=1084 FTP DATA
				...	
465	47.288	UL	54	TCP	D=20 S=1084 ACK=8290278
466	47.459	DL	1460	FTP	Port=1084 FTP DATA
467	47.539	DL	946	FTP	Port=1084 FTP DATA
468	47.549	UL	54	TCP	D=20 S=1084 ACK=829030171
469	47.549	UL	54	TCP	D=20 S=1084 FIN ACK=829030 SEQ=4270996747 LEN=0
470	48.390	DL	54	TCP	D=1084 S=20 ACK=42709967

Table 5.2: FTP signalling and data transfer at four PDCH

5.1.2 RLC/MAC Layer Signalling

The radio interface implementation of the analysed network sends four types of signalling blocks to the MS during a file transfer: PACKET UPLINK ASSIGNMENT, PACKET UPLINK ACK/NACK, FINAL PACKET UPLINK ACK/NACK and finally POWER CONTROL/TIMING ADVANCE. Each of those control messages take one downlink block to be sent instead of data blocks. That reduces the available downlink bandwidth. This bandwidth loss is considerably high due to the nature of a TCP transmission. Sending frequently small TCP ACK packets in the uplink causes frequent set-up and release of UL TBFs accompanied by PUA, PUAN, FPUAK. For the investigated transmissions, following figures were measured. According to table 5.3, 948 RLC blocks at 1 PDCH correspond to 948 block periods or 18.58 seconds, while 265 blocks at four PDCH correspond to $\frac{265}{4} = 66.25$ block periods or 1.325 seconds. Additionally, by comparing the number of PUA messages can be derived from table 5.3 that three to four times more uplink TBFs are established to acknowledge the same amount of data sent in downlink. This effect is explained by the delayed acknowledge implemented in the FTP client.

	1TS	4TS
PUA - Packet Uplink Assignment	281	79
FPUAK - Final Packet Uplink ACK/NACK	281	79
PUAN - PACKET UPLINK ACK/NACK	246	74
PCTA - POWER CONTROL TIMING ADVANCE	121	33
Sum:	929	265

Table 5.3: RLC/MAC signalling blocks for 1 and 4 PDCH

5.1.3 Utilisation of TCP Delayed Acknowledgement

As specified in [RFC1122], TCP uses delayed acknowledgement to reduce the number of packets sent on the media. The client PCs TCP/IP stack takes a common approach to implement delayed ACK. As data is received by TCP on a connection, it only sends an acknowledgement back if one of the following conditions is met.

- No ACK was sent for the previous segment received.
- A segment is received, but no other segment arrives within 200 milliseconds for that connection.

The DL data is transmitted in segments of 1532 bytes (including TCP, IP, LLC header. Each of these segments takes about 29 RLC blocks to be transmitted at coding scheme CS-4, which converts to 580ms at one PDCH and 145ms at four PDCH under idealised conditions, i.e. no overhead signalling. Delayed TCP acknowledge is possible during a four PDCH transmission, at one PDCH the delayed acknowledgement as described above cannot be utilised at all. As the amount of RLC layer signalling sent in downlink is specified by the networks RLC layer implementation and depends on the data volume sent in multiple uplink TBFs, the transmission times will be slightly higher when delayed acknowledgement is not used. The next two figures 5.3-a.) and -b.) show the effect of how this time difference has affect on delayed acknowledgement. A comparison of the two figures illustrates the difference. Using one PDCH, every TCP segment causes a separate TCP ACK to be sent, while during the four PDCH transmission about three out of four TCP ACK are sent for two segments. The duplicate acknowledge is seen in an unsteady sending of acknowledges in the four PDCH case (figure 5.3-b.)). Figures from the example transmissions are given in table 5.4. As a result, TCP

	1 PDCH	4 PDCH
TCP Segments received	279	285
TCP ACK sent	279	175

Table 5.4: Number of TCP-ACK for 1 PDCH and 4 PDCH

delayed acknowledgement reduces the number of TCP ACK to send and therefore

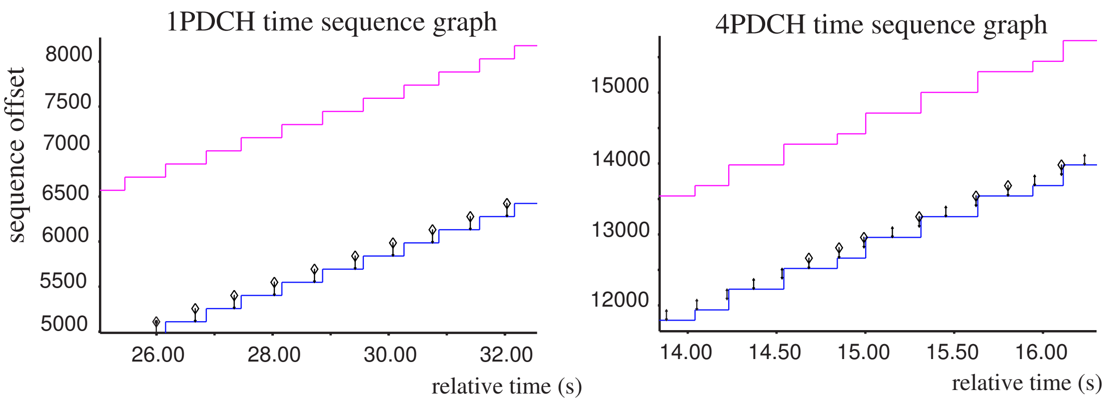


Figure 5.3: Effect of delayed acknowledgement

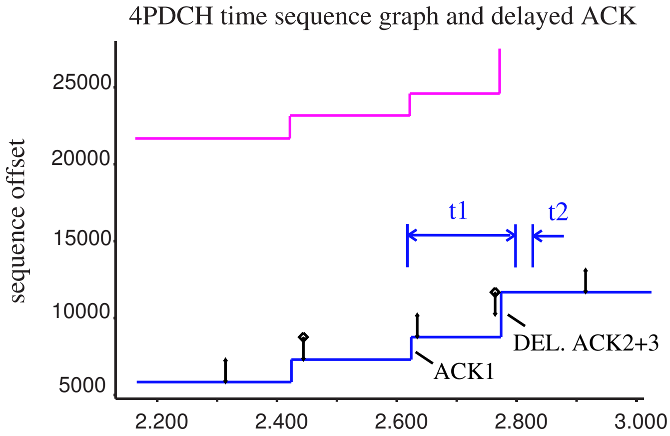
the number of UL TBFs to be established, which further increases the channel efficiency when four PDCH and CS-4 are used.
Looking at table 5.4 shows that more TCP segments are sent when four PDTCH are used and only 79 UL TBFs are established when 175 TCP ACK are sent UL.

5.1.4 UL TBF Set-up Time

Besides the delayed acknowledgement, the reduced number of uplink TBFs is another reason for the higher throughput results at four PDCH. As to be seen in table 5.4 are 175 ACK sent on TCP layer, but only 79 UL TBFs are used to transmit these packets.
During the time an UL TBF is set up and subsequently the TCP ACK is transmitted in this TBF, the next TCP ACK will already be queued in the MS and two TCP packets can be transmitted in one UL TBF. According to section 3.2.3.2, the average uplink TBF set-up delay lasts about 230ms and the transmission of one TCP ACK at CS-4 can be done within 2 block periods when considering a single uplink timeslot transmission. The example in figure 5.4 shows the transmission of even three TCP ACK in one uplink TBF. A single acknowledgement (ACK1) is sent after delay acknowledgement timer expires. Shortly after sending the ACK, the next TCP packet is arriving at the FTP client, the ACK (ACK2) is delayed. Before the delayed acknowledgement timer expires the next TCP packet arrives, triggering a delayed acknowledgement (ACK2+3). Additionally, at that point in time the uplink TBF for ACK1 is set up and the ACK1 together with the delayed acknowledgement already queued in the MS are sent in that TBF.

5.1.5 TCP PUSH Flag

TCP was designed for fixed networks with relatively low RTT compared to high bandwidths, so one of the bottlenecks in a transmission would be the receivers abil-



t1 . . . UL TBF set-up time

t2 . . . time to transmit 2 TCP ACK

Figure 5.4: UL TBF and delayed acknowledgement

ity to process the incoming data, i.e. the receive buffer. The receiver's advertised TCP window gives the sender an indication of how much data he is allowed to send into the transport pipe to the receiver and therefore also the receiver's buffer. Normally the receiver's buffer is emptied when a specific threshold of buffer level is reached; the data will be delivered to the higher layers. If the sender wants to indicate that there is still a lot of data to go, but no space for it, he sets the PUSH flag. Another common reason to set the push flag is to indicate the end of the transmission. The push flag orders the receiver to empty his receive buffer immediately to the higher layers and make space for further data [RFC793].
The FTP server used for comparison of one PDCH and four PDCH sets the PUSH flag when it recognises the window size comes close to the advertised window size. It will try to free the buffer in the receiver, by sending a packet that will fit in the receive buffer for sure, but also setting the PUSH flag in the TCP header. If it is not possible to send an MSS (Maximum Segment Size) sized segment because the advertised window size is already filled up, it will send a smaller segment that fits. Due to TCP delayed ACK, in the four PDCH transmission, the ACK come back to the sender in an unsteady frequency. This, combined with the higher bandwidth, causes the above-explained scenario to happen more often. This results in a higher overhead of the four PDCH transmission and in fact more data is transmitted over the radio interface, as well as a higher number of segments is transmitted. The overhead difference is in the range of some hundreds of bytes and therefore causes a difference of about 100ms or 5 block periods to disadvantage of the four PDCH scenario.

	Number of Pushed Segments	Number of Pushed Segments below MSS
1 PDCH	19	9
4 PDCH	24	14

Table 5.5: Pushed TCP segments and MSS

1TS:	4TS:
total idle blocks 225	total idle blocks 1202
total new blocks 8607	total new blocks 8627
total NAK blocks 0	total NAK blocks 0
total PAK blocks 72	total PAK blocks 236
sum of blocks 8904	sum of blocks 10065

Table 5.6: PCU output UL/DL TBF statistics

5.1.6 TCP Slow Start

Another significant difference appears when comparing the slow start behaviour of one PDCH against four PDCH. While the server's maximum initial congestion window of four segments is more than sufficient for one PDCH, i.e. there is always enough data in the queue to achieve a continuous transmission right from the start. In case of four PDCH, the bandwidth is high enough so that the complete initial window is transmitted before the following segments arrive in the queue. During

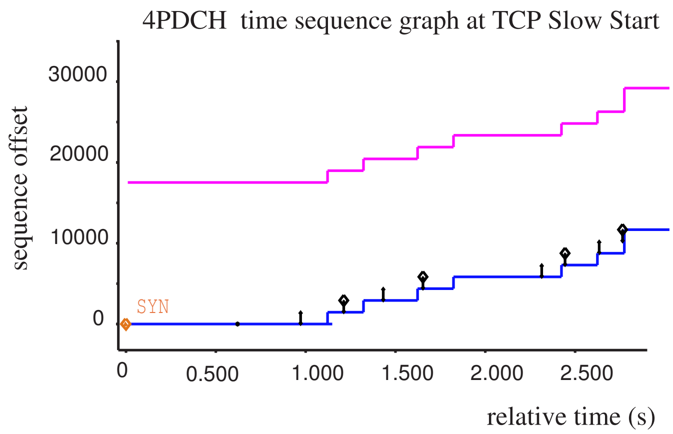


Figure 5.5: Effect of TCP slow start at 4 PDCH transmission

TCP slow start the offered GPRS bandwidth is not fully utilised and a delay of about 300ms to 400ms is introduced, whereas the one PDCH case shows a steady raise in data volume received throughout the entire download.

5.1.7 Resulting Influence - One PDCH vs. Four PDCH

The resulting effect on 4 PDCH transmission are summarised below:

- The major part of the throughput per TS difference between four PDCH and one PDCH is caused by the RLC/MAC signalling overhead.
- TCP Push flag causes TCP packets below MSS (Maximum Segment Size) to be sent, increasing the TCP overhead and number of RLC data blocks respectively. Pushed TCP packets are more frequently at four PDCH
- The RLC/MAC signalling overhead is reduced by TCP delayed acknowledgement, which can be utilised at four PDCH and coding scheme CS-4
- Uplink TBF set-up works in favour of delayed acknowledgement.
- At bandwidths achieved with four PDCH and CS-4, the maximum initial slow start window of four segments is not big enough to circumvent the effects of TCP slow start.

Following data transmission times have to be considered when calculating the throughput per timeslot: The one PDCH FTP achieves an effective throughput of

	1 PDCH	2 PDCH
Transmission of TCP Packets	190.515	44.855
Effect of TCP Push Flag		-0.1
Effect of TCP Slow Start		-0.4
RLC/MAC Signalling	-18.58	-1.325
	171.935	43.03
Total Ttransmission Time:	171.935	172.12

Table 5.7: Download times per timeslot

2382 byte/s or 19.056kbit/s, at four PDCH the effective throughput per timeslot is 2380 byte/s or 19.040kbit/s. The difference in radio link throughput is mainly because of the estimated TCP slow start. The error made is in the range of 2 RLC blocks at CS-4 per file transfer.
The effect of bandwidth reduction as a result of RLC/MAC layer signalling for the reverse link is more distinct for data transmission on one timeslot compared to data transmission on multiple timeslots. While this result is also dependent on the actual implementation of RLC layer signalling in the PCU, the traffic is generated in the uplink direction clearly affects the downlink bandwidth and the maximal throughput in downlink direction.

5.2 Uplink Throughput as Function of MS Multislot Class

A similar effect as described above can also be seen for data transmission in uplink direction, even when doing a file transfer using an MS with one Tx channel but multiple Rx channels, i.e. MS class 2, 4 or 8. The effect on uplink throughput is a result of the different interval of polling for PDAN when multiple downlink timeslots are used. Especially when using higher coding schemes and multiple timeslots the polling interval has to be reduced to receive information on RLC layer block errors as early as possible and to avoid RLC transmit window stalls. According to [3GPP-TS04.60], the maximum number of unacknowledged RLC blocks is limited to 64 blocks. Comparing a data transmission on a single timeslot with one on four timeslots, theoretically a fourth of the polling interval required is to acknowledge the same amount of RLC data blocks. Considering a four PDCH download, an unacknowledged downlink RLC data block has to be re-transmitted within 16 block periods or 320ms to avoid a stalled RLC window. The effect of polling for PDAN on uplink throughput is shown in table 5.8. The increased amount of downlink signalling blocks sent in uplink reduces the throughput for a class 8 MS compared to a single timeslot MS by 4%. The polling interval for a

FTP upload file size: 100kB		
MS multislot class	duration (s)	throughput (bit/s)
1	81.06	10104.5
2	81.44	10057.0
4	82.78	9896.0
8	84.20	9728.6

Table 5.8: Uplink throughput as function of MS multislot class

downlink TBF at a single PDCH is illustrated in table 5.9. Within one GPRS 52-multiframe one PACKET DOWNLINK ACK/NACK messages is sent. Each message acknowledges a maximum of 11 downlink RLC data blocks.

In case of the same uplink transfer but using a class 8 MS, 156 additional signalling blocks are sent in uplink, increasing the upload duration by 3.12 seconds.

The interaction of application behaviour and RLC/MAC layer signalling is in the case of single Tx MS classes less important. When transmitting TCP packets with a maximum segment size (MSS) of 1460 bytes, every packet arriving at the FTP server will trigger sending an acknowledgement, independent on the downlink multislot usage. Again has to be said that the interval of polling for downlink acknowledgements is dependent on the networks RLC/MAC layer implementation, but also on the existence of the delayed downlink TBF release, which increases

TDMA FN	Link	Message Type
1156375	UL	Packet Downlink ACK/NACK
1156376	DL	Downlink Data Block
1156379	UL	Uplink Data Block
1156380	DL	Downlink Data Block
1156383	UL	Uplink Data Block
1156384	DL	Downlink Data Block
1156388	UL	Uplink Data Block
1156389	DL	Downlink Data Block
1156392	UL	Uplink Data Block
1156393	DL	Downlink Data Block
1156396	UL	Uplink Data Block
1156397	DL	Packet DL Dummy Control Block
1156401	UL	Uplink Data Block
1156402	DL	Packet DL Dummy Control Block
1156405	UL	Uplink Data Block
1156406	DL	Packet DL Dummy Control Block
1156409	UL	Uplink Data Block
1156410	DL	Downlink Data Block
1156414	UL	Uplink Data Block
1156415	DL	Downlink Data Block
1156418	UL	Uplink Data Block
1156419	DL	Downlink Data Block
1156422	UL	Uplink Data Block
1156427	UL	Packet Downlink ACK/NACK

Table 5.9: Polling interval for single downlink PDCH

the number of RLC data blocks sent in downlink and subsequently increases the number of PDAN messages. In fact, reducing the downlink allocation during uplink data transfer does not affect the transmission of TCP ACK in the downlink, but it may reduce the RLC/MAC signalling overhead and potentially improve the uplink throughput.

Chapter 6

GPRS Radio Link Bandwidth

In this chapter, I discuss the maximum bandwidth per mobile client as function of coding scheme usage and block error rate. Once the radio resource assignment is completed (TBF set-up) and a given number of timeslots are assigned, the throughput is mainly determined by the coding scheme usage and RLC block error rate.
I briefly summarise the GPRS channel coding structure and the error protection mechanisms on the RLC layer, followed by some considerations on RLC block errors and block retransmissions, as well as RLC acknowledged and unacknowledged mode. The focus in this chapter is to provide throughput performance results and usage of CS-3 and CS-4 in a typical high traffic network.

6.1 The GPRS Channel Coding Schemes

GPRS provides four channel coding schemes. Two *strong channel coding schemes*, CS-1 and CS-2, with high coding redundancy offer high robustness against transmission errors but limited radio link bandwidth; two *weak channel coding schemes*, CS-3 and CS-4, offer less error correction capabilities but higher radio link bandwidth due to reduced coding overhead, [FRCODING00]. Maximising the radio link bandwidth for an error prone wireless link with varying quality makes a channel coding adaptation mechanism necessary.

6.1.1 Properties of GPRS Coding Schemes

The RLC/MAC frame structure for all four channel coding schemes is depicted in figure 6.1. The length of the RLC data block depends on the coding rate used. The MAC header length is one octet, containing the three USF bits, which are used to control uplink access in dynamic allocation mode. The remaining five bits are different for uplink and downlink RLC/MAC blocks. The RLC header is also different for uplink and downlink RLC blocks and it may change in size when more

than one LLC frame is spanned in the RLC data block, [3GPP-TS04.60].
The most channel coding protection and therefore the lowest radio link band-

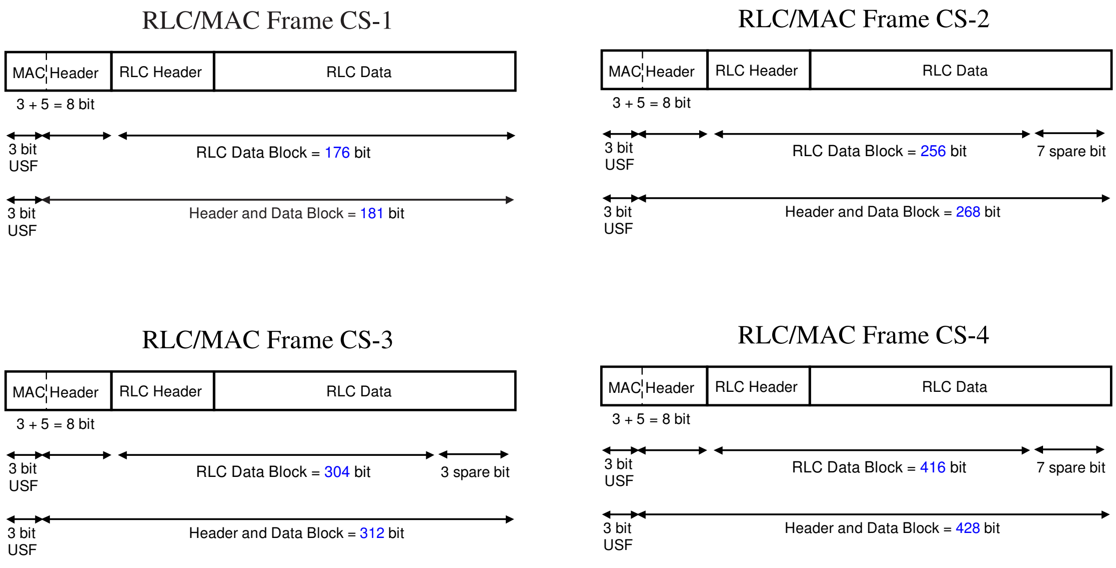


Figure 6.1: RLC/MAC header and payload size

width is offered by coding scheme CS-1, using the same coding as a SACCH frame, [3GPP-TS05.03]. After convolutional encoding, an additional forward error correction (FEC) using a 40 shortened binary cyclic code (FIRE code) is applied for maximum robustness against bursts of bit errors. In contrast, coding scheme CS-4 does not apply any channel coding at all, but therefore the maximum radio link bandwidth is offered. Error detection at CS-4 uses a 16-bit CRC, FEC is not applied. The parity bits of the 16-bit CRC or 40-bit BCS protect the RLC block information field as well as the USF bits. For CS-1,2 and CS-3 the USF bits pass the convolutional encoder; CS-4 is using a 12 bit pre-coding for error protection. For CS-2,3 and CS-4 the USF bits are encoded using the same bit pattern, therefore the MS can decode these 12 bits in the same way. Coding scheme CS-2 and CS-3 use punctured convolutional codes, while the USF bits are never punctured. The minimum Hamming distance between two USF codes is 5. The channel coding for all four coding schemes is illustrated in figure 6.2.
The receiving entity is informed about the coding scheme of each RLC/MAC block by decoding the 8 stealing bits within the four TDMA bursts. The stealing flags bit pattern (q(0) . . . q(7)) for encoding all four coding schemes is shown in table 6.1. The 8-bit encoding has a minimum Hamming distance of 5.

The different RLC payload sizes lead to different nominal data rates for every coding scheme, as to be seen in table 6.2.

	q(0)	q(1)	q(2)	q(3)	q(4)	q(5)	q(6)	q(7)
CS-1	1	1	1	1	1	1	1	1
CS-2	1	1	0	0	1	0	0	0
CS-3	0	0	1	0	0	0	0	1
CS-4	0	0	0	1	0	1	1	0

Table 6.1: Encoding of coding scheme information

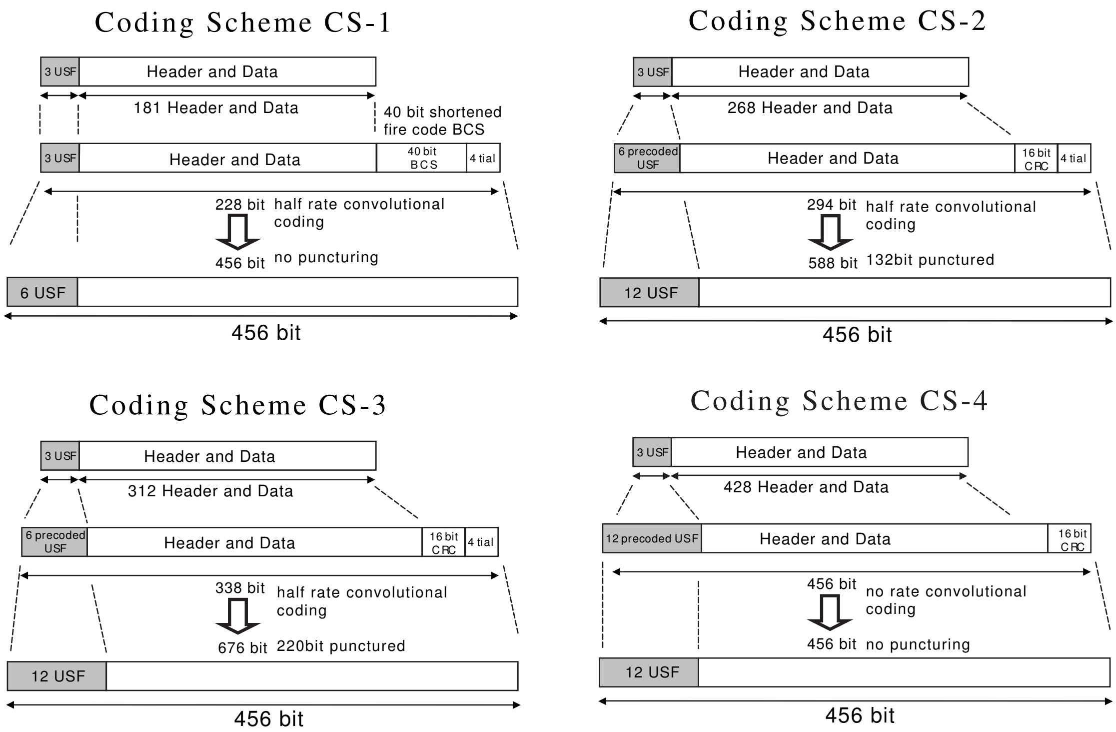


Figure 6.2: GPRS channel coding

Coding Scheme	Data Block Size (bit)	Spare Bits	RLC/MAC Block Size (bit)	Block Period (ms)	Nominal Gross Data Rate (kbit/s)
CS-1	176	0	181	20	9.05
CS-2	256	7	268	20	13.40
CS-3	304	3	312	20	15.60
CS-4	416	7	428	20	21.40

Table 6.2: RLC data block size

6.1.2 RLC Data Block Transmission Errors

An RLC data block is marked as received erroneously when the coding scheme information flags do not determine one of the four coding schemes, the USF bits

are not decodeable, or, when the CRC checksum fails. The coding schemes CS-2 to CS-4 use a 16-bit CRC for detection of transmission errors. The generator polynomial is: $g(D) = D^{16} + D^{12} + D^5 + 1$, as defined in the X25-standard. A description of error detection capability of a 16-bit CRC is given in [Tanenbaum88]. Tanenbaum shows this 16-bit CRC implementation detects all single and double bit errors, all errors with odd number of bits and burst errors of maximum length 16 bit. Bursts with 17 bit errors are detected with a probability of 99.997%, and 99.998% of 18 bit and longer error bursts are detected. The 40 bit CRC at CS-1 used for error detection and error correction implements the polynomial: $g(D) = (D^{23} + 1) * (D^{17} + D^3 + 1)$.
Once bit errors are not detected at RLC/MAC layer, LLC will be able to catch residual bit errors. Additionally at transport layer (TCP/UDP) a checksum is applied and finally it is possible to secure sensitive data against transmission errors on application layer. I assume the error probability of a GPRS link to be about 10^{-6} or even better[1]. During all of the GPRS performance tests I performed, not a single TCP segment checksum failure occurred because of residual bit errors passing the GPRS protocol stack.

6.1.3 RLC Acknowledged Mode

The transmission of RLC data blocks within a TBF can be done in *acknowledged* or *unacknowledged* mode. The RLC mode of operation as well as other RLC/MAC layer functions necessary for control of packet data physical channels are provided by the Radio Resource (RR) sublayer, (see [3GPP-TS04.60, clause 4.]). The RR sublayer provides services for the transfer of upper layer PDUs using a shared medium between multiple mobile stations and the network, RR protocol architecture and RLC/MAC function is depicted in [3GPP-TS04.60], figure 4.1.
The RLC mode is negotiated during the PDP context activation, see appendix E, but it can also be defined during TBF set-up. For uplink TBF set-up using the two-phase access procedure, the RLC mode information is included in the channel request description of either PACKET RESOURCE REQUEST, see table a-4 or PACKET DOWNLINK ACK/NACK, see table a-9. Downlink TBF set-up defines the RLC mode e.g. in PACKET TIMESLOT RECONFIGURE or PACKET DOWNLINK ASSIGNMENT. For message content description, refer to [3GPP-TS04.60], clause 11. The RLC mode cannot be changed during the TBF. The TBF establishment and subsequently the RLC mode is triggered by a request from upper layers to transfer a LLC-PDU. In a one-phase access, the RLC mode defaults to RLC acknowledged mode.
Acknowledged RLC mode is solely employed, since the most commonly used applications require reliable delivery of data. Unacknowledged RLC mode offers

[1] An error probability of 10^{-6} corresponds to reliability class 2 (see [3GPP-TS03.60], section 15.2. Quality of Service Profile)

less delay for time-critical streaming applications, such as voice over IP (VoIP), but it appears that generally the higher delay due to acknowledgement and retransmission of incorrect blocks is more acceptable than discarding N-PDUs due to RLC errors. Similar conclusions can be drawn from the simulation results shown in [TABO2002]. Acknowledged RLC mode uses a selective ARQ with a fixed window size of 64. Every RLC blocks is a number modulo 128 assigned for correct in-sequence reassembly of LLC frames. The acknowledgement interval is controlled by the network. Acknowledgement during downlink TBFs is requested by the network using a polling bit in the MAC header, which causes the MS to send a PACKET DOWNLINK ACK/NACK (see [3GPP-TS04.60], clause 10.2.1). Acknowledging during uplink data transfer is also controlled by the network by sending a PACKET UPLINK ACK/NACK to the MS. The ACK/NACK messages tell the sending entity which RLC blocks have been delivered successfully and which blocks have to be retransmitted. This, together with the channel quality reports included in the acknowledgements, is the only information the sending end receives about conditions on the radio link. Scheduling of retransmissions of negatively acknowledged blocks is incumbent on the sending entity. The coding scheme of retransmitted blocks has to be the same as during the first transmission. The performance implications of acknowledged RLC mode is also discussed in clause 4.2 and [LADPT00].

6.2 Link Level Performance, Laboratory Measurements

In this section, I discuss link level measurements in laboratory environment, under controlled propagation conditions, and at presence of co-channel interference. Additionally I compare the laboratory measurements with simulation results provided by ETSI, [ETSI-TR101.115].

6.2.1 Downlink Coding Scheme Usage, Static Case

The dependency of coding scheme usage at presence of co-channel interference for static MS is shown in figure 6.3. The provided throughput results are measured on application layer. The tests have been done with a four Rx MS (multislot class 8), the test application is a downlink UDP data stream to provide a continuous data flow across the radio interface. The test set-up provides a fading free environment and constant receive signal level. The radio environment is shielded against interference from life GSM networks. Transmission on four timeslots occasionally provokes RLC window stalls due to high BLER, as described in clause 4.2. Test samples with RLC window stalls have not been included in the test results; however, RLC stalls mainly happen at CS-4 and for C/I below 14dB.
For C/I below 10dB it is not possible to use CS-4, since it frequently happens that

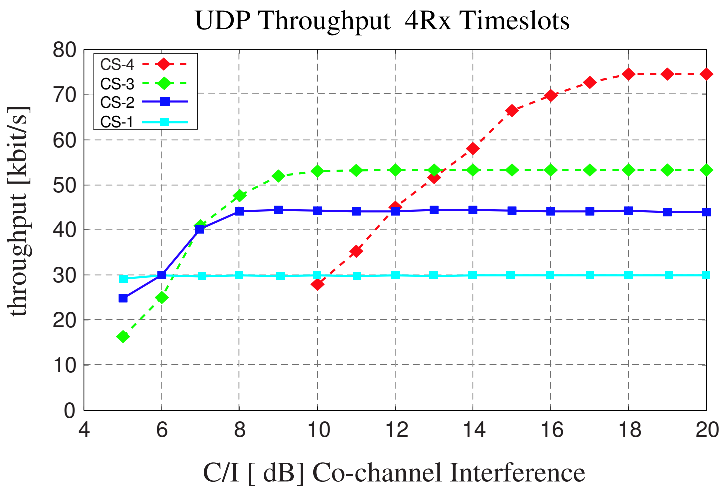


Figure 6.3: UDP throughput in fading-free environment

the TBF is released abnormally, in some cases even the PDP context is lost. CS-1, CS-2 as well as CS-3 can be used at $C/I = 5dB$, but abnormal TBF releases and loss of PDP context happen frequently. At C/I below 5dB, no radio connection could be maintained.

Conspicuous is the high throughput for CS-3 even for high interference. CS-3 can be utilised in a wide range, CS-2 just improves the throughput compared to CS-3 for C/I values below 7dB. At $C/I = 6dB$, the usage of CS-1 provides higher throughput compared to CS-2. CS-4 is superior for C/I above 13dB. Under very low interference conditions with C/I above 18dB, the throughput at CS-4 reaches the theoretical maximum and therefore BLER is zero. The throughput results illustrated in figure 6.3 are summarised in appendix F, table a-12.

6.2.2 Comparison of Simulation and Laboratory Results

In the following, I compare typical GPRS link level simulations as provided by ETSI[2], with the laboratory results shown previously. The aim of this comparison is to provide a prediction of coding scheme usage to be expected in urban environment and to summarise the most deciding effects on radio link quality from the MS point of view, i.e. velocity, environment and C/I.

The measurement results are scaled to provide throughput per timeslot figures, to allow a direct comparison with the simulation results. It has to be considered that the comparison of simulation results and field measurements just provides a qualitative estimation of link level performance due to varying simulation assump-

[2] See: [ETSI-TR101.115], Annex M: GPRS simulation results in TU3 and TU50, no Frequency Hopping

tions and test conditions. Apart from that, static results in a fading channel are not well defined because the performance is expected to depend randomly on the fading process at the point where MS is located. This might be the reason why static simulation results are not very common.

The performance for all four coding schemes for different MS velocities are depicted in figure 6.4. Two distinct effects are common for all four coding schemes.

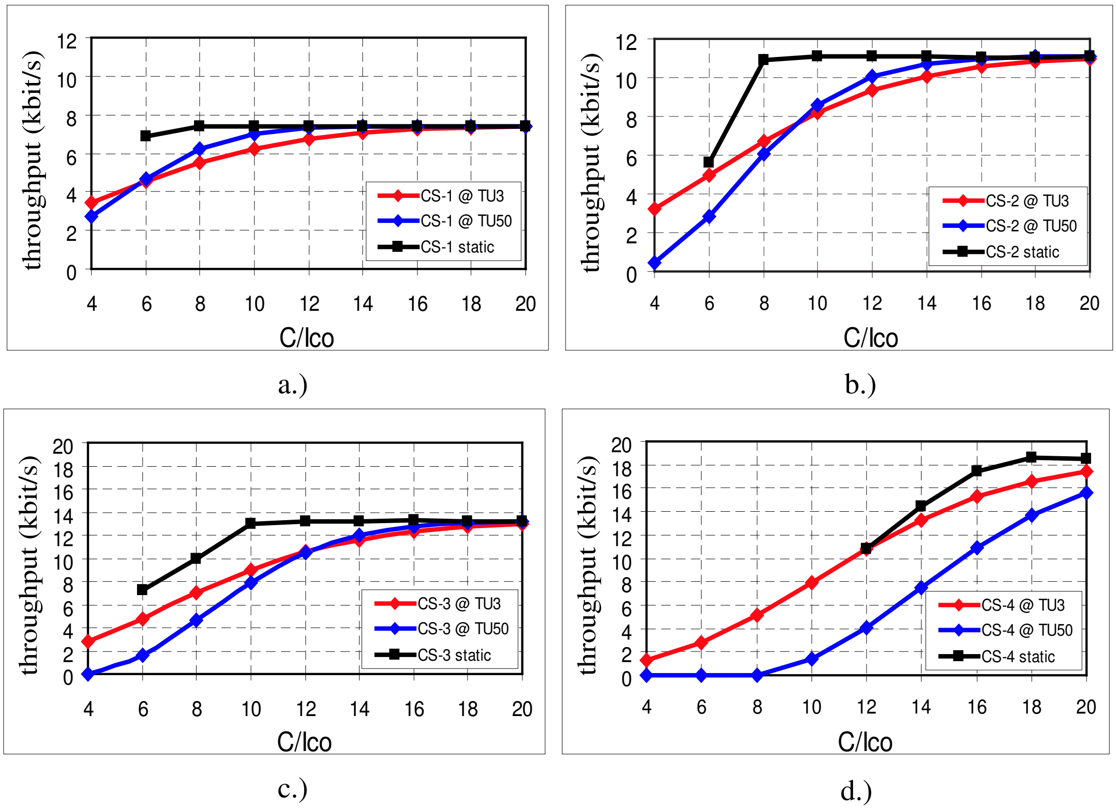


Figure 6.4: Coding scheme vs. C/I at different velocities

Firstly, the static throughput results outperform the results for moving MS and secondly, the fact that the maximum throughput degrades to approximately 40 to 50%. Beyond this value of throughput reduction due to BLER are abnormal TBF releases and even loss of PDP context happening. For CS-1 to CS-3 (the coding schemes that apply convolutional coding), the curves for 3km/h and 50km/h are crossing, while CS-4 shows constant performance degradation with increased velocity. An exact desription of this effect would require to analyse the equalizer and channel coding properties in depth, which is out of the scope of this diploma thesis. Therefore, I just give a brief explanation of the impact of mobility in the following. Mobility affects radio performance in two ways. The primary effect on

performance is signal variation from burst-to-burst. A secondary effect is signal variation within each burst. For moderate speeds (i.e. 50km/h), the variation of the received signal within each burst is negligible, but the signal variation from burst-to-burst results in diversity that can be taken advantage through coding and interleaving.
Now, if there is no channel coding (CS-4), burst-to-burst signal variation degrades performance. The simulation results presented in [TABO2002] also show that at higher velocity the signal variation within a burst will take effect, resulting in additional link performance degradation.

6.3 Link Level Measurements from Life Networks

The following measurement results focus on downlink throughput at CS-3 and CS-4. The aim of the measurements is to provide an estimation of how the weak coding schemes CS-3 and CS-4 perform in a typical GPRS network, especially in urban environment. Measuring exact figures for instantaneous C/I, BLER, coding scheme usage and application layer throughput at CS-4 turned out to be a challenge for typical link layer logging tools, which have been designed for CS-1 and CS-2 mainly. Therefore, I focus on end-to-end throughput and coding scheme usage without providing BLER or C/I measurements.
GPRS was implemented on the BCCH carrier; BCCH frequency reuse $5 * 3$. The application used to collect the throughput measurements is FTP, using an application server directly connected to the GSN, to avoid any influence from external PDNs.

6.3.1 Coding Scheme Usage in Rural Environment

The rural environment tests were conducted in a small village and the surrounding open areas. The cell radius is approximately 4.5km; the inter-site distance to neighbouring cells approximately ranges from 7 to 10km. The antenna is mounted 23 meter above ground and radiates $ERP = 48dBm$. The tests cover on-street static throughput measurements and motion tests with the GPRS-MS inside a car and at a vehicle velocity of 100km/h; the drive test route leads along a country road outside the built-up area.
For the static tests, three representative locations have been selected; one in near distance to the antenna (static 1), another one on the cell border, without major obstacles within the propagation area (static 2), and finally one in a shadowed area at the other end of the village (static 3). The drive test route starts next to the site and goes to the cell border, right at the beginning of another village. The locations for static downloads tests and the drive test route are illustrated in figure 6.5.

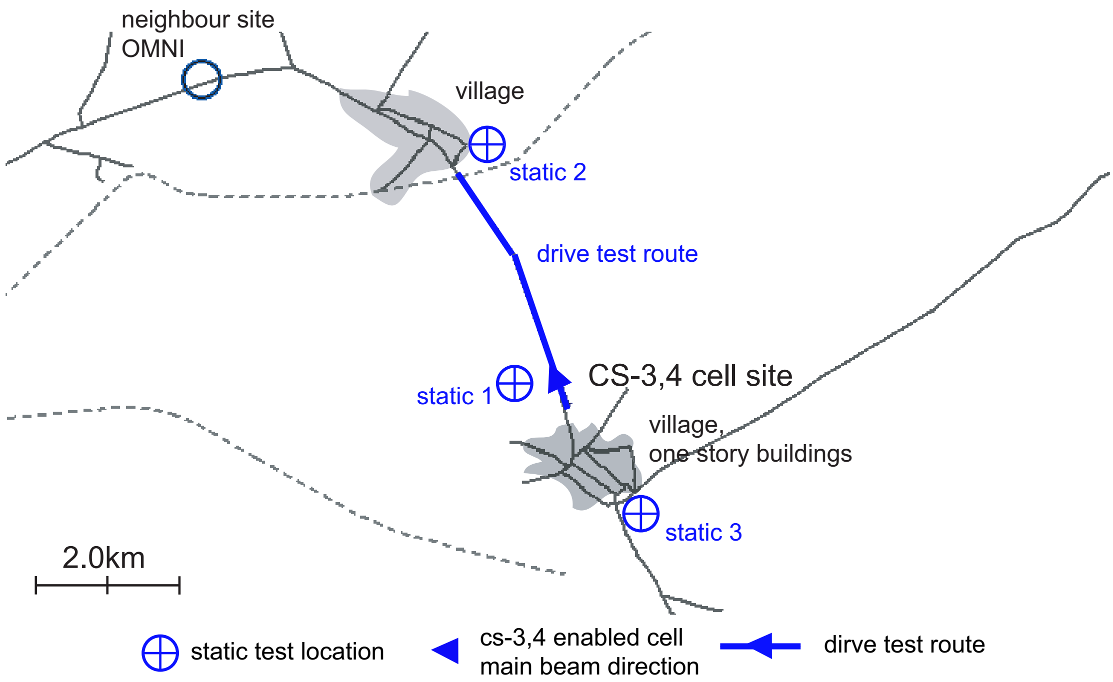


Figure 6.5: Test locations and drive test route, rural environment

6.3.1.1 Static Throughput Measurements, Rural Area

Average FTP throughput results for a class 1 and a class 8 MS are shown in table 6.3. The file size downloaded for the class 8 MS is 400kB, and for the class 1 MS 100kB, each average value shown in table 6.3 comprises 10 FTP file transfers. The test close to the antenna (static 1) obviously provides very good radio conditions; CS-4 can be utilised without being affected by block errors. The downloads 'static 2' still provide very high throughputs close to the theoretical maximum of CS-4, even though no line-of-sight to the antenna exists and the location is on the cell border. Only test 'static 3' in non-line-of-sight and at very low coverage level is significantly below the theoretical maximum of CS-4, but the throughput is still 20% above an error free transmission at CS-2. It can be seen that the results for the class 1 in comparison to the class 8 are consistent, even at a low sample rate by doing 10 FTPs.

6.3.1.2 Motion Tests Rural Environment

The drive test route as shown in figure 6.5 leads from the cell site to the cell border. The route leads through an open area and predominantly provides line-of-sight. An example of throughput results for 100km/h vehicle speed is illustrated in figure 6.6. The throughput graph shows a 2MB download using a class 8 MS. During the entire duration of the test, no other MS has an active TBF in the cell.

			average throughput	
			class 1	class 8
Location	Rxlev	Rxlev (dBm)	(kbit/s)	(kbit/s)
static 1	70	-40	16.85	72.6
static 2	48	-62	16.4	72.5
static 3	25	-85	11.6	54.34

Table 6.3: Static test results rural environment

After 17 seconds static download, the drive test starts. As soon as the vehicle

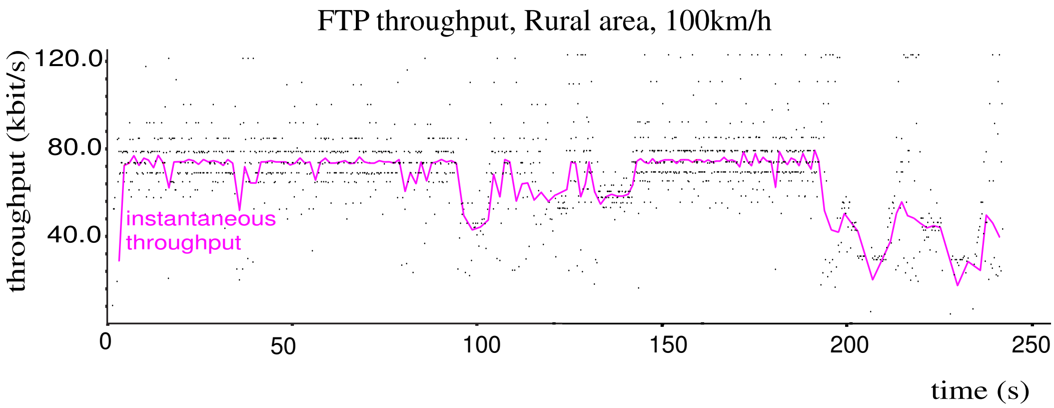


Figure 6.6: FTP throughput rural area, 100km/h

is setting into motion the radio channel characteristics change and the BLER increases. As a result of the sudden increase in BLER, the network changes the coding scheme from CS-4 to CS-3. During the short period of CS-3 a drop in throughput is seen. After switching back to CS-4 the maximum throughput is close to the theoretical maximum of CS-4 at four PDCH again and the average block error rate obviously is very small. Block errors happen though, as to be seen in the TCP throughput graph. The inter-arrival time of TCP/IP packets pretends to be more than 50% above the theoretical maximum on the radio link, indicating interleaved transmission of TCP segments due to RLC/MAC retransmissions. The throughput stays high until the car enters a shadowed area, 2.2km distance from the cell site. Here, the throughput goes down to approximately 48kbit/s. After 170s or approximately 4.2km distance a built-up area is reached and the speed is reduced to 40km/h. After passing the shadowed area the connection to the cell site is non-line-of-site. The throughput goes up to the theoretical maximum again, though. The connection is able to use CS-4 in a distance of further 200m from the serving cell (between 170s and 190s) before CS-2 and CS-1 can be used only. Finally, the MS performs a cell reselection to the stronger neighbour cell. The

total application layer throughput before reselecting is 60.8kbit/s.
Similar results are seen for several repeated tests, see table 6.4.

FTP	1	2	3	4	5
Throughput (kbit/s)	57.68	54.4	60.4	60.88	59.36
FTP	6	7	8	9	10
Throughput (kbit/s)	55.28	58.4	58.4	53.92	59.36
Minimum (kbit/s)	53.92				
Maximum (kbit/s)	60.88				
Standard deviation	2.4				

Table 6.4: Drive test results at 100km/h

6.3.1.3 Summary of Coding Scheme Tests in Rural Environment

Both, static tests and motion tests indicate very high utilisation of CS-4 in rural environment. For the case of a static MS, the maximum throughput at CS-4 can be achieved almost over the entire cell area.
Similar good results are seen for the motion case at 100km/h. In a rough estimation, it can be assumed that in rural areas, when moving at moderate speed up to 100km/h, CS-4 can be utilised within half of the cell radius. In the outer half of the cell area, the coding scheme drops from CS-4 to CS-2 or even CS-1.

6.3.2 Coding Scheme Usage in Urban Environment

The tests in urban environment have been conducted in the coverage area of one macro cell. The cell radius varies from 250m to 500m.

6.3.2.1 Static Throughput Measurements

Similar to the rural performance results a high CS-4 utilisation for static downloads is even in urban area possible. Figure 6.7 illustrates downlink FTP results using a single timeslot MS. The test results have been conducted in a typical, densely build-up European metropolitan city centre, in a high capacity network. CS-4 is utilised in the entire best server area[3] at very low block error rates. With increasing distance, the maximum throughput degrades depending on the coverage level. A noticeable effect is that the MS is still camping on the tested cell even when neighbour cells are in closer distance to the MS compared to the serving cell. This behaviour is caused by the cell reselection behaviour. A GPRS capable MS applies

[3] Best serving area of a cell site is the area where the cell provides highest signal level compared to the surrounding neighbouring cells

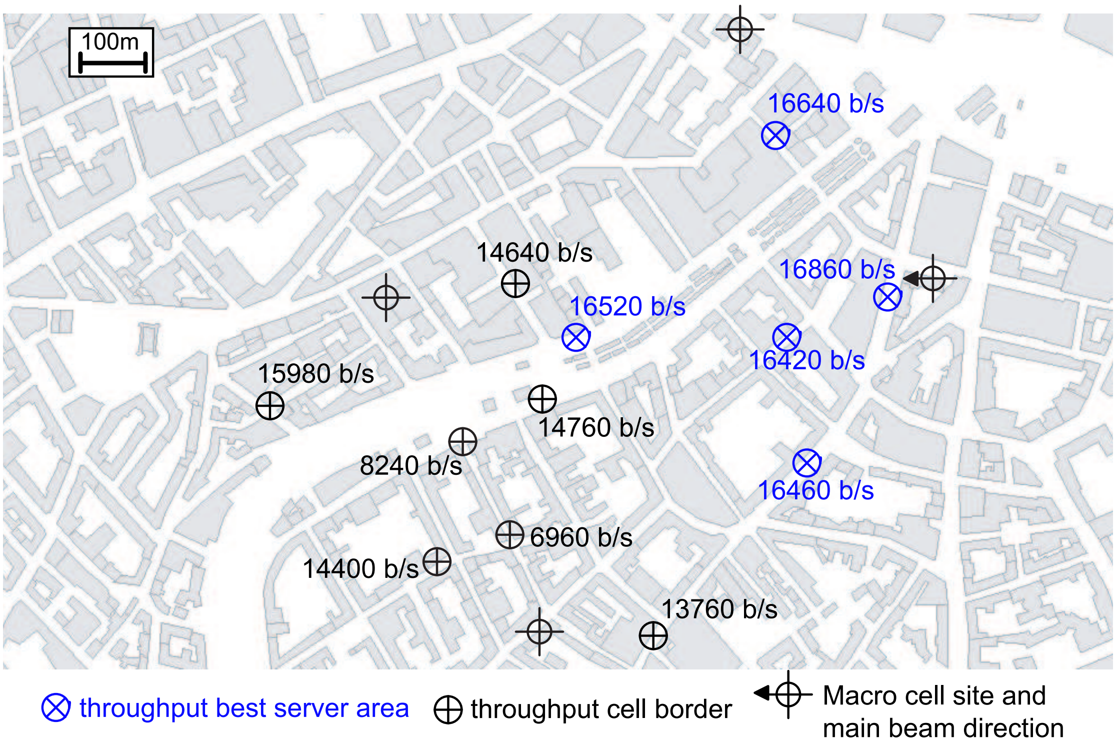


Figure 6.7: Static throughput results, class 1 MS in urban environment

the parameter CELL_RESELECT_HYSTERESIS which defines the minimum difference in signal level before the MS reselects a neighbouring cell, [3GPP-TS05.08]. High values for the reselect hysteresis will force the MS to stay on the serving cell even when stronger neighbours could be selected, subsequently the BLER will raise and the coding scheme usage will be reduced to stronger coding schemes and lower radio link bandwidth when the MS is not camping on the strongest cell.

6.3.2.2 Motion Tests Urban Environment

The motion tests provide a comparison between two drive test routes through the coverage area, see figure 6.8. The tests have been conducted using a class 1 MS and during times with low traffic, to be able to keep the vehicle speed of 50km/h. It can be seen that CS-3 and CS-4 are less effective in the motion case. The throughput along route-1 shows a very high usage of CS-3 and CS-4, indicating that the TU50 conditions do not apply in this environment. Along route-2 CS-3,4 improves the throughput by approximately 45% compared to the theoretical data rate at CS-2. The difference between the two routes are the environmental conditions, while route-1 leads through an area with lower building height, broad streets and open room, route-2 leads through densely build up environment with

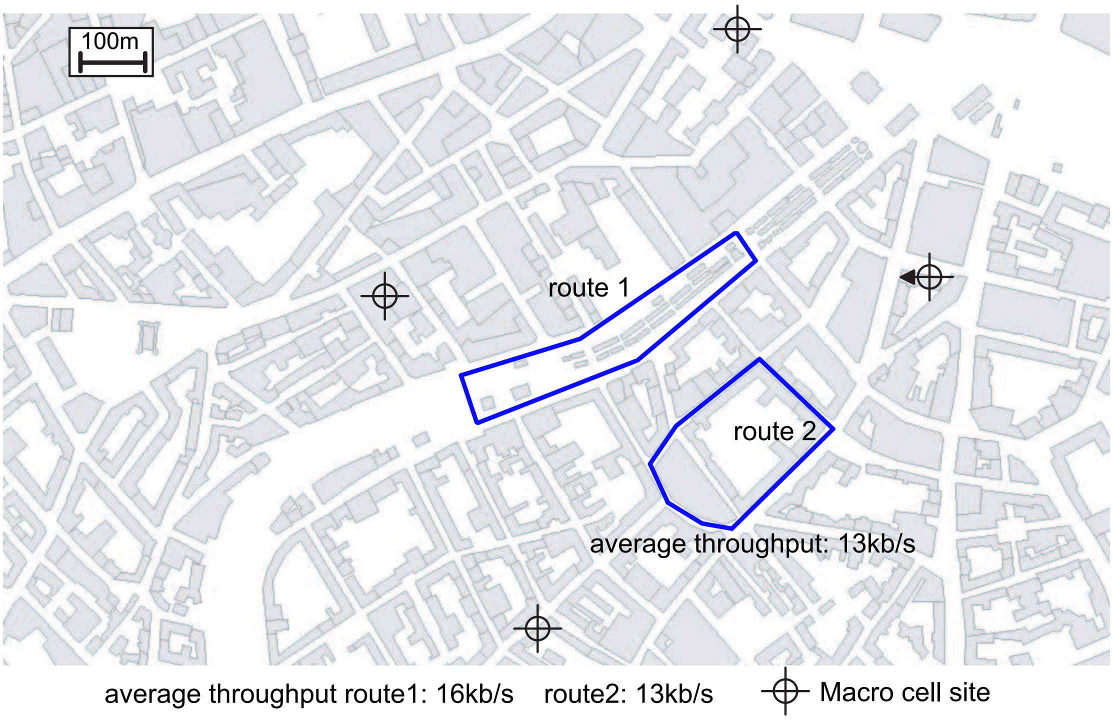


Figure 6.8: Drive test results, class 1 MS in urban environment

small streets. The average Rxlev for route-1 is $-62dBm$ and for route-2 $-66dBm$ accordingly.
The distribution of coding scheme usage is illustrated in figure 6.9. The drive

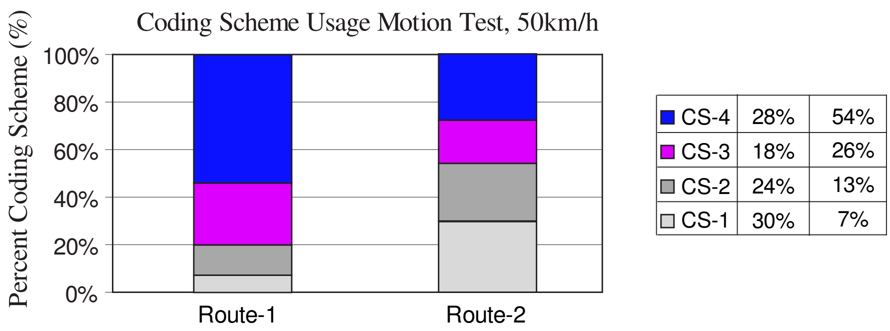


Figure 6.9: Coding scheme distribution, urban environment, 50km/h

test results show that the environmental conditions predominantly determine the radio link performance. The assumption that in urban environment the channel model 'TU' applies is not obvious. Especially street canyons with good coverage or areas with partial line-of-sight and a predominant propagation path will offer

better performance than simulations for TU environment predict.
The throughput improvement of CS-3,4 varies between zero for fast moving MS to 60% for static MSs in good radio conditions.

6.4 Conclusions - Performance of Higher Coding Schemes in Life Networks

Throughput for static GPRS users will greatly improve at higher coding schemes CS-3,4. CS-3,4 provides up to 60% increase in throughput compared to CS-1,2 in almost the entire cell coverage area. This applies to urban as well as rural environment. Only in the outer zone of the cell coverage area degradation in throughput is seen; although CS-3,4 still significantly increases the user throughput compared to a network that supports CS-1 and CS-2 only.
Motion tests in rural environment also show high performance gain of CS-3 and CS-4, at least for MS speeds up to 100km/h.
Motion tests in urban environment show high differences in radio link performance. In areas close to the cell site, CS-3,4 still improves the data throughput compared to CS-2. In far distance from the serving cell, the performance gain from CS-3,4 is limited. The reduced signal strength and environmental conditions highly affect the wave propagation and sometimes not even continuous CS-2 coverage is supported. Especially in the transition zone from one cell's best coverage area to the neighbouring cell's best coverage area, moving GPRS users do not benefit from higher coding schemes CS-3,4.
However, high-traffic networks in urban environment and with small cell-radius often provide good propagation conditions even at the cell border. In street canyons with good illumination with good coverage level, CS-3,4 occasionally improves the throughput even in far distance from the serving cell. Therefore, the field performance results frequently outperform results predicted by simulations based on TU50 environmental conditions.
In dense built-up areas in the outer zone of the cell, at low coverage level, non line-of-sight to the serving cell and distinct multi-path propagation, GPRS performance results are similar to the simulation predictions for TU50 environment. Under these circumstances, CS-3,4 does not improve the throughput compared to CS-1 and CS-2 in urban environment.

Chapter 7

GPRS Link Layer Performance Analysis

The GPRS link layer is split into two protocol layers, the LLC and SNDCP layer. In addition, I include BSSGP in the link layer analysis. Actually, BSSGP does not provide typical link layer protocols functions, it is rather a GPRS proprietary protocol that is terminated in the PCU and SGSN. I include BSSGP in the link layer analysis as it encapsulates and routes LLC frames over the Gb-interface. The GPRS link layer protocols provide functions according to layer 2 of the OSI reference model, i.e. securing underlying protocols against transmission errors and adding reliability. In particular are LLC frames secured by a strong CRC checksum, to detect transmission errors, which have not been captured by the RLC error protection. In-sequence delivery and acknowledged mode by means of a ARQ protocol are securing against LLC frame loss. SNDCP provides compression to increase the efficiency of the bandwidth limited radio link.
In this chapter, I discuss LLC and SNDCP layer properties and I show measurement results of SNDCP compression. First, I summarise the various effects leading to LLC frame loss. I use these results to quantify the benefit of LLC acknowledged mode. Finally, I discuss particular functions of the BSSGP protocol, the flow control in the downlink direction.

7.1 Link Layer Characteristics

The LLC layer provides a highly reliable connection between SGSN and MS. It spans the SGSN, Gb-interface, Um-interface to the MS and works independent of underlying radio interface protocols, e.g. implementing EGPRS will introduce basic changes in the physical layer and RLC layer without affecting LLC functionality. Further functions are provision of one or more logical link connections discriminated between by means of the Service Access Point Identifier (SAPI), error detection and recovery, in sequence delivery of LLC frames, downlink flow control and ciphering, i.e. provide user data and user identity confidentiality. LLC also provides Quality of Service parameters to be used in each PDP Context. The LLC layer protocol support two modes of operation:

- **Unacknowledged operation (Asynchronous Disconnected Mode):** A logical link entity may initiate transmissions without prior establishment of a logical connection. Errors are detected but no error recovery procedures are defined at the LLC layer. Flawed or duplicate frames are rejected, no retransmission at LLC layer is made and higher layer protocols need to provide reliability, depending on the application used. Two transmission modes are distinguished: protected, FCS calculated over entire LLC frame and unprotected, FCS calculated over LLC frame header only.

- **Acknowledged operation (Asynchronous Balanced Mode):** A balanced data link involves two participating entities, and each entity assumes responsibility for the organisation of its data flow and for error recovery procedures associated with the transmission. It provides a reliable service with in-order delivery. A subset of LLC parameters, important for acknowledged mode are:

Parameter	Description
T200	Retransmission timer
N200	Maximum number of retransmissions
kD	Downlink window size
kU	Uplink windows size

Table 7.1: LLC parameters

Acknowledged mode requires high amount of CPU and memory resources in the SGSN and the MS. The answer which mode is to prefer depends on many parameters, such as application protocol behaviours, round trip delay between MS and SGSN, LLC protocol parameters. LLC ACK mode only improves the GPRS performance, when LLC-PDUs are discarded between MS and SGSN.
The most significant effect of losing LLC-PDUs happens due to user mobility, when reselecting into a different serving cell. Other effects leading to loss of LLC frame loss are residual bit errors after RLC error protection or congestion on the frame relay network on the Gb-interface. However, my empirical results show that discarding LLC frames due to CRC failures must be a rare event, independent on the coding scheme used and on the latter is LLC ACK mode not providing an end-to-end performance improvement. Hence, I consider the frequency and performance impact on LLC frame loss due to cell reselection being the deciding factor to conclude about the benefit of LLC acknowledged mode. The effects of user mobility and cell reselections are discussed in the following section.

7.2 Effects of User Mobility on GPRS Link Layer

User mobility and cell changes during an active data transfer can lead to loss of downlink LLC-PDUs. The SGSN always provides sufficient amount of LLC frames to guarantee continuous data flow over the radio link within one TBF, hence the PCU is required to buffer LLC frames (leaky bucket algorithm, see chapter 7.5). Cell reselections between two BSS will always lead to loss of buffered LLC frames since the Gb-interface protocols do not provide forwarding of LLC-PDUs between different BSS. On reselections between cells belonging to the same BSS, it is possible to forward buffered LLC frames from the original source cell to the new target cell. In that case, LLC ACK mode does not improve the GPRS performance. Loss of LLC-PDUs in the uplink direction does not happen, since the MS is able to release the radio contact on transmission of LLC boundaries.

7.2.1 Cell Reselection with LLC-PDU Loss

The effect of LLC-PDU loss on reselection is shown in figure 7.1. It can be seen that the TCP session resumes approximately 6 seconds after the MS released the radio contact in the source cell, when the retransmission time-out in the TCP server sends the discarded TCP segments again. Actually, the MS establishes a radio

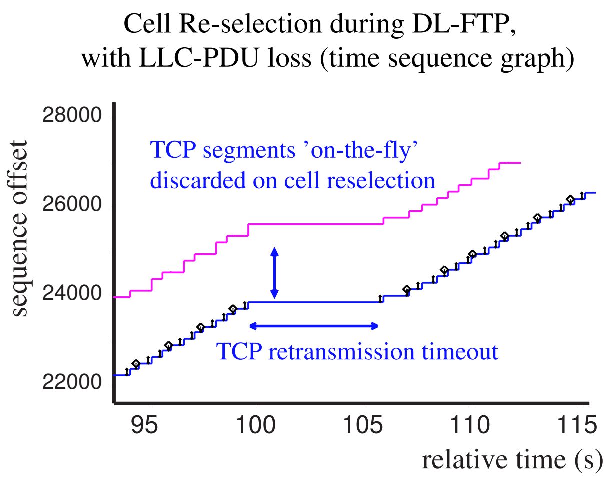


Figure 7.1: TCP segment loss - cell reselection during downlink FTP

link typically two seconds after leaving the source cell. The MS has to decode the system information messages in the new target cell and finally it is able to establish an uplink TBF again. In recent GPRS systems, the forwarding of LLC-PDUs on cell reselect is supported and the performance impact on application layer is reduced. The throughput outage on TCP layer will take about two seconds and no TCP congestion avoidance mechanism is invoked. This implementation of seamless cell reselection reduces the benefit of LLC acknowledged mode, as predominantly reselections between different BSS would cause loss of LLC-PDUs. Considering BSS size of several hundred cells can be assumed that reselection to a different BSS will only happen in about 10% of all cell reselections in the network.

7.3 LLC Acknowledged vs. Unacknowledged Operation

Cell reselection during an active data session always reduces the data throughput since the radio contact between MS and network is lost for about two seconds. Furthermore, loss of LLC frames is possible. This inevitably causes a loss of upper layer PDUs. Especially TCP based applications suffer from lost data since TCP interprets a segment loss as network congestion and therefore reduces the offered throughput. LLC acknowledged mode could improve the upper layer protocol performance by avoiding retransmissions on time-out or congestion control. However, interaction with upper layer protocols retransmission mechanisms have to considered when evaluating LLC acknowledged mode. The different interdependencies of link layer and upper layer retransmissions make an exact evaluation of acknowledged mode rather complicated. Hence, various studies come to different results, e.g. [GPRS2000] concludes that time-out based retransmissions of TCP and LLC retransmissions can worsen the application performance, since LLC retransmissions cause TCP segment retransmissions due to time-out. The LLC parameter settings (see table 7.1) and the TCP implementation used for the simulations are not explicitly stated in this paper. A more detailed description of LLC acknowledged mode and TCP interaction is given in [TCPLLC]. The simulation results show optimum settings for LLC parameters based on the LLC-PDU size. A proper setting of the LLC retransmission timer, T200, is able to avoid spurious TCP retransmissions. An optimised setting for N200, which defines the maximum number of retransmissions, is also essential to avoid a LLC connection release, followed by LLC discard and subsequent LLC connection re-establishment. In opposite to [GPRS2000] the simulations show possible improvements of TCP based applications and LLC acknowledged mode. However, the optimised parameters depend on the LLC-PDU size.

The fact that recent BSS implementations provide forwarding of LLC frames within a PCU reduces the necessity of LLC acknowledged mode. Considering the required resources in the SGSN and the MS can be said that the cost of implementing LLC acknowledged mode in the SGSN is not justified by the limited performance gain seen by the GPRS user.

Similar conclusions are stated by [RFC3366]. LLC-ARQ is only able to improve the TCP performance when optimised persistence of LLC retransmissions is chosen. Unwanted interactions of multiple connections across shared high delay links are mentioned. Additionally, the impact of LLC retransmissions on other UDP or TCP flows may lead to unwanted jitter and spurious retransmissions. From the GPRS perspective, I conclude that reliability should be provided by the RLC protocol, by implementing RLC acknowledged mode. The event of LLC frame loss on cell reselection should be avoided by forwarding LLC-PDUs from the source cell to the new serving target cell. Finally, retransmissions of unavoidably lost LLC frame on reselection, e.g. reselection between two BSS, have to be handled by higher layer protocols, such as transport layer or application layer. Up-to-date TCP implementations making use of the fast recovery mechanism (see 8.1.1.3) have to ensure minimum effect of LLC frame loss on upper layer protocol performance.

7.4 SNDCP Layer Performance Analysis

In this section, I briefly discuss header and data compression provided by SNDCP. Beside compression, SNDCP provides multiplexing of N-PDUs, LLC link establishment for acknowledged mode and XID negotiation. Compression saves transmission time over the radio link on the expense of MS and SGSN computing time. Both compression entities, SGSN and MS are challenged when enabling compression, resulting in additional delays or even TCP segment loss. Most of the mobile stations do not support compression, even when requesting compression on PDP context activation via modem commands (see [3GPP-TS07.07]):

```
at+cgdcont=1,"IP","APN","0.0.0.0",1,1
```

Where the final '1,1' enables data and header compression. But not only the MS, also most of the GPRS core networks tested do not enable compression, or only a limited number of connections can be using compression due to computing time limits.

7.4.1 Link Layer Compression - Overview

The initial release 1997 GPRS specifications[1] define TCP/IP (IPv4) header compression as specified in [RFC1144]. In addition, SNDCP supports V.42bis data compression, as defined in [ITU-T-V42BIS]. The way negotiation of compression is defined is open to add new compression methods in later releases. In the release 1999 set of specifications [3GPP-TS04.65] additional TCP/IP and UDP/IP header compression for IPv4 and IPv6 is defined according to [RFC2507].

7.4.2 Known Performance Issues on Error Prone Links

TCP/IP header compression according to RFC1144 uses a delta compression method, saving static header information, which does not change from one packet

[1] See 3GPP-TS04.65, version 6.7.0

to the other. TCP options active when TCP segments are lost, such as selective acknowledgement, explicit congestion notification, or TCP time-stamp are not compressible. In addition, segment losses cause synchronisation loss between the two compression entities, reducing TCP retransmission mechanisms. These issues are addressed in the release 1999 specifications of SNDCP, which includes an explicit request for retransmission of an uncompressed packet to allow decompressor resynchronisation [RFC2507].

7.4.3 Link Layer Compression - Tests

I performed a limited number of functional tests of SNDCP compression using one type of GPPRS handset only. In fading free environment and with CS-1/CS-2 only, a maximum throughput improvement of 140% has been achieved with data and header compression, when downloading a plain text file. When compressing the text file before download, data compression on SNDCP layer is not able to improve the data throughput. A maximum throughput improvement of 2.5% has been seen for maximum MSS of 1460 byte. This result is in accordance of the expected throughput improvement of header compression.
A detailed analysis of performance gain due to SNDCP compression could not be completed, as frequent MS resets were result in lost PDP context and connection drop.

7.4.4 Conclusion on Link Layer Compression

Compression only improves the data throughput when data with redundancy is transmitted. Typical HTTP pages embed images in compressed formats, or include compression on application layer, such as HTTP/1.1 does, [RFC2616]. Improving end-to-end security by transmitting data over an encrypted tunnel through the GPRS network, e.g. when using IP security [RFC2411], is also not offering room for data compression. Thus, GPRS link layer compression offers limited performance gain.
Further considerations on compression are discussed in [PILC-LINK]. The conclusion drawn is that the benefit of application data compression is limited when encrypted or already compressed data is transmitted. It is also concluded that end-to-end compression is beneficial for every network element in the path, while subnetwork layer compression only improves for a single subnetwork.
Applying all these results on GPRS, I conclude that TCP/IP or UDP/IP header compression can be of benefit for GPRS, when a release 1999 implementation of header compression can be utilised, especially for networks with limited GPRS radio link bandwidth offered. The actual implementation of header compression depends on the cost of implementation in the GSN and the availability of the MS supporting compression. Application data compression should be done on application layer, e.g. by a performance enhancing proxy, to reduce the transmitted

data volume over the entire GPRS network.

7.5 Gb-Interface Performance Analysis

I include the BSSGP protocol in the GPRS link layer analysis, since LLC-PDUs are encapsulated in BSSGP when transmitted over the Gb-interface. BSSGP is a GPRS specific protocol, which is terminated in the SGSN and the PCU, thus it is not visible for the end user. The performance impact of BSSGP comes from the flow control between PCU and SGSN in downlink direction. BSSGP does not have any performance impact in uplink direction. In the following, I discuss the performance impact of the flow control in downlink direction and I summarise potential shortcomings of the flow control algorithm.

7.5.1 Gb-Interface Flow Control

The flow control is required since application layer PDUs arrive at the SGSN on a high bandwidth backbone network with minimum delay. On the Gb-interface, the available bandwidth will be reduced to several hundred kbit/s and finally the radio interface will further limit the available bandwidth. In addition, the bandwidth on the radio interface strongly varies based on RF conditions and network traffic load. The aim of the flow control is firstly to avoid overloading the PCUs input buffers and secondly, fair distribution of resources among GPRS users. The interface requirements of the flow control and a possible implementation are defined in [3GPP-TS08.18].
The flow control is done according to the *leaky bucket algorithm*. The PCU informs the SGSN about the current data throughput across the radio interface and the maximum available buffer space per cell and per MS in the cell. The SGSN is responsible to use these parameters to estimate the buffer level in the PCU, and decide if the PCU is able to receive further LLC-PDUs or if it should be delayed until the PCU buffer level gets lower.
The algorithm proposed in the TS uses a two-stage approach to manage the data flow, as shown in figure 7.2 and [3GPP-TS08.18]. The BVC flow control avoids flooding the PCU buffers while MS flow control ensures fair sharing of PCU buffer resources. Figure 7.2 b.) illustrates the sharing of the entire BVC bucket among the active users in the cell. Actually, the MS flow control avoids situations where a streaming application occupies the entire PCU buffer while other MSs in the cell suffer high queuing delays. This is also important to support LLC scheduling disciplines such as FIFO, or *Earliest Deadline First* and finally support QoS for GPRS.

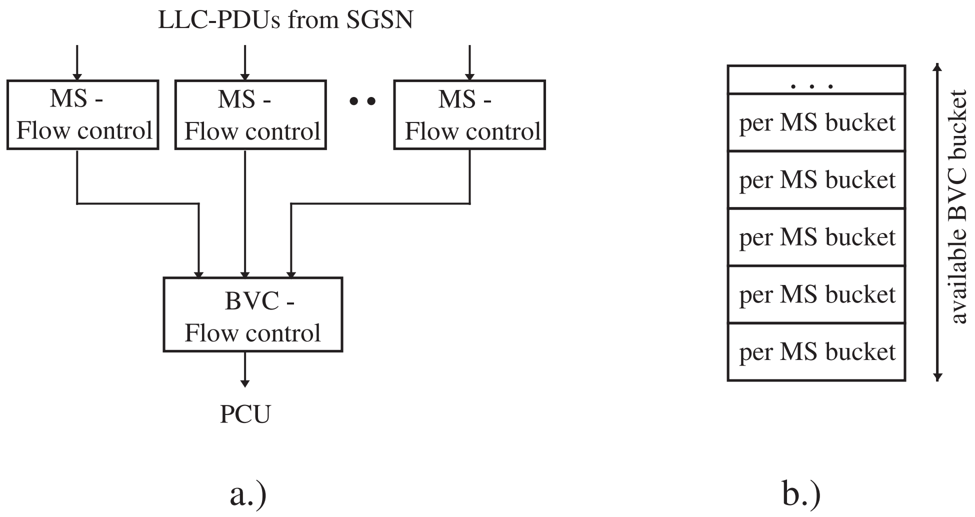


Figure 7.2: Two-step approach of BSSGP flow control

7.5.2 Flow Control Measurement Reports

For BVC flow control according to release 1999 of the TS, the following information is sent from the PCU to the SGSN:

- The maximum bucket size (Bmax)
- The leak rate parameter (R) according to the data throughput on the radio link
- The current bucket full ratio, if the feature Current Bucket Level (CBL) is negotiated
- The default MS bucket size (Bmax_default_MS)
- The default MS leak rate (R_default_MS)

The CBL feature is an enhancement of the release 1997 specification to provide re-synchronisation of the SGSN calculated bucket fill ratio and the actual bucket level at the PCU. This feature is also a mandatory requirement on cell reselect, to synchronise the SGSN bucket level counters after forwarding of LLC frames from the source cell to the new target cell. I am discussing the CBL feature in the following section, 7.5.3. Additionally MS flow control related information is sent. The default values for MS flow control are applied by the SGSN as long as no MS flow control message is received by the SGSN. Finally, the PCU can include an optional delay value measurement, reporting the queuing delay of LLC-PDUs.
The MS flow control message has similar structure as the BVC flow control message, containing:

- The TLLI identifying the MS
- The maximum bucket size (Bmax) for this MS
- The leak rate parameter (R) to be applied to the bucket
- The bucket full ratio, if the Current Bucket Level (CBL) feature is negotiated

The SGSN is required to acknowledge flow control messages within a 100ms time constraint.

7.5.3 Potential Shortcomings of BSSGP Flow Control

The initial implementation of BSSGP flow control of release 1997 of the TS was susceptible to drift of the PCU bucket level calculated by the SGSN. The main effect causing a drift in the calculation of bucket size is the limited dynamic in the feedback process between PCU and SGSN. The PCU is limited in sending flow control messages; at most, every second a flow control message can be sent in order to reduce the amount of control messages on the Gb-interface. In addition, it is required to define a minimum threshold of leak rate has to be exceeded before the PCU is allowed to send a new flow control message. Small changes in throughput on the radio interface will not cause a flow control message to be sent by the PCU. These effects will inevitably leading to a drift in the SGSNs bucket level estimation. Re-synchronisation was only given when the cell was not carrying ans GPRS traffic for a period of time, until the SGSN could be sure that the bucket level is zero.

The result of a drifting bucket level estimation towards higher bucket levels result in reduced throughput, since the SGSN throttles the data rate to reduce the bucket level. The SGSNs reduced sending rate can cause LLC frame loss in the SGSN, as internal buffers can be flooded. This scenario could happen when one user starts a long data download, or when several users cause a continuous downlink data flow in the cell. Figure 7.3 shows negative effect of a drifting BSSGP flow control procedure. The tests shows a FTP download of a 200kB file after the the cell was continuously active for approximately forty minutes. During that period, the SGSNs estimation of the PCU bucket level drifted to higher values of bucket fill level. After 14 seconds, the SGSN starts reducing the data rate in order to relief the PCU bucket. As a result of the reduced throughput, the PCU reports lower leak rate, what in turn causes the SGSN to throttle the data rate again to avoid buffer overflow. Finally, the SGSN discards PDUs due to buffer overflow within the SGSN. This deadlock situation is relieved after five to 10 seconds. The transmission resumes after TCP retransmission time-out, when the TCP flow control recovers from TCP segment loss.

These negative effects have been addressed by different equipment vendors in modified flow control implementations. Some BSS vendors reported very high leak

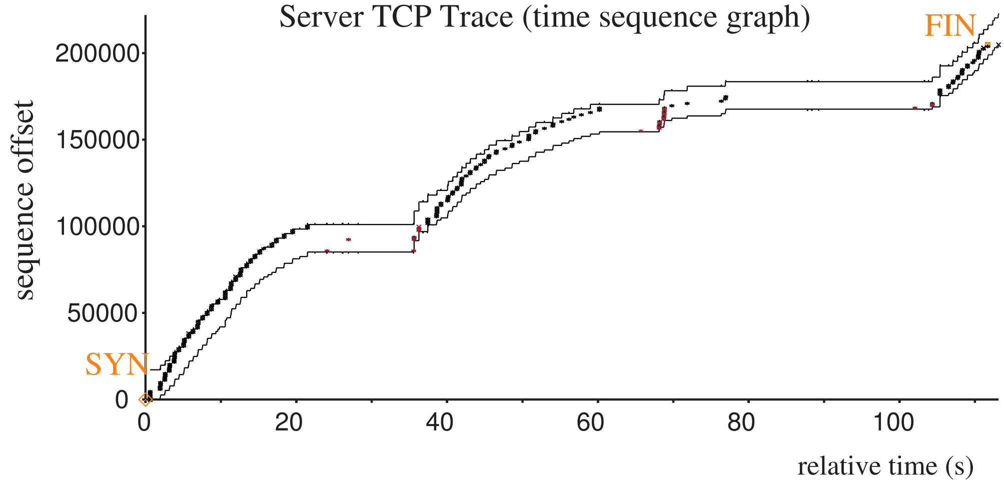


Figure 7.3: Gb-interface flow control affecting downlink throughput

rates, slightly above the maximum data rate depending on the number of timeslots available in the cell. This implementation of BSSGP flow control works similar to the Xon/Xoff protocol but effectively avoids throughput reduction due to wrong bucket estimations.

In the proceeding development of GPRS the CBL (Current Bucket Level) feature was introduced in the release 1999 specifications. The CBL feature provides an additional parameter in the flow control messages, which informs the SGSN about the actual bucket fill level. Not only the drift in bucket level calculation, but also the support of seamless cell reselection, providing forwarding of LLC-PDUs on cell reselection, made CBL feature for BSSGP flow control mandatory. Resynchronisation of bucket calculation is done with every flow control message.

Another problem of BSSGP flow control and buffering in the PCU can arise when the client advertised TCP window is very high, e.g. 64kB and above. Considering a full TCP window of 64kB queued in the PCU and four radio timeslots at coding scheme CS-2 will result in 11.6 seconds queuing delay in the PCU. This can cause spurious retransmissions of TCP segments or PDU discard on lifetime expiry. This may be seen when cell reselections in packet transfer mode between 3G and 2G RAN are happen. A TCP connection in a 3G network may require TCP window size of 64kB, or even TCP window scaling, increasing the TCP window size above 64kB.

Chapter 8

GPRS System Level Performance

In this chapter, I discuss the performance of different applications and their associated transport protocols when used on a GPRS network link. The focus in this chapter lies on applications above TCP/IP. Particularly, I analyse FTP and HTTP, since these applications are frequently used and they require high attention to achieve optimised application performance over GPRS. Applications like MMS or WAP are built above UDP rather than TCP and they provide optimised protocol suites for wireless access, making optimisation less crucial compared to TCP based applications.
In the following, I show the most important TCP parameters to be considered when used over GPRS, followed by a short review of applications based on the TCP/IP protocol suite. Finally, I discuss GPRS network enhancements to optimise the application performance over GPRS links.
The majority of applications run above TCP/IP. The slow migration to IPv6 also shows the lasting nature of a wide spread technique like TCP/IPv4. Moreover, the efforts to establish IPv6 across the Internet give additional reason to focus on TCP on top of IPv6. The common understanding today is that end-to-end usage of ATM has major advantages compared to TCP/IP. However, due to the persistence of TCP can be forecast that TCP/IP will also play an important role in next generation networks.

8.1 TCP Performance Analysis

TCP is an acknowledgement driven protocol, which provides a reliable connection oriented, full-duplex connection. TCP supports flow control, which prevents the sending entity from overrunning the buffer capacity in the receiving entity. In addition, TCPs slow start and congestion control avoids overrunning the buffer capacity of network elements within the end-to-end connection path. Detailed descriptions of TCP are [Tanenbaum88] or the TCP standard [RFC793]. RFC793 has been updated by [RFC3168], defining the *Explicit Congestion Notification* (ECN). ECN enhances TCP flows by enabling the sender to reduce the transmission rate when an intermediate network entity in the transmission path sets the congestion notification. Hence, TCP segment loss due to congestion is avoided. The support of ECN is negotiated during set-up of the TCP connection. Unfortunately, ECN is not commonly used in the Internet.

TCP was originally designed to work above wired links, which usually provide a very low error probability. The congestion avoidance mechanism interprets a packet loss as congestion in an intermediate network element or link, causing the serving entity to throttle the transmission rate. Together with the high latency of wireless links such as satellite links or GPRS, TCP congestion avoidance is heavily affecting the data throughput when packet losses due to transmission errors in the wireless link happen. The TCP insufficiencies were revealed in the mid 80's, when introducing of data transmission on high bandwidth, high latency satellite links, having a round trip delay of up to 1.5 seconds. GPRS turned out to be even more challenging since the latency typically ranges from 1 to 2 seconds and the bandwidth varies between 10kbit/s and 70kbit/s. In addition are access delay and bandwidth varying due to the TDMA frame structure and retransmission due to BLER. Flushing LLC_PDUs on user mobility is causing multiple TCP segments to be lost.

Various improvements of TCP have been introduced in the past to improve TCP for wireless links. The suggestions range from modifying TCP functions, e.g. selective acknowledgement [RFC2018] which is well established now, to split the TCP connection into a connection optimised for wireless link and a connection for the wired link, e.g. indirect TCP, [I-TCP95] or [SNOOP95].

In the following, I briefly summarise the most important TCP parameters - independent of the particular TCP implementation, such as TCP New RENO, [RFC2582] - to be considered when used on high latency and error prone, low bandwidth wireless link.

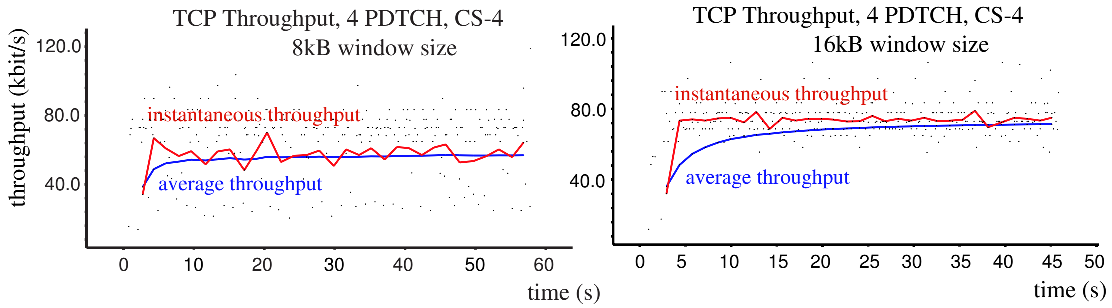
8.1.1 Optimising TCP/IP for Wireless Links

The most important TCP parameters affecting application performance over GPRS are summarised in the following. I focus on TCP parameters, which are common in the majority of Internet servers.

8.1.1.1 TCP Window Size

The advertised receive window (TCP window size) is most important TCP parameter of the client host. The sender is able to inject as many data into the network as the destination host advertises in a TCP segment sent back to the serving host (e.g. TCP acknowledgement). The TCP window has to be big enough to be able to utilise the available link bandwidth, on the other hand are unrealistically large

TCP window sizes causing high round trip time due to high delays on buffering in the network entities. In the worst case TCP segments are discarded on buffer overrun.
Recent operation systems use a default TCP window size of 16kB, which is sufficient to provide full throughput using CS-4 and 4 PDTCH. Figure 8.1 is comparing a 400kB file transfer on 4 timeslots at CS-4. The left graph showing the FTP throughput at 8kB TCP window stays about 21% below the maximum throughput at 16kB window size. TCP window sizes above 16kB are not recommended



TCP Receive Window	Duration (s)	Throughput (kbit/s)
8kB	57.25	57.25
16kB	45.15	72.56

Figure 8.1: 8kB vs. 16kB receive window size

on GPRS links to avoid additional queuing delays in the PCU. Increasing the TCP window to 64kB inevitably leads to retransmission time-outs due to the high queuing delay in the PCU, causing unnecessary TCP retransmissions. TCP window scaling, providing TCP window sizes above 64kB, is not required in GPRS networks, which also applies to EGPRS or even UMTS at 384kbit/s. In either case, the bandwidth-delay-product will stay below 64kB. UMTS connections up to 2Mbit/s are likely to require TCP window scaling in order to support maximum TCP performance.

8.1.1.2 MTU - Maximum Transmission Unit and MSS - Maximum Segment Size

The MTU is an IP parameter, defining the maximum size of a data packet of one physical frame on the network. Each transmission unit consists of a header and a data segment. The maximum size of the data segment is defined as MSS (Maximum Segment Size). Following equation can be derived:

$$MTU = MSS + TCPheader + IPheader.$$

The MTU or rather the MSS has to be considered when defining the advertised TCP window. The minimum TCP/IP protocol overhead can be achieved when setting the receive window to multiples of the MSS, e.g. for an MTU of 1500 (e.g. Ethernet), 20 byte IP header and 20 byte TCP header (without additional TCP options) the MSS will be 1460 bytes. The optimum link utilisation will be at a TCP receive window of 16.06kB (11 segments) or 17.52kB (12 segments). The server will then be able to send maximum size TCP segments (full-sized segments, see [RFC2581]) until the TCP window is filled up, minimising the TCP overhead. The effect is illustrated in the TCP server trace in appendix G.

8.1.1.3 TCP Congestion Control

The primary design goal of TCP is to avoid overloading the network and cause buffer overruns. Subsequently TCP interprets packet losses as network congestion. The congestion avoidance provided by TCP, as defined in [RFC2581], covers four major components.

- Slow start
- Congestion avoidance
- Fast retransmit
- Fast recovery

In addition, the start behaviour of TCP after a relatively long idle period is defined in RFC2581. I focus on slow start and fast retransmit since these features allow parameter optimisation in common TCP implementations.

TCP Slow Start In the beginning of a data transfer, TCP increases the number of segments sent but not yet acknowledged step-by-step, rather than dispatching a burst of data that may overload the network. Every acknowledgement received by the sender causes transmission of two new segments until the receivers advertised TCP window is filled up. TCP slow start is a mandatory feature to avoid packet losses in congested networks. Nevertheless, the probing for network bandwidth on TCP slow start has impact on connections, in particular over 4 PDTCH and CS-4. Slow start limits the number of TCP segments on start-up to 4 segments, as shown in figure 8.2. With 4 PDTCH and CS-4, no continuous data stream during TCP slow start can be provided. The maximum throughput is reached after 3 to 5 seconds. This effect is limiting the maximum application layer throughput over a high latency link when small objects are transferred in a TCP session. The time sequence graph from the client shows that 20kB are transferred within 4 seconds, resulting in a throughput of 40kbit/s. This is approximately 50% of the maximum throughput of CS-4 on 4 PDTCHs. Typical TCP implementations are even more cautious and start with two segments initially, rather than using four segments, making the effect of TCP slow start even more distinct.

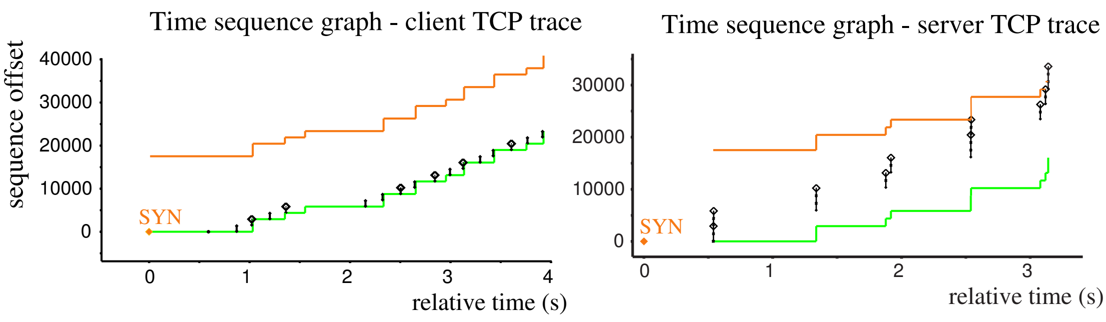


Figure 8.2: Effect of TCP slow start

Fast Retransmit and Fast Recovery Every segment received out-of-order causes the TCP destination to send a duplicate ACK. With fast retransmit, the sender is able to retransmit the missing segment after a certain number of duplicate ACK is received, instead of waiting of expiry of the retransmission time-out. Fast retransmit is typically set to 3, meaning that 4 identical ACK trigger the retransmission. Effects such as segment duplication or out-of-sequence delivery have to be considered to avoid unnecessary retransmissions, reducing the effective radio link bandwidth. Since about 11 or 12 segments are on the fly with a 16kB receive window, fast retransmit can be increased to 5, without losing the continuous data stream across the radio link. Fast retransmit is mostly implemented together with fast recovery, which aims to avoid slow start when only one segment is lost and no network congestion is assumed. Fast recovery is an algorithm to faster inflate the senders TCP window after a segment loss. The algorithm is based on the assumption that receipt of duplicate ACK indicates that after the missing segments further segments have been transmitted successfully, hence the amount of acknowledgement pending data in the network is reduced.

8.1.1.4 TCP Selective Acknowledgement

The TCP retransmission mechanism without selective acknowledgement turned out to have catastrophic impact on the session throughput when multiple segments of one TCP window are lost. TCP selective acknowledgement (SACK) [RFC2018] enables the TCP receiver to inform the sending entity about successfully received out-of-order segments. Thus, the sender only has to retransmit not acknowledged segments. The information about selectively acknowledged segments is included in the TCP options field. The peer's ability to process SACK is included in the three-way handshake of the TCP control session (SYN and SYN-ACK message).
The following TCP sequence graph in figure 8.3 shows the function of SACK on example of two segment losses. Every segment out-of-order is immediately acknowledged, delayed acknowledgement is not applied during selectively acknowledging.

The retransmitted segment arrives at the client after a full TCP window worth of

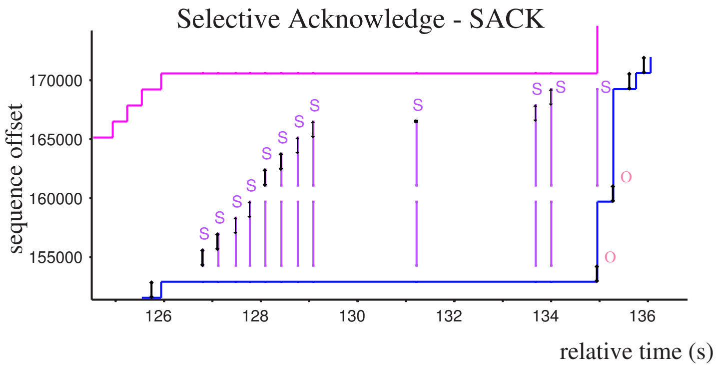


Figure 8.3: Example for selective acknowledgement

data is passing the PCU buffer and the radio link. At arrival of the retransmitted segments, all out-of-order segments are acknowledged. The TCP throughput is zero for approximately five seconds, during the phase of retransmitting the two missing segments. The server is sending a very small TCP segment (detected after 131.2 seconds on the client) indicating that the amount of acknowledgement pending data reached the advertised TCP window size. The same situation without selective acknowledgement would have lasted for an additional period of retransmission time-out until the transmission continues.
An example of multiple TCP segments lost within one TCP window and without implementing SACK is shown in figure 8.4. The first retransmission is triggered by fast retransmit. The retransmission of the following two TCP segments is purely triggered by TCP retransmission time-out, causing almost three seconds of inactivity in the data flow. The examples above confirm that TCP selective

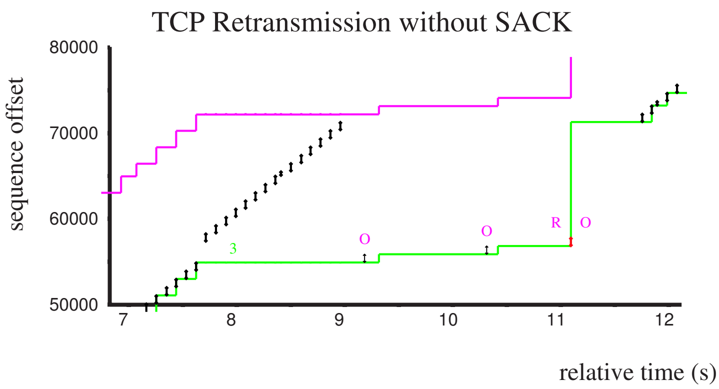


Figure 8.4: TCP retransmission without SACK

acknowledgement is a mandatory TCP feature on a high delay, low bandwidth

GPRS link.

8.1.2 TCP Parameter Optimisation

A summary of parameters to be optimised over GPRS in order to improve TCP performance is shown in the following.
The initial slow start has to be set to four segments to maximise the TCP throughput when transferring small files, especially when four PDTCHs are assigned to the MS and when CS-3/CS-4 is implemented in the network. The initial retransmission time-out has to be high enough to avoid spurious retransmissions on starting up the TCP session. An initial retransmission time-out of three seconds advisable when considering standby mode as worst case in terms of system latency. The advertised TCP receive window has to be big enough to achieve maximum TCP throughput. For GPRS CS-4 and 4 PDTCH a TCP window of 16kB is sufficient. The TCP window should be set to multiples of the MSS to be able to provide full-size TCP segments, thus reducing the TCP/IP protocol overhead. Fast retransmit should be increased as much as possible to avoid unnecessary retransmissions on out-of-sequence delivery. Triggering fast retransmission after receiving 5 duplicate ACK is cautious to avoid spurious retransmissions, but still not leading to intermittent transmission on the radio link. Finally, TCP selective acknowledgement is an important feature to recover quickly after multiple segment losses within one TCP window.
It has to be noted that TCP parameter optimisation is generally limited since access to all servers and clients connected to the network is not possible. From the GPRS network operator perspective, parameter recommendations to GPRS users can be provided, or on the network side, parameter settings in a TCP proxy server can be implemented to improve the TCP performance for a wide range of Internet hosts and applications, see clause 8.3.

8.2 Application Performance Analysis

In the following, I discuss the performance of typical mobile applications and their associated underlying protocols when used over GPRS. A major focus is the performance when higher coding schemes CS-3 and CS-4 or EGPRS coding schemes are enabled in the network.

8.2.1 FTP Performance Analysis

The File Transfer Protocol (FTP) provides reliable data transfer among hosts. FTP is standardised in [RFC959]. The standardisation of FTP makes the user (either directly a user or an application using FTP) independent of the host's

operating system. FTP is widely spread and can be fully controlled from the wireless client, what makes FTP a good application for benchmarking wireless networks.
FTP uses a control connection to exchange commands and replies based on the Telnet protocol. This control session uses server TCP-port 21 by default. The data transfer uses a single, full duplex TCP connection between the client and the FTP server addressed by TCP-port 20. From the GPRS performance perspective, FTP is very efficient since it uses just a single TCP connection to transfer the actual data, thus requiring small protocol overhead. Typical throughput graphs for an FTP download of 400kB are shown in figure 8.1.
It has to be noticed that at the beginning and at the end of a data transfer exchange of control messages is required, e.g. to indicate the successful end of a data transfer. The exchange of control messages is reducing the effective throughput seen by the user when multiple files are transferred or when an application uses several instances of FTP. Figure 8.5 is illustrating the FTP control message exchange and the TCP data session establishment in the beginning of the file transfer. The receipt of the first data segment happens about 1.5 seconds after requesting the file transfer.

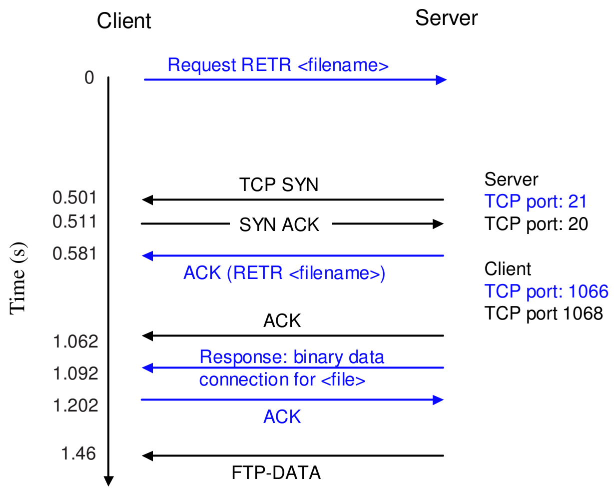


Figure 8.5: FTP - establishment of file transfer

The FTP connection as shown in figure 8.5 is already established and the user is logged in. Considering the entire procedure to start the FTP session and log on to the server would take at least two more round trips. This control session overhead makes the FTP unsuitable for interactive data transfer sessions, such as during web browsing. However, the maximum instantaneous throughput of the TCP data transfer session is 74kbit/s, indicating that the FTP data transfer utilises the entire bandwidth of the radio link.

The transfer of user data uses a single TCP session, hence the same effects of TCP slow start apply for FTP as for TCP, shown in clause 8.1.1.3. Figure 8.6, below is showing a FTP download of a 400kB file again. TCP slow start takes about four seconds until TCP is fully utilising the available bandwidth of the GPRS radio link. While the effect of TCP slow start is not distinct when transferring big files (relatively long transfer time compared to the slow start duration), the throughput at transfer of small files will suffer from TCP slow start. At 4 PDTCH and CS-4, TCP will not finish the slow start procedure when transferring files of about 20kB and the average throughput will reach about 50% of the maximum. The effect of

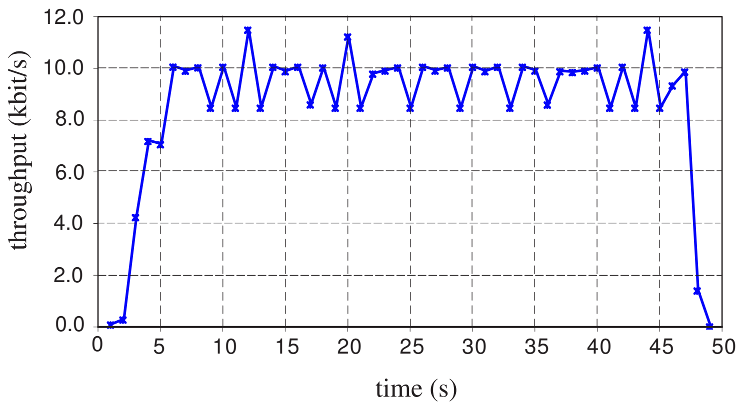


Figure 8.6: FTP throughput, CS-4, 4Rx

slow start on FTP performance is more distinct when higher coding schemes are used. Due to the increased TCP window size required at higher link bandwidth, the slow start mechanism will take a longer period of time until the TCP window is filled up and the maximum link bandwidth can be utilised.

8.2.2 HTTP Performance Analysis

HTTP, the Hypertext Transfer Protocol is the major protocol for information exchange used in the World Wide Web. HTTP has been developed to overcome the limitations and inefficiencies of FTP. Both protocols, FTP and HTTP use TCP as a reliable, connection oriented underlying transport protocol, the default port for the web server is TCP-port 80. HTTP is an application layer protocol for request-response interaction. HTTP is not only transferring files, such as FTP does, but also script outputs or dynamically generated query results from the server, enhancing the web browsing capabilities from read only to dynamic modification of web content, e.g. providing personalised web content, interaction with client input or back end applications. In any web transaction, two or more entities are involved. One is the web client, commonly called browser, and the other entity are one or several web servers. The browser communicates with a web server using one or several TCP connections. HTTP is the protocol used between client and server to communicate over the TCP connection. Once the client has established a TCP connection to the server, it issues an HTTP request through the TCP data stream, and reads back the server's HTTP response from the same TCP data stream (request-response interaction). The basic concept of request - response is illustrated in table 8.1, below.

Simple HTTP Request and Reply	
Request:	GET http://www.server.com/page.html HTTP/1.1 Request Method: GET
Response:	HTTP-Version: HTTP/1.1 200 OK Content-Length: 10100 Connection: close Content-Type: text/html <body>

Table 8.1: HTTP request - response

8.2.2.1 HTTP/1.0

A typical web page consists of a HTML (Hyper Text Markup Language) document with many embedded images or objects. The initial specification (HTTP/1.0, see [RFC1945]) was handling multiple requests, creating a separate TCP connection for the HTML document and for each and every embedded object. Since the embedded objects are typically small, the transfer finishes before the TCP slow start is finished. In addition, every TCP connection requires signalling overhead, leading to inefficient utilisation of resources. On the other hand, multiple TCP connections can send a high amount of data into the network, requiring buffer space as much as several TCP windows, potentially causing network congestion.

8.2.2.2 HTTP/1.1

The performance limitations of HTTP/1.0 have been addressed in the specification of HTTP/1.1, see [RFC2616]. HTTP/1.1 opens one connection for the transaction and keeps this connection open - *persistent connection* - which addresses the limitations of multiple consecutive connections, i.e. high signalling overhead, TCP slow start and buffer overrun in the transport pipe. In addition, HTTP/1.1 offers compression to reduce the transferred data volume. Within the persistent connection multiple request - response operations can be run simultaneously - *pipelining* - to utilise the offered TCP connection efficiently.
Pipelining is an essential feature in high latency GPRS networks since fetching each object separately requires one additional round trip for every object transferred, as illustrated in figure 8.7.

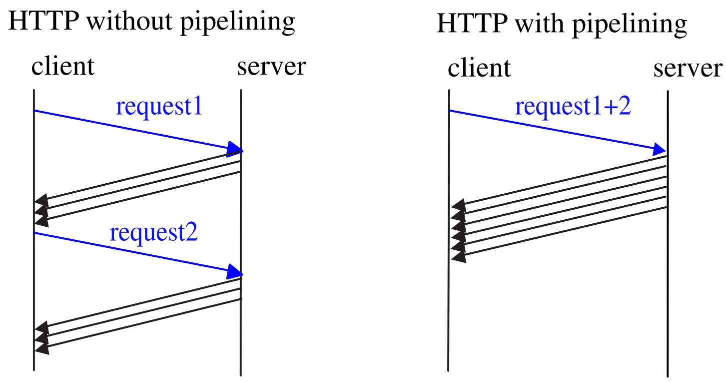


Figure 8.7: HTTP pipelining

8.2.2.3 HTTP Parameter Optimisation for GPRS Networks

Optimisation of HTTP parameters for GPRS networks is rather limited. The most important HTTP parameter on the browser side is the availability of request pipelining. Recent browser implementations already offer pipelining by default. The main feature affecting the web client as well as the server side (or on the proxy or cache) is the use of persistent connection. Several HTTP servers provide "Keep-Alive parameters" accordingly. The negotiation of the keep-alive duration is done in the GET-request and HTTP response message. The smaller value announced in both messages will be used for the persistent TCP connection. The example shown in table 8.1 will not use persistent connections, to be seen in the connection: close command sent by the server. Persistent TCP is probably not implemented in the HTTP-proxy or web server hosting this web site.

8.2.2.4 HTTP Tests over GPRS

The effect on HTTP performance over GPRS without persistent connections is illustrated in the following examples. The first web access is done with a fully HTTP/1.1 compliant browser, accessing the site the first time, without cached information on the client. The server does not support persistent connections, the entire download takes 21.6 s, comprises 176 TCP segments transferred across 12 TCP connections.

In a second test, I use a different browser, also HTTP/1.1 compliant, and I access the same web site. The site access is also done the first time through that browser to avoid caching. The entire download takes 17.9 s, comprises 257 TCP segments transferred across 19 TCP connections. Comparing both examples can be seen that the transferred data volume as well as the TCP connections and the transfer time highly vary, depending on the link conditions, packet retransmissions,

TCP connection info				
number	client TCP port	server TCP port	packets sent UL	packets received DL
1	1442	80	10>	11<
2	1443	80	10>	14<
3	1444	80	5>	5<
4	1445	80	7>	6<
5	1446	80	9>	8<
6	1447	80	6>	6<
7	1448	80	6>	5<
8	1449	80	12>	12<
9	1450	80	6>	5<
10	1451	80	6>	5<
11	1452	80	6>	5<
12	1453	80	6>	5<

Table 8.2: HTTP/1.1 non-persistent connection, test1

TCP connection info				
number	client TCP port	server TCP port	packets sent UL	packets received DL
1	1455	80	11>	11<
2	1456	80	6>	14<
3	1457	80	12>	5<
4	1458	80	7>	6<
5	1459	80	9>	8<
6	1460	80	7>	6<
7	1461	80	6>	5<
8	1462	80	13>	12<
9	1463	80	6>	5<
10	1464	80	6>	5<
11	1465	80	6>	5<
12	1466	80	6>	5<
13	1467	80	6>	5<
14	1468	80	5>	5<
15	1469	80	6>	5<
16	1470	80	6>	5<
17	1471	80	5>	5<
18	1472	80	6>	5<
19	1473	80	6>	5<

Table 8.3: HTTP/1.1 non-persistent connection, test2

and the HTTP implementation in the server and browser. For web access over GPRS can be concluded, that a HTTP proxy optimised for wireless network links and up to date implementations of the browsers are vital for optimised HTTP performance, especially since HTTP is one of the most important applications used over GPRS.

8.2.2.5 Maximum HTTP Performance over GPRS

The maximum performance of HTTP can be achieved by having a single, persistent TCP connection and when using HTTP pipelining. In that case, web access via HTTP performs similar to FTP and the performance effects are as for described for TCP, see clause 8.1. However, HTTP slightly reduces the maximum throughput compared to FTP due to the additional HTTP protocol overhead. The maximum instantaneous throughput will be 71.2kbit/s at HTTP compared to 73.6kbit/s at FTP, when using four PDTCH and CS-4.

8.3 Performance Enhancing Proxies

In the following, I summarise the most important performance enhancements for the use in GPRS networks and I am showing an implementation example, comparing the performance with and without proxy.

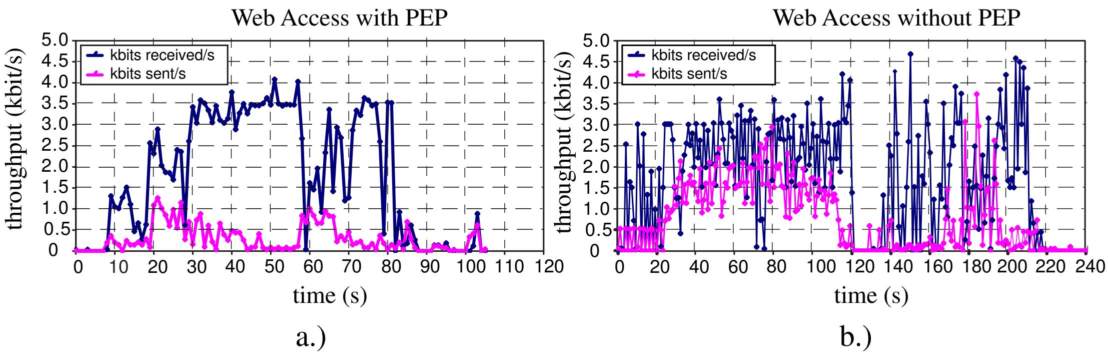
Several types of proxies can be distinguished, based on the protocol layers the proxy is acting on, or the location of the proxy in the network. PEPs can be working client less, i.e. all performance enhancements are controlled from the server side, or it can require a client software fulfilling protocol routines together with the stationary proxy server. Other typical performance enhancements are caching and pre-fetching of web pages to reduce access latency, content transformation to reduce image resolution or colour depth or compression of images, protocol headers or payload or protocol translation. A detailed description of different PEP implementations can be found in [RFC3135].

Finding an optimal solution for a given GPRS network has to consider the performance limitations that should be improved with the PEP. The following list shows the major performance limitations affecting GPRS performance that should be addressed when implementing a PEP.

- TCP slow start reducing the maximum throughput for short TCP sessions.
- TCP congestion control interprets packet losses due to transmission errors as congestion and throttles the data rate.
- Cell reselection on user mobility can cause multiple TCP segments to be dropped and annoying TCP or application disconnections.
- Many servers connected to the Internet rely on obsolete TCP protocol implementations or TCP parameters are optimised for wired connections rather than high latency and slow wireless links.
- HTTP/1.1 still not commonly used, many web servers implement HTTP/1.0 which is known to limit the performance in wireless networks.

- Highly presentation oriented web pages take a very long download time in wireless networks.

The possible benefit of implementing a PEP in a GPRS network is shown in the following example. The same web page is downloaded with a PEP and in a second



a.) b.)

Figure 8.8: Comparison, web download with and without PEP

test without a PEP. The PEP implements a protocol booster, using UDP data stream rather than TCP. An additional TCP connection controls the UDP stream and adds reliability. The PEP compresses images with configurable quality reduction; it pre-fetches the requested web page and provides caching for faster access. This implementation requires a booster software on the client to convert the UDP stream back into TCP/IP and HTTP. The data transmission rate on the client with PEP is shown in figure 8.8 a.). The GPRS connection provides CS-2 and four PDTCH, with a maximum data rate of 44kbit/s on TCP layer. After 106 seconds, the data transmission finishes and the browser displays the web page being downloaded successfully. After disabling the PEP, the same download takes about twice as long. 120s to 130s after starting the download, a 10-second gap is seen in the throughput graph. Similar to the test with PEP, the maximum application layer throughput is approximately 28kbit/s. The performance improvement for the example shown below is about 50% in terms of web page access time. However, the complexity of the requested web page and the fact that the image resolution is reduced by the PEP indicates, that a 50% improvement cannot be expected for every web page access over GPRS.
In summary it may be said that performance enhancing proxies are mandatory to improve the GPRS link characteristics based performance reductions of HTTP and TCP/IP and from my perspective even more important, avoid application disconnects over the wireless link.

Chapter 9

Summary and Conclusions

The aim of this thesis was the evaluate GPRS performance capabilities. I approach GPRS performance in different phases. First, I define the basic key performance metrics, i.e throughput, the amount of application layer information that can be transferred over the GPRS link in a certain period of time; and latency, the time a data packet needs to travel from one end of the GPRS network to the other. Based on measurement results of these key metrics it becomes obvious that the radio link is predominantly determining GPRS performance.
In the second phase, I analyse the GPRS protocol functions to identify the basic causes for latency and throughput limitations of GPRS networks.
The third analysis phase focuses on GPRS performance capabilities from the system level point of view, considering different applications used on a GPRS link, and finally I identify methods for end-to-end performance improvements.

Analysing the GPRS Protocol Stack

The analysis of the GPRS protocol functions shows that the RLC/MAC layer has decisive impact on the system performance. The latency of a GPRS network mainly comes from the protocol functions to set up a connection over the radio link (TBF set-up). In chapter 3, I show that the latency of a GPRS network is a result of the rigid TDMA frame structure, resulting in high waiting time until TBF set-up messages can be scheduled for transmission. In addition, the TDMA frame structure introduces high variability in the TBF set-up times.
When re-using the existing GSM signalling channels for GPRS, for a given control channel configuration the set-up of an uplink TBF can take between 130ms and 470ms, depending on the MS requesting multiple uplink timeslots or just a single uplink timeslot. Setting up a downlink TBF takes between 187ms and can last up to 3150ms when the MS is in standby state. An improvement of TBF set-up times can be achieved when using the optional PBCCH/PCCCH feature, since the control channel structure on the PCCCH is more flexible and more suitable to

GPRS traffic behaviour than the original voice signalling channels are. However, the deployment of PBCCH is limited since not all MS types support PBCCH. The RLC/MAC layer also determines the bandwidth of the radio link and subsequently the maximum throughput over GPRS. The key parameters determining the GPRS bandwidth are coding scheme usage, the RLC/MAC block error rate, and the MS multislot assignment. The multislot assignment is the most influential parameter. I discuss the performance implications of multislot usage in chapter 4. The specifications define 29 multislot classes; out of these 29 classes only class 1 to class 12 are commonly deployed. These multislot classes work in semi-duplex mode, which limits the maximum multislot assignment to four timeslots in one link direction, since the mobile station's front end has to re-synchronise the frequency when switching between send and receive, and, in addition, neighbour-signal-strength measurements have to be performed during every TDMA multiframe. Besides these timing restrictions, the uplink multislot usage is further limited due to the implementation of the dynamic uplink allocation, which limits these multislot classes to two uplink timeslots only. Extended dynamic allocation will allow using three uplink timeslots. The theoretical maximum of four uplink timeslots for MS class 12 will most likely not be feasible due to specification issues for the timing definitions. The maximum multislot assignments are either four downlink and one uplink timeslot, or three downlink and two uplink timeslots. For a network supporting coding scheme CS-1 and CS-2 a maximum downlink throughput of 44kbit/s[1] and a maximum uplink throughput of 33kbit/s can be expected. A network supporting CS-3, CS-4 will provide maximum throughput results of 74kbit/s for four downlink PDTCH and 54kbit/s for two uplink PDTCH. The deployment of higher coding schemes CS-3 and CS-4 has attracted increased attention in the past. Chapter 6.1 shows that CS-4 allows up to 60% throughput improvement compared to CS-2. Static users will greatly benefit from CS-4, in a very wide range of the cell coverage area and for different environments. Moving GPRS users will still be able to get high throughput in rural environment. Only in the outer third of the cell coverage area the performance of CS-4 drops significantly. But as soon as the mobile client is moving in urban environment, the CS-3 and CS-4 usage can only be used in close distance to the cell site, with good coverage level and low interference, and particular line-of-sight to the base station. Moving GPRS users in urban environment will experience an average throughput improvement due to higher coding schemes of 30%, even for moderate speed.

The next protocol entities above RLC are the link layer protocols, LLC and SD-DCP. The LLC protocol provides reliability, acknowledged mode, and retransmission of discarded LLC frames. SNDCP mainly provides TCP header and data compression as link layer features. In chapter 7, I show that compression as well as acknowledged LLC mode are not used in commercial GPRS networks. Ac-

[1] Throughput results on application layer based on TCP/IP, without compression.

knowledged LLC mode requires high amount of memory and CPU resources in the SGSN. In addition can competitive retransmissions on link layer and TCP not be excluded. It is a common approach in recent GPRS networks to improve reliability on the RLC layer, which is operated in acknowledged mode, and minimise LLC frame loss on cell reselection. The minimised number of retransmissions are taken over by higher layer protocols such as TCP or application protocols. A performance enhancing proxy, which is optimised for use in wireless networks, prevents TCPs congestion control from reducing the throughput on LLC frame loss on the wireless link.
Similar considerations apply to the SNDCP compression algorithms. Link layer compression requires a high amount of CPU resources in the MS and SGSN. Compression should be handled by higher protocols rather than in the GPRS link layer. Lastly, only few MS types support TCP header and data compression.

Another performance relevant protocol is BSSGP, see section 7.5. BSSGP acts on the Gb-interface and it provides flow control in downlink direction, to adapt the data transmission from a high bandwidth IP backbone to a low bandwidth radio link. It turned out that the feedback mechanism of the leaky bucket algorithm is a challenge for the SGSN, in terms of accuracy of predicting the PCU buffering state, but also when LLC-PDUs are forwarded within the PCU on cell reselection. The performance issues of the Gb-interface flow control have been addressed in the release 1999 of the technical specifications, which provides an additional parameter in the flow control messages.

Other GPRS protocols acting on the IP backbone have minor impact on GPRS performance compared to the radio protocols. The protocol layers identified as being performance relevant for the most part are shown below and in figure 2.4.

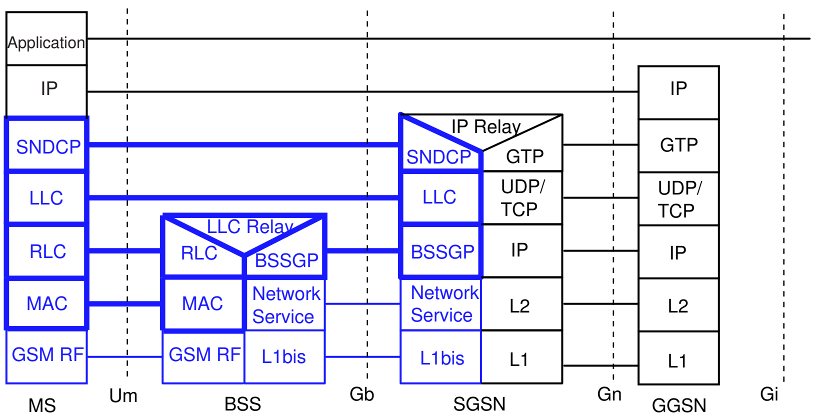


Figure 9.1: Protocol layers mainly determining GPRS performance

GPRS Performance - System Level Considerations

From the system level point of view a GPRS network is a highly latent and low-bandwidth link. In addition, both, the bandwidth as well as the latency, show high variability. The reliability of GPRS networks can be reduced on user mobility, since cell reselection between two BSS areas is inevitably leading to loss of LLC frames, which have already been buffered and scheduled for transmission in the source cell. These properties of GPRS networks interact with the higher layer application protocols. Especially TCP, the most commonly used transport protocol is challenged on GPRS networks. TCP was originally designed for wired networks with low latency and high bandwidth. TCP interprets lost TCP segments, e.g. due to cell reselection, as network congestion and reduces the data throughput, an effect that is clearly noticeable on a low bandwidth network. In addition, the TCP slow start, which was also designed to avoid network congestion, is also clearly reducing the throughput, especially when a small amount of data is transferred using that TCP session.
It can also be seen that legacy TCP implementations result in low application throughput, but, even worse, disconnected TCP sessions and therefore disconnections on application layer happen. While reduced data throughput across GPRS is unpleasant but acceptable for the end user performance perception, a disconnected session is very annoying for the GPRS user. The only option to improve the TCP performance over GPRS is to implement a performance enhancing proxy. Several concepts are available, e.g. optimise the TCP connection for a wireless RAN[2], provide caching, or compression. Providing GPRS network access for typical office applications such as web access, corporate email, or access to file servers makes a performance enhancing proxy mandatory to overcome the shortcomings in the transport layer as well as application protocols.

Evolution of GPRS

The initial GPRS implementations have been very limited in terms of functionality. Only a subset of the Release 1997 specifications was implemented in the beginning. After resolving interoperability issues between RAN and GSN vendors or MS vendors, and with further development of release 1997 features, quality of service (QoS) is now approaching. A further development step will be the introduction of NACC, the network assisted cell change, as defined in release 4 of the 3GPP specifications. NACC aims to reduce cell reselection times by supplying the MS with information about the target cell before the MS leaves the source cell, which enables the MS to quickly connect to the new cell. This feature requires changes in the network and in the MS. NACC most likely will be deployed within the next year.

[2] Radio Access Network

The major GPRS enhancement, which is going to be deployed commercially in Europe in the near future, is EDGE. EDGE is a release 1999 feature that mainly introduces changes in the radio part of the network and the RLC/MAC layer of the GPRS protocol suite. Enhanced channel coding strategies and forward error correction capabilities offer about three times the code rate of GPRS coding scheme CS-4. EDGE is fully backward compatible with GPRS, the only disadvantages are the increased backhaul bandwidth on the A_{bis}-interface and that only recent BTS equipment is EDGE compatible. EDGE is not commercially deployed in European GSM networks now, which is mostly be rooted in the limited availability of EDGE capable handsets. EDGE has high potential as capacity solution in areas with limited spectrum, as backup in areas without UMTS coverage or for network operators without UMTS licence. However, a general deployment of EDGE in Europe is still unclear now. It mainly depends on the investments in 2G infrastructure compared to the deployment of UMTS infrastructure.
Beside the discussions on the infrastructure side are wireless applications often discussed. Push-to-talk (PTT), a walkie-talkie kind of application, is being standardised now. PTT will be a bridge between non-real-time SMS and voice calls, and it appears to receive high attention from network operators, especially in the US. PTT can be implemented independently of the underlying network, but again, the PTT functionality has also to be implemented in the handsets. Similar to MMS or EDGE, the MS is the limiting factor of commercial deployment.

GPRS Compared to Other Wireless Packet Data Networks

GPRS played a pioneer role in the wireless data network landscape and it still plays the pioneer role for features like QoS, which are neither in GPRS nor in UMTS available now. But while GPRS is still not fully developed and usage of wireless data services in 2G networks is still low compared to voice services, GPRS is obviously going to be overtaken by the proceeding deployment of UMTS. Improved GPRS features are now deployed with UMTS at the same time. UMTS recently suffered from lack in handsets (UE, User Equipment) and the grade of service for circuit-switched voice calls is not reaching the performance targets of 2G networks. Especially the limited functionality of handover between 2G and 3G limits commercial deployment, since full UMTS coverage cannot be provided now, and UMTS coverage borders are often resulting in dropped calls even though 2G service coverage would be available. The position of GPRS compared to UMTS will mainly depend on the improvement of the 3G infrastructure and especially improvements on the UEs but also on the predicted end-of-life for 2G networks as investments in legacy network infrastructure has to done carefully to guarantee return of investment. Other wireless access technologies such as WLAN have been included in the service portfolio of wireless network operators. WLAN is very successful in home our enterprise environment. The success in wireless data networks depends on how seamless mobility is supported, the availability of handsets and the offered capacity. The current market conditions point on a coexistence of 2G, UMTS and WLAN for wireless data networking. I am assuming that most likely, 2G and UMTS are predominantly used for circuit-switched voice traffic. GPRS will remain being the network wide backbone for data services, while other technologies are going to handle data traffic in hot spot areas at higher data rates or urban environment.

Appendices

A. Two-Phase Access - Decoded Messages

```
SYSTEM INFORMATION TYPE 13
Protocol Discriminator:
6 Radio resource management message
Skip Indicator:0 This message is not to be ignored
Message Type:0 System Information Type 13
SI 13 Rest Octets
 BCCH_CHANGE_MARK:77
 SI_CHANGE_FIELD:0 Update of unspecified
 SI message or SI messages
 RAC:5
 SPGC_CCCH_SUP:1 SPLIT_PG_CYCLE
 is supported on CCCH in this cell
 PRIORITY_ACCESS_THR:6 Packet access
 is allowed for priority level 1 to 4
 NETWORK CONTROL ORDER:0 NC0:
 MS controlled cell re-selection, no measurement reporting
 NMO:1 Network Mode of Operation 2
 T3168:1 (1000msec)
 T3192:0 (500msec)
 DRX_TIMER_MAX:0 (0)
 ACCESS_BURST_TYPE:1
 11 bit access burst shall be used
 CONTROL_ACK_TYPE:0
 default format is four access bursts
 BS_CV_MAX:6
 PAN_DEC:1
 PAN_INC:1
 PAN_MAX:1
 maximum value allowed for counter N3102 is 8
 ALPHA:0
 T_AVG_W:9
 T_AVG_T:13
 PC_MEAS_CHAN:0 BCCH
 N_AVG_I:4
```

Table a-1: System information 13 message

```
CHANNEL REQUEST
- Establishment Cause Description
 One phase packet access with request for
 single timeslot uplink transmission,
 one PDCH is needed
- Binary:0x7C
```

Table a-2: Channel request, two-phase access

```
IMMEDIATE ASSIGNMENT
-Protocol Discriminator:6 Radio resource management message
 -Skip Indicator:0 This message is not to be ignored.
 -Message Type:63 Immediate Assignment
 -Page Mode:3 Same as before
 -Dedicated Mode or TBF
  TMA:0 No Meaning
  Downlink:0 No Meaning
  T/D:1 This message assigns a temporary block flow (TBF)
 -Packet Channel Description
  Channel Type:1
  TN:7
  TSC:2
  ARFCN:123
 -Request Reference
  FN:20618
  Request Reference:124
  Description:One phase packet access with request for
  single timeslot uplink transmission; one PDCH is needed
 -Timing Advance
  Timing Advance:0
  Description:0.000000 microseconds
 -IA Rest Octets
  ALPHA:0 a = 0.0
  GAMMA:0
  TBF_STARTING_TIME:20648
```

Table a-3: Immediate assignment, two-phase access

```
PACKET RESOURCE REQUEST
MAC Header
Retry:  MS sent channel request message once (0)
Access Type:  Mobility Management procedure (3)
Temporary Logical Link Identity:  2147485469 (0x8000071d)
Mobile Radio Capability
 Access technology type:  GSM E (1)
  RF Power Capability:  4
  A5/1:1 A5/2:1 A5/3:0 A5/4:0 A5/5:0 A5/6:0 A5/7:0
  Controlled early classmark sending indicator:  Implemented (1)
  Pseudo synchronisation capability:  Not Present (0)
  Voice group call service capability:
   No VGCS Capability or no notifications wanted (0)
  Voice broadcast service capability:
   No VBS capability or no notifications wanted (0)
 Multislot capability
  GPRS Multislot Class:  RX=3 TX=1 SUM=4
  TTA=3 TTB=1 TRA=3 TRB=1 TYPE=1 (4)
  GPRS Extended Dynamic Allocation Capability:  Not Implemented (0)
 Access technology type:  GSM 1800 (3)
  RF Power Capability:  1
  Controlled early classmark sending indicator:  Implemented (1)
  Pseudo synchronisation capability:  Not Present (0)
  Voice group call service capability:
   No VGCS Capability or no notifications wanted (0)
  Voice broadcast service capability:
   No VBS capability or no notifications wanted (0)
 Channel Request Description
  Peak Throughput Class:  9
  Radio Priority:  Radio priority 1 (0)
  RLC Mode:  Acknowledged (0)
  LLC PDU Type:  Not SACK or ACK (1)
  RLC Octet Count:  19
  C_Value:  44
```

Table a-4: Packet resource request, two-phase access

```
PACKET UPLINK ASSIGNMENT
MAC Header
Relative Reserved Block Period:
Uplink block with TDMA frame number =
(N+13) mod 2715648 (0)
Supplementary/Polling:
Relative Reserved Block Period is not valid (0)
Uplink State flag:  7
Page Mode:  Normal Paging (0)
Temporary Logical Link Identity:  2147485469 (0x8000071d)
Channel Coding Command:  CS-2 (1)
TLLI Block channel coding:  The MS shall use CS-1
for any RLC data block containing a TLLI
in the RLC data block header (0)
- Packet Timing Advance
  Timing Advance Index:  10
  Timing Advance Timeslot:  7
- Frequency Parameters
  Training sequence code:  2
  ARFCN: 123
- Dynamic Allocation
  Extended Dynamic Allocation:  0
  USF Granularity:  One RLC/MAC block (0)
  Uplink Temporary Flow Indicator:  16
  USF timeslot number7:  0
```

Table a-5: Packet uplink assignment, two-phase access

```
UPLINK DATA BLOCK
MAC Header
 Countdown Value:  0
 Stall indicator:  MS RLC transmit window is not stalled (0)
 Retry:  MS sent channel request message once (0)
 Temporary Flow Identifier:  16
 TLLI Indicator:  Not Present
 Block Sequence Number:  0
```

Table a-6: Uplink data block, two-phase access

B. One-Phase Access - Decoded Messages

```
UPLINK DATA BLOCK
 MAC Header
  Countdown Value:  5
  Stall indicator:  MS RLC transmit window is not stalled (0)
  Retry :  MS sent channel request message once (0)
  Temporary Flow Identifier:  0
  TLLI Indicator:  Present
  Block Sequence Number:  0
 Temporary Logical Link Identity:  3301166573 (0xc4c3cded)
```

Table a-7: Uplink data block with TLLI

```
IMMEDIATE ASSIGNMENT
 Protocol Discriminator:6 Radio resource management message
 Skip Indicator:0 This message is not to be ignored.
 Message Type:63 Immediate Assignment
 Page Mode:0 Normal paging
 Dedicated Mode or TBF
  TMA:0 No Meaning
  Downlink:0 No Meaning
  T/D:1 This message assigns a temporary block flow (TBF)
 Packet Channel Description
  Channel Type:1
  TN:7
  TSC:0
  ARFCN:123
 Request Reference
  FN:3203
  Request Reference:121
  Description:One phase packet access with request for
  single timeslot uplink transmission; one PDCH is needed
 Timing Advance
  Timing Advance:0
  Description:0.000000 microseconds
 IA Rest Octets
  TFI_ASSIGNMENT:0
  POLLING:0 The mobile station shall transmit one RLC/MAC block
  USF:0
  USF_GRANULARITY:0
  CHANNEL_CODING_COMMAND:1 CS-2
  TLLI_BLOCK_CHANNEL_CODING:1 The MS shall use coding scheme as
  specified by the CHANNEL_CODING_COMMAND field
  GAMMA:17
  TIMING_ADVANCE_INDEX:9
```

Table a-8: Immediate assignment, one-phase access

C. Concurrent Uplink TBF Set-up

```
PACKET DOWNLINK ACK/NACK
 Message Details
MAC Header
Retry:  0
Downlink Temporary Flow Indicator:  25
ACK/NAK Description
  Final ACK Indicator:  No retransmission requested
  - all RLC data in TBF is acknowledged (1)
Channel Request Description
  Peak Throughput Class:  2
  Radio Priority:  Radio priority 2 (1)
  RLC Mode:  Acknowledged (0)
  LLC PDU Type:  Not SACK or ACK (1)
  RLC Octet Count:  67
Channel Quality Report
  C_Value:  47
  RXQual:  6
  Signal variance:  (3F) : >15.75 dB^2 to 16 dB^2
  I_LEVEL_TN6:  I_LEVEL_2
  I_LEVEL_TN7:  I_LEVEL_6
```

Table a-9: Uplink TBF request included in PDAN

D. Mobile Multislot Classes

Table a-10 is an extraction from [3GPP-TS05.02], describing the MS multislot classes and their timing parameters.

Multislot class	Max. number of slots Rx	Tx	Sum	Min. number of slots T_{ta}	T_{tb}	T_{ra}	T_{rb}	Type
1	1	1	2	3	2	4	2	1
2	2	1	3	3	2	3	1	1
3	2	2	3	3	2	3	1	1
4	3	1	4	3	1	3	1	1
5	2	2	4	3	1	3	1	1
6	3	2	4	3	1	3	1	1
7	3	3	4	3	1	3	1	1
8	4	1	5	3	1	2	1	1
9	3	2	5	3	1	2	1	1
10	4	2	5	3	1	2	1	1
11	4	3	5	3	1	2	1	1
12	4	4	5	2	1	2	1	1
13	3	3	NA	NA	a)	1	a)	2
14	4	4	NA	NA	a)	2	a)	2
15	5	5	NA	NA	a)	3	a)	2
16	6	6	NA	NA	a)	2	a)	2
17	7	7	NA	NA	a)	1	0	2
18	8	8	NA	NA	0	0	0	2
19	6	2	NA	3	b)	2	c)	1
20	6	3	NA	3	b)	2	c)	1
21	6	4	NA	3	b)	2	c)	1
22	6	4	NA	2	b)	2	c)	1
23	6	6	NA	2	b)	2	c)	1
24	8	2	NA	3	b)	2	c)	1
25	8	2	NA	3	b)	2	c)	1
26	8	4	NA	3	b)	2	c)	1
27	8	4	NA	2	b)	2	c)	1
28	8	6	NA	2	b)	2	c)	1
29	8	8	NA	2	b)	2	c)	1

Table a-10: Definition of MS multislot timing requirements

Key:

a) = 1 with frequency hopping
= 0 without frequency hopping

b) = 1 with frequency hopping or change from Rx to Tx
= 0 without frequency hopping and no change from Rx to Tx

c) = 1 with frequency hopping or change from Tx to Rx
= 0 without frequency hopping and no change from Tx to Rx

$T_t x$ = Time to get ready to transmit
$T_r x$ = Time to get ready to receive
Index a = A measurement report is required to be made by the MS
Index b = No measurement report has to be made by the MS
Type 1 MS are not required to transmit and receive at the same time.
Type 2 MS are required to be able to transmit and receive at the same time.

E. Negotiation of RLC Mode

```
ACTIVATE PDP CONTEXT ACCEPT
Protocol Discriminator:10 GPRS Session management messages
Transaction Identifier
  TI Flag:1 The message is sent to the side that originates the TI
  TI Value:0
Message Type:66 ACTIVATE PDP CONTEXT ACCEPT
Negotiated LLC SAPI:11
Negotiated QoS
  Delay Class:4 Delay Class 4(best effort)
  Reliability Class:3 Unacknowledged GTP and LLC
  acknowledged RLC
  protected data
  Peak throughput:9 up to 256000 octet/s
  Precedence class:2 Normal Priority
  Mean throughput:8 20000 octet/h
Radio Priority:4 Priority Level 4
Spare half octet:0
PDP address
  Dynamic PDP addressing:0
  PDP Type organisation:1 IETF allocated address
  PDP Type Number:33 IPv4 address
  Address:123.123.123.123
Protocol Config Options
  Configuration Protocol:0 PPP
  Protocol Id:32801 IPCP
  IPCP Compression Protocol:0 Unrecognised Compression protocol
  Additional data :0xD388116B8306D38812ABC0231D
  Protocol Id:0 Unrecognised protocol identifier
```

Table a-11: Activate PDP context accept

F. Measurement Results - UDP Throughput vs. C/I

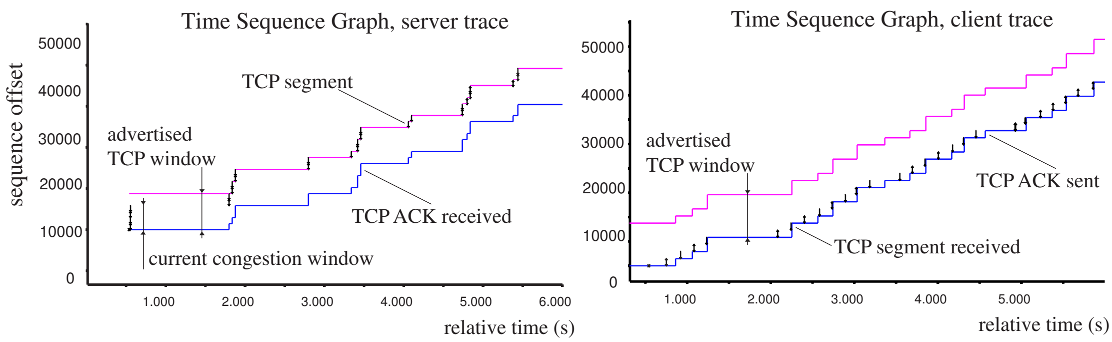
C/I_{co}	CS-4 [bit/s]	CS-3 [bit/s]	CS-2 [bit/s]	CS-1 [bit/s]
20	74100	52800	44400	29680
19	74074	53440	44000	29680
18	74534	52560	44240	29600
17	72727	52880	44320	29680
16	69767	53040	44080	29600
15	66298	52880	44080	29680
14	54054	52800	44400	29680
13	47619	52960	44160	29680
12	43010	52880	44320	29600
11	35294	52560	44320	29680
10		51920	44320	29600
9		46800	44320	29680
8		37440	43520	29680
7		27680	41120	29600
6		28960	22560	27520
5		28640	25360	28320

Table a-12: UDP throughput vs. coding scheme, 4 Rx

G. Description of TCP Segment Graphs

TCP Time-Sequence Graph

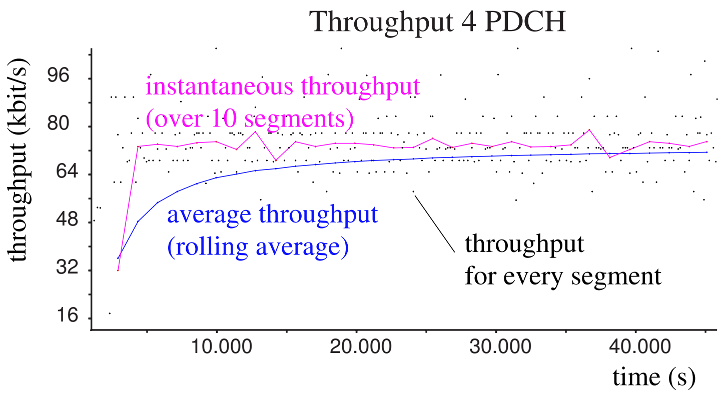
The TCP sequence graphs shown in this thesis have been created with tcptrace, see: http://www.tcptrace.org. A brief description of graph format is given below. The black vertical arrows indicate the point in time when a TCP segment



is detected by the receiving host or when it is passing through the TCP protocol layer in the sending host. The length of the vertical line is determined by the TCP segment size in byte, y-axis corresponds to start- and end-sequence number. In the server graph, the lower, blue line indicates receipt of TCP ACKs. The upper, pink line is parallel to the ACK-line the vertical offset is determined by the client's advertised TCP window. The current congestion window will be below, or as big as the advertised TCP window. The congestion window is a flow control parameter in the server. The client TCP graph shows receipt of TCP segments and the blue line shows sending of TCP ACKs accordingly. The client graph is not showing the congestion window, which is a TCP protocol parameter of the sending entity.

TCP Throughput Graph

The throughput graphs are also generated by tcptrace. The black dots denote the



throughput for every TCP packet by means of (TCP segment size) / (inter-arrival time of TCP segments). The instantaneous throughput is a considering a 10 TCP segments average, the average throughput is the rolling average of TCP segments over time.

Abbreviations

2G	Second Generation mobile communication networks
3G	Third Generation mobile communication networks
3GPP	Third Generation Partnership Project
A	
AFN	Absolute Frame Number
AGCH	Access Grant Channel
AGCH	Access Grant Channel
ARQ	Automatic Repeat Request
B	
BCCH	Broadcast Common Control Channel
BDP	Bandwidth Delay Product
BEC	Backward Error Correction
BLER	Block Error Rate
BSC	Base Station Controller
BSN	Block Sequence Number
BSS	Base Station System
BSSGP	Base Station System GPRS Protocol
BTS	Base Transceiver Station
C	
C/I	Carrier to Interference Ratio
CCCH	Common Control Channel
CDR	Charging Data Record
CPU	Central Processing Unit
CR	Channel Request
CRC	Cyclic Redundancy Check
CS-x	Coding Scheme (x =1,2,3,4)
D	
DL	Downlink
DRX	Discontinuous Reception
E	

EDGE	Enhanced Data rates for GSM Evolution
ERP	Effective Radiated Power
ETSI	European Telecommunication Standards Institute
F	
FCS	Frame Check Sequence
FEC	Forward Error Correction
FIFO	First In First Out
FPDAN	Final Packet Downlink ACK/NACK
FPUAK	Final Packet Uplink ACK/NACK
FTP	File Transfer Protocol
G	
GERAN	GSM/EDGE Radio Access Network
GERAN	GPRS/EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GTP	GPRS Tunnelling Protocol
H	
HLR	Home Location Register
HSCSD	High-Speed Circuit-Switched Data Service
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
HTTP	Hypertext Transfer Protocol
I	
ICMP	Internet Control Message Protocol
IP	Internet Protocol
K	
kB	Kilobyte
kb/s	Kilobit per second
km/h	Kilometer per hour
L	
LLC	Logical Link Control
LTN	Long Thin Network
M	
MAC	Media Access Control
MB	Megabyte
Mb/s	Megabit per second
MCS-x	Modulation and Coding Scheme (x =1..9)

MM	Mobility Management
MMS	Multimedia Messaging Service
MS	Mobile Station
MSS	Maximum Segment Size
MSS	Maximum Segment Size
MTU	Maximum Transmission Unit
N	
NCH	Notification Channel
N-PDUs	Network-Packet Data Units
P	
PACCH	Packet Associated Control Channel
PC	Personal Computer
PCA	Packet Control Acknowledgement
PCH	Paging Channel
PCTA	Packet Control Timing Advance
PCU	Packet Control Unit
PDA	Packet Downlink Assignment
PDAN	Packet Downlink ACK/NACK
PDCH	Packet Data Channel
PDN	Packet Data Network
PDTCH	Packet Data Traffic Channel
PDU	Packet Data Unit
PEP	Performance Enhancing Proxy
PPP	Point-to-Point Protocol
PSK	Phase Shift Keying
PTR	Packet Timeslot Reconfigure
PTT	Push to Talk
PUA	Packet Uplink Assignment
PUAN	Packet Uplink ACK/NACK
Q	
QoS	Quality Of Service
R	
RA250	Rural Area, 250km/h MS speed
RACH	Random Access Channel
RAN	Radio Access Network
RLC	Radio Link Control
RLP	Radio Link Protocol
RR	Radio Resource

RTD	Round Trip Delay
RTD	Round Trip Delay
RTT	Round Trip Time
S	
SACK	Selective Acknowledgement
SAPI	Service Access Point Identifier
SGSN	Serving GPRS Support Node
SNDCP	Subnetwork Dependent Convergence Protocol
T	
TBF	Temporary Block Flow
TCH	Traffic Channel
TCP	Transport Control Protocol
TDMA	Time Division Multiple Access
TE	Terminal Equipment
TFI	Temporary Flow Identify
TLLI	Temporary Logical Link Identity
TN	Timeslot Number
T-PDU	Payload (PDU) passed through a GTP tunnel
TS	Technical Specification
TS	Timeslot
TU3	Typical Urban, 3km/h MS speed
TU50	Typical Urban, 50km/h MS speed
U	
UDP	User Data Protocol
UE	User Equipment
UL	Uplink
USF	Uplink State Flag
USF	Uplink State Flag
V	
VoIP	Voice over IP
W	
WAP	Wireless Application Protocol
WLAN	Wireless Local Area Network
X	
XID	EXchange IDentification command/response

Bibliography

[BER-FER98] Wigard, J.; Nielsen, T.T.; Michaelsen, P.H.; Mogensen, P., "BER and FER Prediction of Control and Traffic Channels for a GSM Type of Air-Interface", Vehicular Technology Conference, 1998. VTC 98. 48th IEEE, Volume: 2, 18-21 May 1998, Pages:1588 - 1592 vol. 2.

[BER-WNMM98] Wigard, J.; Nielsen T.; Michaelsen P. H.; Mogensen P., "BER and FER prediction of control and traffic channels for a GSM type of air-interface", in Proc. IEEE VTC'98, pp.1588-92, Ottawa 1998.

[COST-207] Digital Land Mobile Radio Communications - COST207 Final Report, Commission of the European Communities, 1989.

[COST-231] COST231 Final Report - Digital Mobile Radio Towards Future Generation Systems, Commission of the European Communities, 1998, ISBN 92-828-5416-7.

[COST-259] COST259 Final Report - Wireless Flexible Personalized Communications, May 2001, ISBN 0-471-49836-X.

[3GPP-TS03.60] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Digital cellular telecommunications system (Phase 2+); "General Packet Radio Service (GPRS); Service description", Stage 2, Version 7.9.0., Release 1998, September 2002.

[3GPP-TS03.64] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; "General Packet Radio Service (GPRS), Overall description of the GPRS radio interface"; Stage 2; Version 8.10.0., Release 1999, February 2002.

[3GPP-TS04.18] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; "Mobile radio interface layer 3 specification, Radio Resource Control Protocol", Version8.14.0., Release 1999, 3GPP, April 2002.

[3GPP-TS04.60] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE, Radio Access Network, General Packet Radio Service

(GPRS); Mobile Station (MS) - Base Station System (BSS) interface; "Radio Link Control/ Medium Access Control (RLC/MAC) protocol", Release 1999, Version 8.14.01. 3GPP, May 2002.

[3GPP-TS04.65] 3rd Generation Partnership Project; Technical Specification Group Core Network, Digital cellular telecommunications system (Phase 2+), General Packet Radio Service (GPRS); Mobile Station (MS) - Serving GPRS Support Node (SGSN); "Subnetwork Dependent Convergence Protocol (SNDCP)", Release 1999, Version 8.2.0. 3GPP, September 2001.

[3GPP-TS05.02] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network, "Multiplexing and multiple access on the radio path", Release 1999, Version 8.10.0. 3GPP, August 2001.

[3GPP-TS05.03] Technical Specification; 3rd Generation Partnership Project; Technical Specification Group GERAN; "Channel Coding", Release 1999, Version 8.6.1. 3GPP, January 2001.

[3GPP-TS05.08] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE, Radio Access Network; "Radio subsystem link control", Release 1999, Version 8.12.0. 3GPP, November 2001.

[3GPP-TS05.10] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network, Digital cellular telecommunications system (Phase 2+), "Radio subsystem synchronisation", Release 1999, Version 8.9.0. 3GPP, February 2002.

[3GPP-TS07.07] 3rd Generation Partnership Project; Technical Specification Group Terminals, "AT command set for GSM Mobile Equipment (ME)", Release 1997, Version 6.5.0. 3GPP, March 2001.

[3GPP-TS08.18] 3rd Generation Partnership Project; Technical Specification Group GSM EDGE Radio Access Network; General Packet Radio Service (GPRS); Base Station System (BSS) - Serving GPRS Support Node (SGSN); BSS GPRS Protocol (BSSGP), Release 1999, Version 8.10.0. 3GPP, May 2002.

[3GPP-TS29.060] 3rd Generation Partnership Project; Technical Specification Group Core Network; General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface, Release 1999, Version 3.16.0. 3GPP, March 2003.

[3GPP-TS23.060] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) Service description; Stage 2, Release 1999, Version 3.10.0, January 2002.

[3GPP-TS32.015] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication Management; Charging management; Call and event data for the Packet Switched (PS) domain, Release 1999, Version 3.11.0, March 2003.

[3GPP-TS44.060-540] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; General Packet Radio Service (GPRS); Mobile Station (MS) - Base Station System (BSS) interface; Radio Link Control/Medium Access Control (RLC/MAC) protocol, Release 5, Version 5.4.0, December 2002.

[3GPP-TS44.064] 3rd Generation Partnership Project; Technical Specification Group Core Network; Mobile Station - Serving GPRS Support Node (MS-SGSN); Logical Link Control (LLC) layer specification, Release 4, Version 4.3.0, March 2002.

[GAHL94] Gahleitner R., "Radio Wave Propagation in and into Urban Buildings", Dissertation, TU Wien, 1994.

[GPRS2000] Pang, Q.; Bigloo A.; Leung V.C.M.; Scholefield C.; "Performance Evaluation of Retransmission Mechanisms in GPRS Networks", Wireless Communications and Networking Conference, 2000, WCNC 2000 IEEE, Volume: 3, 23-28 September 2000, Pages:1182 - 1186 vol. 3.

[GPRSLL97] Ludwig, R.; Turina, D., "Link layer analysis of the General Packet Radio Service for GSM", Universal Personal Communications Record, 1997. Conference Record., 1997 IEEE 6th International Conference on, Volume: 2, 12-16 Oct. 1997, Pages:525 - 530 vol. 2.

[GSM03.02] Digital cellular telecommunications system (Phase 2+); "Network architecture", TS 100 522, Version 6.1.0, Release 1997, July 1998.

[GSM04.22] Digital cellular telecommunications system (Phase 2), GSM 04.22, "Radio Link Protocol (RLP) for Data and Telematic Services on the MS/BSS Interface and the BSS/MSC Interface", Version 8.0.0. Release 1999, ETSI, July 1999.

[EGPRS01] Molkdar, D.; Featherstone, W.; Larnbotharan, S.; "An overview of EGPRS: the packet data component of EDGE", Electronics & Communication Engineering Journal , Volume: 14 Issue: 1 , Feb 2002 Page(s): 21 - 38.

[ETSI-TR101.115] Technical Report; Digital cellular telecommunications system (Phase 2+); "Background for Radio Frequency (RF) requirements", (GSM 05.50 version 8.2.0 Release 1999), Version 8.2.0., April 2000.

[ETSI-TS101.627] Digital cellular telecommunications system (Phase 2+); Subscriber and equipment trace; GSM 12.08, version 5.1.1, Release 1996, July 1998.

[FRCODING00] ,Flament, M.; Unbehaun, M.; "Frequency reuse and coding for GPRS multi-slot operation", International Zurich Seminar on Broadband Communications, 2000. Proceedings , 2000, Page(s): 127 - 131.

[I-TCP95] A. Bakre, B.R. Badrinath, "I-TCP: Indirect TCP for Mobile Hosts," Proc. 15th International Conference on Distributed Computing Systems (ICDCS), May 1995.

[ITU-T-V42BIS] ITU-T, Recommendation V.42 bis: "Data compression procedures for data circuit- terminating equipment (DCE) using error correcting procedures", ITU 1990.

[LADPT00] , Gutierrez, P.J.A.; Wigard, J.; Andersen, P.N.; Damgaard, H.C.; Mogensen, P., "Performance of Link Adaption in GPRS", Vehicular Technology Conference, 2000 , IEEE VTS-Fall VTC 2000. 52nd, Volume: 2, 24-28 Sept. 2000, Pages:492 - 499 vol. 2.

[OEZC01] , H. Özcelik , "Dynamische Zuteilung von Funkressourcen", Diplomarbeit, Technische Universität Wien, May 2001

[PILC-LINK] , Internet Draft, Performance Implications of Link Characteristics, Phil Karn, "Advice for Internet Subnetwork Designers", File: draft-ietf-pilc-link-design-12.txt, July, 2002 (Note: draft expired).

[RFC793] Request for Comments: 793, Van Jacobson, "Transmission Control Protocol", DARPA Internet Program, Protocol Specification, September 1981.

[RFC959] Request for Comments: 959, J. Postel; J.K. Reynolds, "File Transfer Protocol", Network Working Group, October 1985, (Obsoletes RFC0765) (Updated by RFC2228, RFC2640, RFC2773) (Also STD0009) (Status: STANDARD).

[RFC1122] Request for Comments: 1122, R. Braden Editor, "Requirements for Internet Hosts – Communication Layers", Internet Engineering Task Force, Network Working Group, October 1989.

[RFC1144] Request for Comments: 1144, Van Jacobson, "Compressing TCP/IP Headers for Low-Speed Serial Links", Internet Engineering Task Force, Network Working Group, Febrary 1999.

[RFC1945] Request for Comments: 1945, T. Berners-Lee, R. Fielding, H. Frystyk, "Hypertext Transfer Protocol - HTTP/1.0" , Internet Engineering Task Force, Network Working Group, May 1996 (Status: INFORMATIONAL).

[RFC2018] Request for Comments: 2018, M. Mathis; J. Mahdavi; S. Floyd; A. Romanow, "TCP Selective Acknowledgment Options", Category: Standards Track, Internet Engineering Task Force, Network Working Group, October 1996.

[RFC2068] Request for Comments: 2068, R. Fielding, J. Gettys, J. Mogul, H. Frystyk, T. Berners-Lee, "Hypertext Transfer Protocol - HTTP/1.1", Internet Engineering Task Force, Network Working Group, January 1997, (Obsoleted by RFC2616) (Status: PROPOSED STANDARD).

[RFC2411] Request for Comments: 2411, R. Thayer, N. Doraswamy, R. Glenn, "IP Security, Document Roadmap", Network Working Group, November 1998, (Category: Informational)

[RFC2460] Request for Comments: 2460, S. Deering, R.Hinden, "Internet Protocol, Version 6 (IPv6) Specification", (Obsoletes RFC1883) (Status: DRAFT STANDARD), December 1998.

[RFC2507] Request for Comments: 2507, M. Degermark, B. Nordgren, S. Pink, "IP Header Compression", Network Working Group, February 1999, Category: Standards Track.

[RFC2581] Request for Comments: 2581, M. Allman; V. Paxson; W. Stevens, "TCP Congestion Control", Internet Engineering Task Force, Network Working Group, (Category: Standards Track), (Obsoletes: 2001), April 1999.

[RFC2582] Request for Comments: 2582, Floyd, S. and T. Henderson, "The New Reno Modification to TCP's Fast Recovery Algorithm", Internet Engineering Task Force, Network Working Group, April 1999.

[RFC2616] Request for Comments: 2616, R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, "Hypertext Transfer Protocol - HTTP/1.1", Internet Engineering Task Force, Network Working Group, June 1999, (Obsoletes RFC2068) (Updated by RFC2817) (Status: DRAFT STANDARD).

[RFC2757] Request for Comments: 2557, G. Montenegro; S. Dawkins; M. Kojo; V. Magret; N. Vaidya, "Long Thin Networks", January 2000, (Status: INFORMATIONAL).

[RFC3135] Request for Comments: 3135, J. Border, M. Kojo, J. Griner, G. Montenegro, Z. Shelby, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations", Internet Engineering Task Force, Network Working Group, June 2001, (Status: INFORMATIONAL).

[RFC3168] Request for Comments: 3168, K. Ramakrishnan, S. Floyd, D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", Internet Engineering Task Force, Network Working Group, (Updates: 2474, 2401, 793, Obsoletes: 2481) (Category: Standards Track), September 2001.

[RFC3366] Request for Comments: 3366, G. Fairhurst, L. Wood, "Advice to link designers on link Automatic Repeat reQuest (ARQ)", Internet Engineering Task Force, Network Working Group, BCP62, (Category: Best Current Practice), August 2002.

[SNOOP95] H. Balakrishnan, S. Seshan, E. Amir, R. H. Katz; "Improving TCP/IP Performance over Wireless Networks", Computer Science Division, University of California at Berkeley; in Proc. 1st ACM Intl Conf. on Mobile Computing and Networking (Mobicom), November 95.

[STEINB01] Martin Steinbauer; "The Radio Propagation Channel - A Non-Directional, Directional, and Double-Directional Point-of-View", Technische Universität Wien, Dissertation, November 2001.

[TABO2002] Manfred Taferner, Ernst Bonek; "Wireless Internet Access over GSM and UMTS", Springer-Verlag, 2002, ISBN: 3-540-42551-9.

[Tanenbaum88] Andrew.S., "Computer Networks", Prentice Hall, 1988, ISBN: 0-13-162959-x.

[TCPLLC] Luying Zhou; Pat S.Y. Chan; R. Radhakrishna Pillai, "Effect of TCP/LLC protocol interaction in GPRS networks", Kent Ridge Digital Labs, 21 Heng Mui Keng Terrace, Singapore 119613, http://www.sciencedirect.com

[VERBO] Veres, A.; Boda, M., "The chaotic nature of TCP congestion control", INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE , Volume: 3 , 2000.

Acknowledgement

Firstly, I would like to thank Motorola Austria, especially Paul Nowak, and Prof. Dr. Ernst Bonek, head of the Mobile Communications Group from the Institute of Communications and Radio-Frequency Engineering of Vienna University of Technology, for making this diploma thesis possible.
I also express my appreciation to Prof. Bonek for the positive feedback on the content of the manuscript, the technical comments to the topic and for encouraging me to finish this diploma thesis. I also appreciated the valuable inputs to improve my presentation skills and my English. I always enjoyed joining mobile communications group meetings and get updates on research activities. Special thanks to Elmar Trojer for reviewing the manuscript.

My parents deserve my deepest gratitude and respect for supporting me with all their efforts and never losing faith in the way I finish my study. I owe you so much, thank you!

Thank you Gabriele, for your passion and your patience.

To Gabriele
And to my children,
Christina and my unborn son
With Love

last page intentionally left blank