Abstract

Enabling wireless Internet access is one of the upcoming challenges for mobile radio network operators. The General Packet Radio Service is the packet-switched extension of GSM and was developed to facilitate access to IP-based services compared to existing circuit-switched services provided by GSM. Besides an overview on the basic concept, network architecture, and protocols of GPRS, this article discusses the performance the end user perceives when retrieving information from the Web using this access technology. This discussion is based on results obtained from a protocol simulator comprising not only the radio interface protocols, but also the relevant Internet protocols as well as a characteristic application model. The results show that GPRS provides bandwidth-efficient support for bursty applications like Web access.

Wireless Internet Access Based on GPRS

ROGER KALDEN, INGO MEIRICK, AND MICHAEL MEYER,
ERICSSON RESEARCH,
ERICSSON EUROLAB DEUTSCHLAND

In recent years, Internet technology has emerged as the major driving force behind new developments in the area of telecommunications networks. The volume of packet data traffic has increased at extreme rates. In order to meet these changing traffic patterns, more and more network operators adapt their strategies and plan to migrate to IP-based backbone networks. Clearly, the Internet will dominate our daily life in the future much more than today.

Meanwhile, mobile networks face a similar trend of exponential traffic increase and growing importance to users. In some countries, such as Finland, the number of mobile subscriptions has recently exceeded the number of fixed lines. This tremendous success was not expected in the 1980s, when today's second-generation mobile communication systems were designed.

The combination of both developments, the growth of the Internet and the success of mobile networks, suggests that the next trend will be an increasing demand for mobile access to Internet applications. It is therefore increasingly important that mobile radio networks support these applications in an efficient manner. Thus, mobile radio systems currently under development include support for packet data services. The most widely deployed standard for second-generation mobile radio networks is the Global System for Mobile Communications (GSM) [1]. Networks based on this standard will be extended in the near future with the General Packet Radio Service (GPRS), which provides data rates up to 160 kb/s.

When discussions about GPRS started in the early 1990s, applications such as road transport telematics and financial services were driving the demand. The high costs for circuit-switched GSM connections prevented the widespread use of mobile data transmission for such services. In recent years, however, end-user applications such as Web browsing and e-mail are becoming increasingly popular; therefore, the Internet has dominated the standardization of GPRS. Internet applications are predicted to contribute the largest share of the expected traffic volume.

In brief, GPRS can be described as a service providing optimized access to the Internet, while reusing to a large degree existing GSM infrastructure. Advanced mobile terminals using multiple slots will offer more convenient and faster Internet access than today's technology. The GPRS concept allows volume-oriented charging, which permits users to have cheap, permanent connections to the Internet.

This article concentrates on two objectives. First, it provides an overview of GPRS, including technology, architecture, and applicability; second, it presents some simulation results that exhibit the GPRS performance. In particular, the article analyzes Web browsing, presumably the most important application of GPRS.

The following section describes in detail the GPRS architecture and protocols. We then introduce the simulator and its capabilities, traffic models, and parameters. Next, the simulation results are presented and discussed; and finally, conclusions are drawn.

GPRS Architecture and Protocols

Overview

The increased demand for data transmission has also affected mobile networks. As a consequence, two enhancements for GSM were standardized: High Speed Circuit-Switched Data (HSCSD) [2] and GPRS [3]. The primary goal of both services is to provide a bearer service with higher data rates.

HSCSD and GPRS use new coding schemes and have the capability to allocate more than one time slot to one user. In contrast to HSCSD, which is a circuit-switched service (as the 9.6 kb/s data service in GSM), GPRS is a packet-switched service. This provides a lot of advantages, as exhibited later, but also requires changes to the existing GSM networks worldwide.

Packet Switching in GSM

The basic idea of GPRS is to provide a packet-switched bearer in a GSM network. As impressively demonstrated by the Internet, packet-switched networks make more efficient use of the resources for bursty data applications and provide more flexibility in general.

The packet-switched principle is used throughout the GPRS network. The GPRS backbone connecting the dedicated GPRS nodes in the public land mobile network (PLMN) is based on the Internet Protocol (IP). On the air interface the resources are assigned to mobile stations only temporarily on a per-packet basis. In contrast to circuit-switched GSM bearers, where time slots are assigned to one user for the entire duration of a call, in GPRS the radio resources are only assigned for the duration of one or a few IP packets.

Introducing GPRS will enable the following:

 1070-9916/00/$10.00

Parameter	Values				
Precedence	High, normal, low				
Reliability	Packet loss probability: e.g., 10^{-9}, 10^{-4}, 10^{-2}				
Delay for packets of 128 octets	Class	1	2	3	4
	Mean (s)	<0.5	<5	<50	Best effort
	95% (s)	<1.5	<25	<250	Best effort
Maximum Bit rate	8 kb/s–2 Mb/s[1]				
Mean Bit rate	0.22 b/s–111 kb/s				
[1] Current GPRS limit 160 kb/s					

Table 1. *The QoS profile.*

- Circuit- and packet-switched services in one mobile radio network
- Efficient use of the scarce radio resources
- Fast setup/access times
- Efficient transport of packets in the GSM network
- Connectivity to other external packet data networks, based on IP or X.25
- User differentiation based on quality of service (QoS) agreements
- Volume-based charging

Applications

GPRS can act as a mobile access network to the Internet. Due to its efficient support of bursty traffic, GPRS is anticipated to be used for Web browsing, e-mail, traffic telemetric systems, points of sale, and various other vertical applications. In addition, the Short Message Service (SMS) of GSM will be supported with GPRS for improved flexibility and capacity. To enable a wide range of applications, GPRS will provide point-to-point and point-to-multipoint connections.

In order to use the scarce radio resources more efficiently and to support a number of applications with different requirements, GPRS provides several QoS profiles, allowing operators to create schemes for charging differentiation.

The QoS profiles for GPRS phase 1 are characterized by five different parameters as listed in Table 1 [3]. In GPRS phase 2 the ambition is to align the GPRS QoS profiles with the QoS profiles in third-generation wireless networks. Since the QoS profile can be seen as the logical interface between the GPRS system and the application, this alignment would enable roaming between GPRS and Universal Mobile Telecommunications Service (UMTS) networks transparently. The standardization of QoS in GPRS phase 2 is currently ongoing.

Integration in GSM

As mentioned before, GPRS can be seen as an extension to GSM. Therefore, GPRS is embedded in the physical channel — frequency-/time-division multiple access (FDMA/TDMA) — structure of GSM, but employs dedicated protocols, which will be explained below.

GPRS can be implemented in existing GSM systems using the same cell structure. Depending on the coding scheme, not even new frequency planning is necessary. As a consequence only minor changes will be required to introduce GPRS in an existing GSM network.

GPRS Network Architecture

In order to introduce GPRS, modifications of the GSM network are required. Some of the nodes already implemented in current GSM systems can be shared between GPRS and GSM. Only two new node types, serving GPRS support node (SGSN) and gateway GPRS support node (GGSN), have to be introduced. In addition, this new technology requires the development of new mobile terminals.

Figure 1 depicts all nodes and interfaces defined for GPRS. The GGSN is the gateway node between an external packet data network (IP) or packet-switched data network (X.25/X.75) and the GPRS core network. In the case of an external IP network, the GGSN is seen as an ordinary IP router serving all IP addresses of the mobile stations (MSs). This node may include firewall and packet-filtering mechanisms. Additionally, its task is to assign the correct SGSN for a mobile station depending on the location of the MS.

The SGSN interfaces between the GPRS backbone and the radio access network, and switches the packets to the correct base station subsystem (BSS). Its tasks include ciphering and authentication, session management, mobility management, and logical link management to the MS. It also provides a connection to the databases, such as the home location register (HLR), in the mobile switching center (MSC).

The BSS[1] consists of two nodes. First, the base station controller (BSC),[2] including the packet control unit (PCU), sup-

[1] *Only GPRS-relevant functionality is described.*

[2] *Logically, the PCU belongs to the BSC, but physically it can be positioned at the SGSN, BSC, or BTS.*

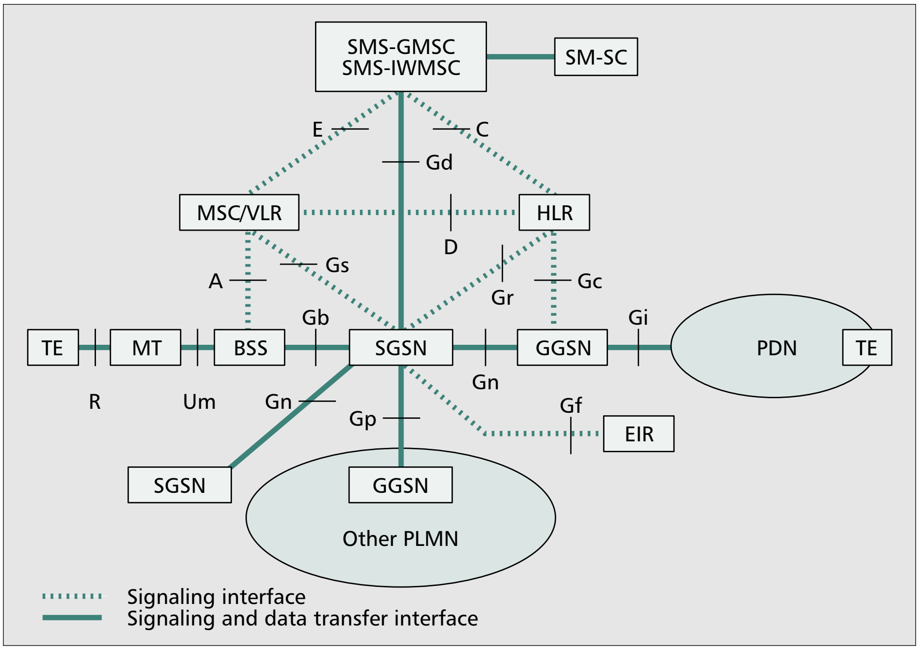


Figure 1. *The GPRS logical node architecture [4].*

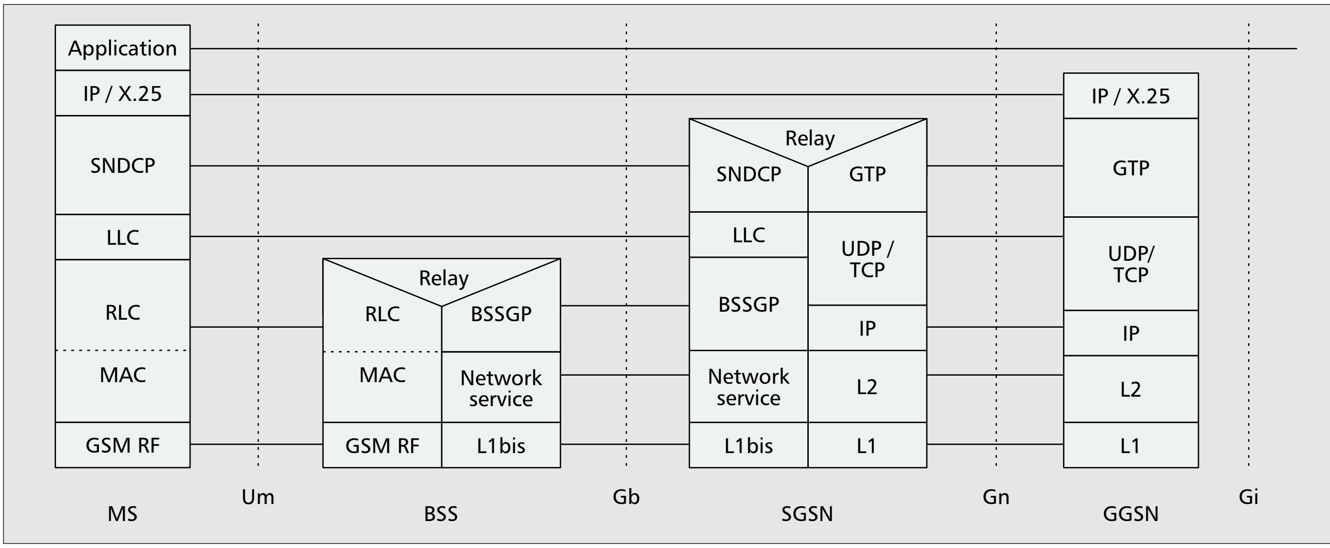


Figure 2. *The GPRS transmission plane [4].*

ports all relevant GPRS protocols for communication over the air interface. The PCU's function is to set up, supervise, and disconnect packet-switched calls, including support for cell change, radio resource configuration, and channel assignment. Second, the base transceiver station (BTS) is only a relay station without protocol functions. It performs the modulation of the carrier frequencies and demodulation of the signals.

The MSC/visitor location register (VLR), HLR, and SMS-Center are functional entities of the initial circuit-switched GSM. These nodes are enhanced by additional interfaces for interworking with GPRS. The MS is equipped with the GPRS protocol stack and is the means of connecting the user to the GPRS network. The GPRS standard foresees MSs that can connect to either circuit- or packet-switched services, or to both services simultaneously [3].

Protocols

On the network level, GPRS supports IP and X.25 protocols to be used by an end-to-end application. According to Fig. 2, IP or X.25 packets are forwarded through the GPRS PLMN network using dedicated protocols. Although the GPRS network consists of several different nodes, it represents only one IP hop.

A peculiarity of GPRS is that, independent of the packets transported, IP is used as the network layer protocol for the GPRS backbone (e.g., to connect SGSN and GGSN). The GPRS Tunnel Protocol (GTP) enables tunneling multiprotocol data packets through the GPRS backbone between GPRS support nodes [5]. The GTP utilizes either TCP or UDP depending on whether a reliable connection is needed (e.g., for X.25 packets) or not (e.g., for IP packets).

The main task of the Subnetwork-Dependent Convergence Protocol (SNDCP) is to carry network layer protocol data units (IP/X.25) in a transparent way [6]. The introduction of new network layer protocols does not require changing all GPRS protocol layers; only SNDCP will be affected. Furthermore, SNDCP provides data compression (e.g., V.42 bis) and header compression (e.g., TCP/IP header compression) in order to improve channel efficiency.

The logical link control (LLC) protocol operates across the Gb and the Um interface, providing a logical link between the MS and its SGSN [7]. Typical LLC functions comprise ciphering, flow control, and sequence control. In addition, if the LLC protocol is used in acknowledged mode, it provides detection and recovery of transmission errors; in unacknowledged mode it signals unrecoverable errors. LLC is used by SNDCP for the transfer of network layer packet data units (PDUs), by the SMS protocol to transfer SMS messages, and by GPRS mobility management to transfer control data.

The radio link control/medium access control (RLC/MAC) protocol layer located in the PCU provides services for the transfer of LLC PDUs using a shared medium between multiple MSs and the network [8]. The functions of the RLC protocol include segmentation and reassembling of LLC PDUs. It can be operated in either acknowledged or unacknowledged mode in accordance with the requested QoS. In acknowledged mode, checksum-based detection of erroneous RLC PDUs and retransmission of them is deployed.

The MAC protocol realizes the different logical channels needed to share the common transmission medium by several MSs [9]. It allows one MS to use several physical channels (time slots) in parallel, but also the multiplexing of several MSs over one physical channel.

The physical link layer provides a number of physical channels to the RLC/MAC layer. Its functionality includes forward error correction, interleaving, monitoring of radio link signal quality, and power control procedures.

The lowest layer on the Um interface, the physical radio frequency (RF) layer, performs transmission and reception of modulated waveforms on the carrier frequencies and is identical to the traditional GSM RF layer.

Figure 3 depicts the segmentation corresponding to the different protocol layers. The figure shows how an application PDU, transmitted across the GPRS air interface, is segmented and encapsulated in several subprotocol PDUs resulting in considerable header overhead. Additional overhead results from signaling at each protocol layer. In total approximately 20–30 percent of the GPRS air interface capacity is spent on protocol overhead and is not available for payload transmission. Simulation results presented later support this calculation.

Air Interface

The Um interface is considered one of the central aspects of GPRS, because it mainly determines the performance of GPRS. To provide more details about data transmission in GPRS, the air interface is explained in more detail in the following paragraphs.

Logical Channels

GPRS uses the same TDMA/FDMA structure as GSM to form physical channels [10]. For the uplink and downlink direction many frequency channels with a bandwidth of 200 kHz are defined through FDMA. They are further subdivided into TDMA frames with a length of 4.615 ms. Each TDMA frame is further split up into eight TSs of equal size.

As an extension to GSM, GPRS uses the same frequency bands as GSM and its derivatives (e.g., GSM1800, PCS1900), and both share the same physical channels (i.e., time slots). Each time slot can be assigned to either GPRS, transmitting packet-switched data, or GSM, handling circuit-switched calls. Time slots used by GPRS are called the packet data channel (PDCH).

Scheme	Code rate (convolutional coding)	RLC data block without RLC header (octets)	Resulting LLC throughput rate (kb/s)
CS-1	1/2	20	8
CS-2	2/3	30	12
CS-3	3/4	36	14.4
CS-4	1	50	20

Table 2. *GPRS coding schemes.*

The basic transmission unit of a PDCH is called a *radio block*. To transmit a radio block four time slots (TSs) in four consecutive TDMA frames are utilized. As can be seen in Fig. 4, all bursts of TS 0 (cross-patterned) belong to PDCH 0. A PDCH is structured in multiframes comprising 52 TDMA frames, which corresponds to a duration of 240 ms. Every 13th burst (so-called *idle burst*) is not used to transmit data, leaving 12 radio blocks in one multiframe. Thus, the mean transmission time per radio block is 20 ms. A radio block contains 456 bits, but due to forward error correction fewer payload bits can be transmitted. The structure and also the number of payload bits of a radio block depend on the message type and coding scheme. GPRS foresees four coding schemes based on convolutional coding [11]. As a consequence, the protection level and throughput change as indicated in Table 2.

A radio block starts with the MAC and RLC headers, while the tail always forms a block check sequence (BCS), which is used to detect errors that cannot be corrected by forward error correction.

Depending on the message type transmitted in one radio block, a sequence of radio blocks forms a logical channel (I.e., each PDCH can carry several logical channels). An example is a Packet Data Traffic Channel (PDTCH) transporting user data. Figure 4 shows the structure of a radio block in case of a PDTCH. It consists of a MAC header, RLC header and the actual RLC payload data [8].

Some of the logical channels realized by the MAC protocol are briefly described:

- **Packet random access channel (PRACH), uplink** — Common channel used by the MSs to initiate an uplink transfer.
- **Packet paging channel (PPCH), downlink** — The BSC uses this channel to page MSs prior to downlink data transmission.
- **Packet access grant channel (PAGCH), downlink** — Resource assignments for up- and downlink transfers are sent on this channel.
- **Packet broadcast control channel (PBCCH), downlink** — On the PBCCH GPRS system-specific information is broadcast.
- **Packet data transfer channel (PDTCH), up- and downlink** — Data packets are sent on this channel. An MS can use one or several PDTCHs.
- **Packet associated control channel (PACCH), up- and downlink** — This channel conveys signaling information related to a given MS and the corresponding PDTCHs (e.g., RLC acknowledgments).

Medium Access Control

To support the packet-switched principle of GPRS, the resources of one PDCH are assigned only temporarily to one MS. The BSC controls resources in both the downlink and uplink directions. The recipients of radio blocks sent on the downlink to an MS are identified by the MS address in the header of the MAC block. Since all radio blocks on the downlink originate from the BSC, no concurrent access on one PDCH can occur.

In contrast, uplink PDCHs are shared between several MSs. To avoid access conflicts in this direction, the BSC transmits in each downlink radio block header an USF indicating the MS allowed to transmit on the corresponding uplink PDCH. Although this concept prevents collisions for

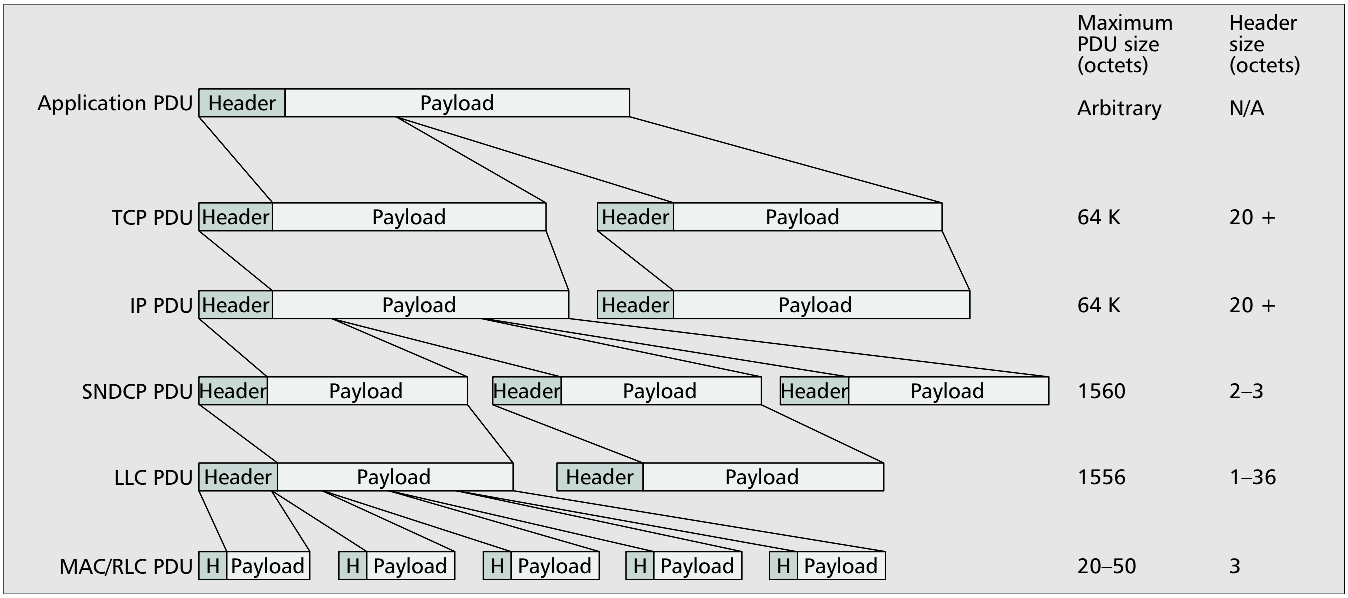
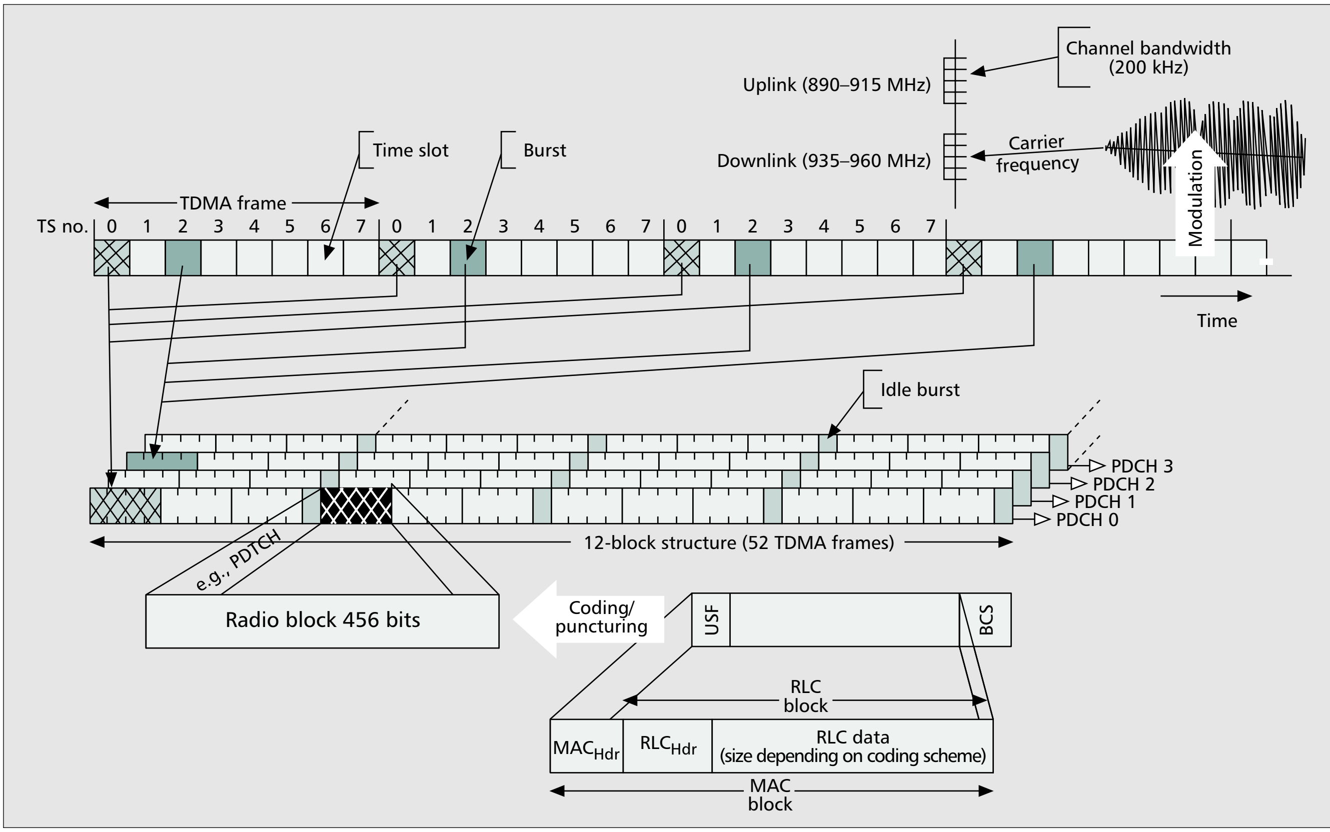


Figure 3. *PDU segmentation.*



■ **Figure 4.** *PDCH and PDU segmentation.*

data transfers, there is still the chance of collision between channel requests of MSs that want to initiate a data transfer. For this problem the PRACH is used to resolve concurrent access on the uplink radio resources. It is a common control channel employing a mechanism similar to slotted Aloha to arbitrate access for the MSs.

Once an MS is successful with its channel request and resources according to this request are available, a temporary block flow (TBF) will be established. When a TBF is established, resources (PDTCH, buffers, etc.) are assigned to an MS, and data transmission can start. As soon as all data for one MS is successfully transmitted, the TBF is released. Using this very flexible packet-oriented resource allocation scheme allows TBF durations ranging from some milliseconds up to several minutes, depending on the amount of data that needs to be transmitted.

One PDCH can carry several PDTCHs for different MSs, resulting in better link utilization. Also, one MS can use several PDTCHs on different PDCHs in order to achieve higher transmission rates. The number of TSs an MS can use is determined by its multislot class.

Figure 5 shows how the resources on the air interface are used. It depicts TBF establishment, including packet channel request and assignment. Additionally, it shows the usage of the USF on the downlink to indicate who is allowed to transmit on the uplink.

Each radio block in Fig. 5 shows the USF (i.e., which MS is allowed to use the corresponding radio block in the uplink direction), the message type (Dt, PUA, PCR, PCA, Ack), and which MS is the owner of the data transmitted (numbers in brackets). As depicted, a radio block might transport data for one MS in the downlink direction and at the same time indicate by means of the USF a different MS that can have access on the uplink direction. This happens in block 2 on TS 1 in downlink direction. A data block (Dt) for MS 3 is transmitted, and MS 1 shall use block 2 on TS 1 in the uplink direction.

Two TSs for the uplink and downlink directions, used as PDCHs by GPRS, are displayed. In this scenario it is assumed that MS 1 has already established a TBF for the uplink direction (TS 0 and TS 1) and transmits data. MS 3 is assumed to have a TBF for the downlink direction only on TS 1. At this point MS 2 also wants to establish a TBF on the uplink.

As can be seen in Fig. 5, the radio blocks in the downlink direction that carry the USF and the uplink radio block for which the USF is granted are shifted by an offset, because the MS needs to receive the USF before transmitting on the uplink.

Block 2 on TS 0 in the downlink direction includes USF = R, indicating that the next radio block belongs to the PRACH. All MSs that want to establish a TBF have to use this logical channel first.[3] In the depicted scenario only MS 2 sends a packet channel request (PCR) on the PRACH. The BSC responds with a packet uplink assignment (PUA) informing the MS which resources (i.e., PDTCHs) have been assigned to accommodate its traffic demands. After reception the TBF is established, and the MS listens to the USF on TS 0 and TS 1 in the downlink direction. According to Fig. 5 three radio blocks are transmitted by MS 2 in two different PDCHs. After these blocks have been transmitted the BSC responds with an Ack message indicating that all data PDUs have been correctly received. A final radio block is assigned to MS 2 in order to transmit the packet control acknowledgment (PCA) for the release of its current TBF.

Radio Link Control

The transmission over the radio link is secured by the RLC protocol. It can be used in acknowledged mode or in unacknowledged mode. The acknowledged mode employs a selective repeat-automatic repeat request (SR-ARQ) mechanism for efficient retransmission of erroneous blocks. The window

[3] *There are exceptions; see [9].*

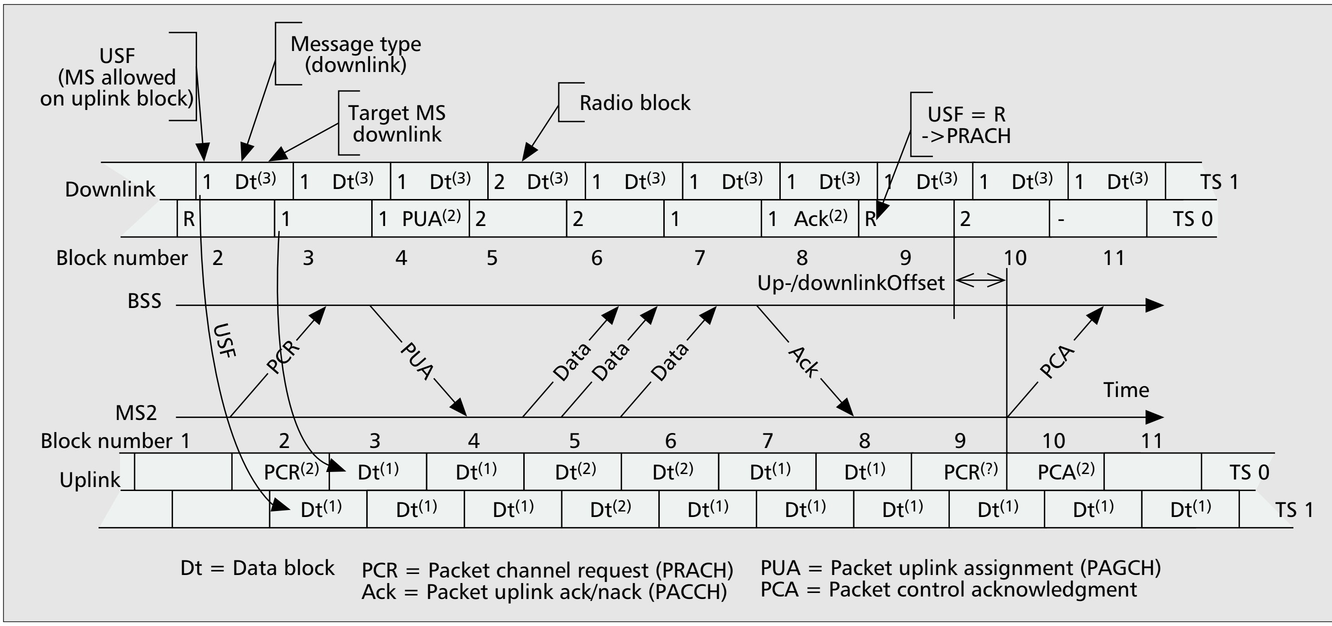


Figure 5. *USF usage and TBF establishment.*

size of the corresponding sliding window is 64 RLC blocks. Frequently sent Ack messages convey a bitmap indicating which blocks need to be retransmitted. In Fig. 5 an error-free transmission is assumed, and hence the final acknowledgment indicates that no retransmissions are necessary and the TBF can be released.

The GPRS Simulator

Since it is sometimes difficult to treat complex telecommunications systems analytically, simulations are appropriate and necessary to obtain an understanding of the system behavior and provide significant insight to performance considerations. The simulator described in this article evolved over a period of three years in parallel with the ongoing standardization process of GPRS. An adequate simulator design requires a clear idea of which items shall be studied in order to apply appropriate models and abstractions. The purpose of developing the simulator was to obtain a tool that allows the following to be investigated:

- Protocol interactions between layered protocols (e.g., RLC and TCP) [12]
- Performance characteristics from a system perspective, such as capacity and throughput
- End-to-end performance from a user perspective with achievable bit rates and observable delays

The simulator should allow verification of standards and analysis of alternative protocols. In addition, it should enable the selection of preferable parameters.

From the beginning it was assumed that system performance is considerably affected by the air interface as well as the Internet infrastructure, which introduces packet delays and losses. The additional delays introduced by the GPRS core network are assumed to be rather small compared to the air interface delays. Therefore, only the transmission delays in the GPRS backbone network are taken into account. In addition, it was required to find some suitable traffic models capable of modeling user behavior and application characteristics. It is obvious that radio conditions and the physical layer design basically determine the behavior of an air interface. However, it is impossible to use detailed physical layer models in intensive protocol simulations because of the heavy computational effort.

All these requirements and assumptions have influenced the simulator design described below. Following the typical methodology for protocol simulations, an event-driven simulator was developed. In order to meet the specified goals, the simulator was built to comprise all nodes relevant to the transmission plane, such as the terminal equipment, MS, BSS, SGSN, GGSN, and finally the Internet host. Since the simulator models only one cell and one carrier frequency, mobility models and handover scenarios are excluded. The complete protocol stack has been implemented, except the protocols responsible for the transmission in GPRS core network, because of their minor influence on the considered study items. This means that TCP, IP, SNCDP, LLC ,and RLC/MAC have been implemented in detail (Fig. 2). Traffic models for typical Internet applications such as Web browsing, e-mail transfer, and file transfer have been developed. The models and protocol implementations employed in the presented simulations are detailed below.

The Web Traffic Model

For the simulations presented in this article only the Web traffic model was used. This traffic model consists of two parts, the client and server models. While the client model determines user behavior, the server model is responsible for the generated amount of traffic. The fundamental statistical parameters were determined by measurements at a dialup fixed-line modem pool of an Internet service provider, since appropriate measurements for wireless terminals are not available yet.

The user behavior is modeled as an on-off model composed of active (requesting data) and idle (reading a page) states for each user. An exponential distribution models the duration of the idle period. The active period is determined by the requested amount of data and the instant system performance (e.g., load, radio conditions) itself. Thus, the system load exercises an influence on the application behavior. In congested situations the active period is extended.

The server model takes into consideration the number of objects (texts, icons, pictures, etc.) on a Web page and the size distribution of these objects. The number of objects is considered to be geometrically distributed.

Based on the measurement results two different object-size distributions were used, one for HTML objects and another

	HTML	Other
Probability of file type	0.51	0.49
Lower bound	100 bytes	100 bytes
Upper bound uniform distribution	4200 bytes	1024 bytes
Alpha	1.36	1.06
Mean file size	12,888 bytes	12,000 bytes
Truncation bound	131072 bytes	
File not found probability	0.03	
File not found response size	30 bytes	
Document cached probability	0.04	
Document cached response size	45 bytes	
Probability for geom. distribution	0.1	
Mean idle time (expo. distribution)	90 s	

Table 3. *Parameters of the Web traffic model.*

for all other object types. In addition, it was derived that the object-size distribution should preferably be modeled by a combined uniform and Pareto distribution. The uniform distribution models the size of small objects with size between a lower and an upper bound, as indicated in Table 3, while the Pareto distribution models the size of larger objects. Finally, there is a parameter to truncate the Pareto distribution in order to model the unlikelihood that users fetch very large objects over relatively slow links. Table 3 summarizes the parameters for the Web model.

The TCP Implementation

Transmission Control Protocol (TCP) [13] is the transport layer protocol used in the Internet for non-real-time applications such as the applications considered in the given simulator. TCP provides an end-to-end connection between two peers, including window-based flow control and retransmission backoff techniques to avoid congestion in the Internet. Additionally, it ensures data reliability by an end-to-end ARQ mechanism. Due to the wide variety of link characteristics in the Internet, TCP operates with adaptive timers based on round-trip time measurements. This ensures that it also adapts to the link conditions provided by GPRS.

All relevant TCP mechanisms (for details see [13]), such as slow start, congestion avoidance, and retransmission based on both timeouts and fast retransmit, have been implemented. The employed values for the maximum segment size and maximum window size of TCP are 536 bytes and 8 kbytes, respectively.

The IP Implementation

The purpose of IP [13] is mainly the routing of packets to the appropriate destination. Since our simulated network topology merely consists of a chain of nodes, no routing functionality is required. Consequently, just the IP protocol header, which enlarges the transmitted packets, has to be considered.

The Internet Model

The Internet transmission delay and packet loss parameters were very important to our simulation studies. It is rather difficult to model these performance parameters very realistically [14]. For our purpose, it is more important to be able to study situations with packet losses and the reaction of TCP to these events. The transmission delay introduced by the Internet backbone increases the round-trip delay and therefore influences transaction performance. The simulator allows us to specify a mean loss rate and, based on the simplified assumption of a Gaussian delay distribution, the mean delay and delay variance of IP packets.

The SNDCP Implementation

SNDCP comprises the functionality needed in order to adapt network layer PDUs to GPRS. If necessary, it performs packet segmentation and reassembly procedures in order to meet the maximum payload size of LLC frames (1520 bytes). In addition, header and data compression algorithms are situated in this protocol layer. The simulator supports TCP/IP header compression, but effects are not analyzed in this article.

The LLC Implementation

The LLC protocol in the simulator supports the acknowledged mode (selective ARQ) as well as the unacknowledged mode, but only the latter is considered here.

The RLC/MAC Implementation

The RLC/MAC protocol layer is the most complex part of the simulator implementation and includes the logical channels PDTCH, PACCH, PAGCH, PRACH, and PBCCH. The PBCCH is considered a bandwidth consumption source without further functionality. The PRACH is implemented with the specified persistence control mechanism and a capture model for simulating concurrent access situations. Procedures such as dynamic TS allocation (including the USF mechanism) and full signaling for establishing a TBF are implemented. Depending on the coding scheme, the RLC layer decomposes LLC frames into a number of RLC/MAC PDUs. RLC can operate in acknowledged and unacknowledged modes as specified in the standards [9]. The simulation results illustrated in the next section are based only on the RLC acknowledged mode.

The Radio Model

As mentioned above, physical layer simulations are computationally too demanding to be combined with already intensive protocol simulations. Therefore, the following approach was applied to model the radio channel. Separate physical layer simulations were performed, including effects such as fast fading, frequency hopping, channel coding, and equalization in order to derive the block error rate for several carrier-to-interference (C/I) ratios and different coding schemes. These results are shown in Fig. 6, and were used as input parameters in our simulator. In order to specify the link quality, a lognor-

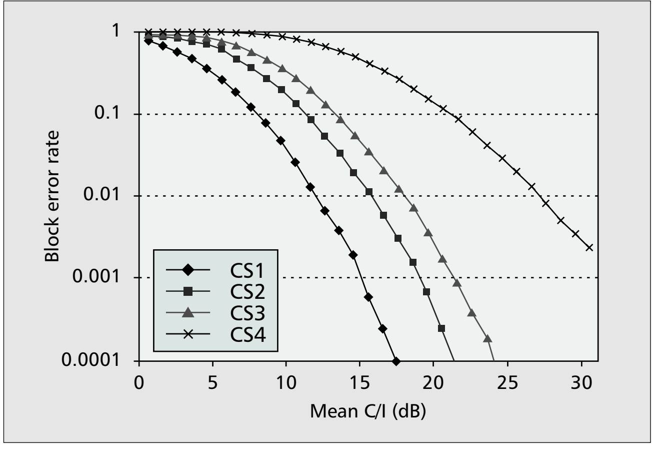


Figure 6. *RLC block error rates versus mean C/I for the GPRS coding schemes.*

mal C/I distribution is considered from which the C/I for a certain user and PDCH is chosen.

Parameters	Load iteration	Channel quality iteration
Number of Web users	10 – 60	50
Mean C/I value	8 + CS • 4 dB	6 – 27 dB
Variance of C/I value	7 dB	7 dB
Coding scheme	1, 2, 3, 4	1, 2, 3, 4
Number of PDCHs	8	8
Multislot capability downlink (MSC)	1, 2, 4	1, 2, 4
Mean Internet packet loss rate	2%	2%
TCP segment size	536 bytes	536 bytes
Simulation duration	45 min	45 min

Table 4. *Main simulationparameters for two different iteration sets.*

Simulation Results

Since there is no GPRS system in large-scale operation at this time, no measurements are available to analyze the system performance. With the implemented simulator, detailed studies were carried out analyzing the performance of Web browsing through GPRS facilities. This analysis can help operators determine what is to be expected from certain allocated resources, and can provide estimates of the service characteristics observed by end users. The first issue is studied by analyzing system performance in terms of throughput under various load conditions for the different coding schemes. Through this analysis system capacity can be evaluated. From a user perspective the most interesting result is the perceived data rate, or how long it takes to download the requested information.

Before describing the simulation results in detail, it should be pointed out that the remainder of this article only considers results measured at the application layer. This is for two reasons: first, the application layer measurements describe what the user really receives; second, the observation of the application layer inherently includes all interactions and overheads caused by the underlying protocol stack. The main simulation parameters are listed in Table 4 and are used in all simulations unless otherwise stated. There are two sets, one for varying load and another for varying channel quality.

The number of PDCHs specifies how many of the eight TSs in a TDMA frame are reserved for GPRS traffic by the network operator on one GSM frequency. The mobile terminals can operate in multislot mode, which allows the use of several TSs of a TDMA frame. On the uplink only one TS is used, whereas in the downlink one, two, and four TSs may be used, which is indicated by the notation MSC1, MSC2, and MSC4, respectively. The higher number of downlink TSs reflects the highly asymmetric traffic pattern generated by Web browsing.

The obtained results for system throughput and packet bit rate are discussed below. The packet bit rate is defined as the size of an object divided by the experienced delay of that object. Due to the asymmetry of Web traffic, we concentrate on downlink performance, since the network load in the uplink direction is not critical.

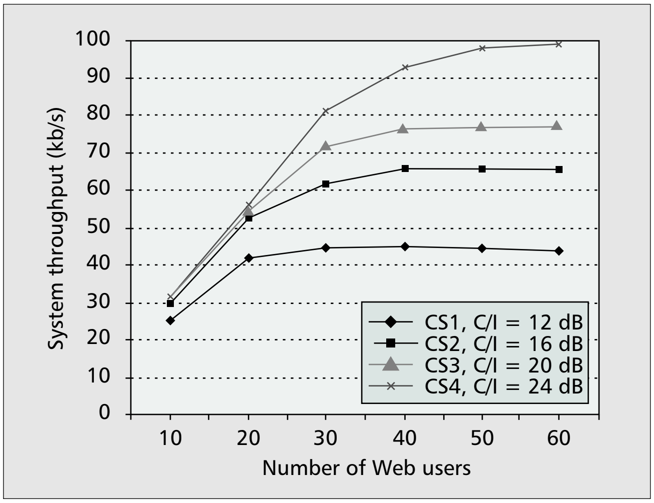


Figure 7. *Aggregated system throughput for varying load.*

System Throughput Performance

Figure 7 illustrates the system throughput at the application layer for four different coding schemes vs. the number of Web users,[4] applying the parameter set for load iteration. The considered multislot capability is four. Note that in order to allow for a fair comparison, the mean C/I value for each coding scheme is different. The values are chosen in such a way that each coding scheme is operated in a specific that provides for low block error rates (Fig. 6) and a good balance between FEC and ARQ.

It is obvious that the system saturates and gets congested as the number of Web sessions increases. The maximum system capacity ranges from 44 kb/s for CS1 to 66 kb/s for CS2, 75 kb/s for CS3, and 100 kb/s for CS4.

Additionally, Fig. 7 allows us to calculate the overhead for all involved protocols of GPRS and for each coding scheme. For example, the link layer data rate per PDCH for CS4 is 20 kb/s, resulting in 160 kb/s for 8 PDCHs, while only a throughput of 100 kb/s is achieved for the application. In total, the overhead caused by protocol headers and signaling sums up to approximately 38 percent.[5]

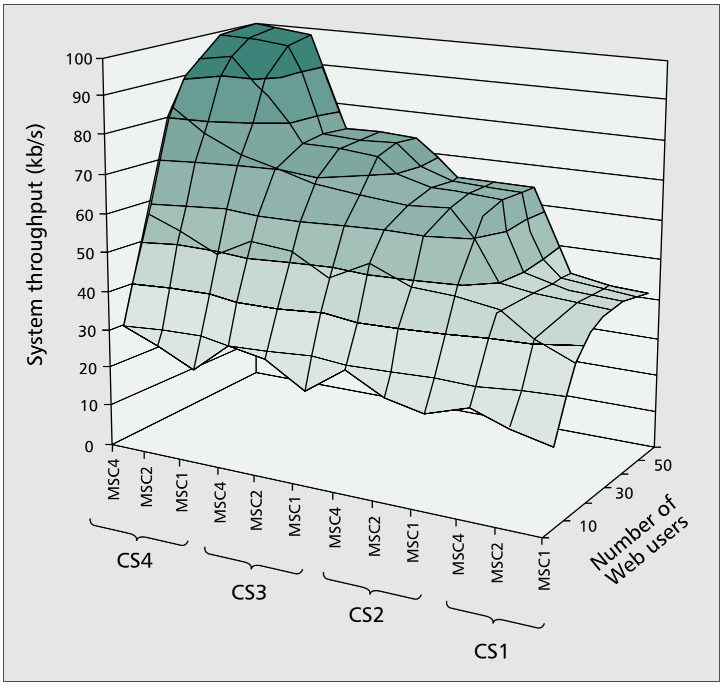
In order to show the dependence of the multislot capability on system throughput, Fig. 8 depicts the system throughput vs. the number of active Web processes for the four coding schemes and three multislot capabilities.

It can be observed that the system saturation level is almost independent of the applied multislot capability and that four horizontal planes (one for each CS) on different levels are built up. For too many active Web applications the 8 PDCHs are fully loaded, and the multislot capability does not affect system throughput. However, in lower load traffic scenarios the throughput at the application level gets slightly higher as the multislot capability is increased. This is due to interactions between the traffic model and system load. When the system is not fully loaded, users benefit more from advanced multislot capabilities. Because they receive their data sooner, they can request new data earlier, resulting overall in increased data volume for a given period of time.

Next, the influence of changing channel qualities is analyzed using the second parameter set from Table 4. The sys-

[4] *Please note that the quantitative results differ from previously published results [15, 16], since more and more details were included in the simulator. Nevertheless, the main conclusions of those articles are still valid.*

[5] *Note that not only overhead contributes to a smaller mean throughput value, but also idle periods in the traffic flow.*



■ **Figure 8.** *Aggregated system throughput for varying load and different combinations of multislot capability and coding scheme.*

tem throughput at the application layer for different coding schemes and MSC4 over a range of mean C/I values is illustrated in Fig. 9. As expected, the results show a strong dependency between throughput and underlying channel conditions.[6] The most robust coding scheme, CS1, has the best performance only for a mean C/I below 6 dB, at which it is unrealistic to operate GSM. CS2 offers superior system throughput between 6 and 10 dB. In the range between 10 and 17 dB CS3 performs optimally, while CS4 is the best for very good channel conditions (i.e., more than 17 dB). CS2 shows very good behavior for low and medium channel conditions. It improves throughput compared to CS1 by 30–50 percent and is only slightly worse (10–15 percent) than CS3, even for C/I values of more than 15 dB. CS4 (the scheme without forward error correction) is very efficient for very good channel conditions. Overall, CS2 is the preferred choice under normal conditions, while CS4 should be used in areas with very good radio quality.

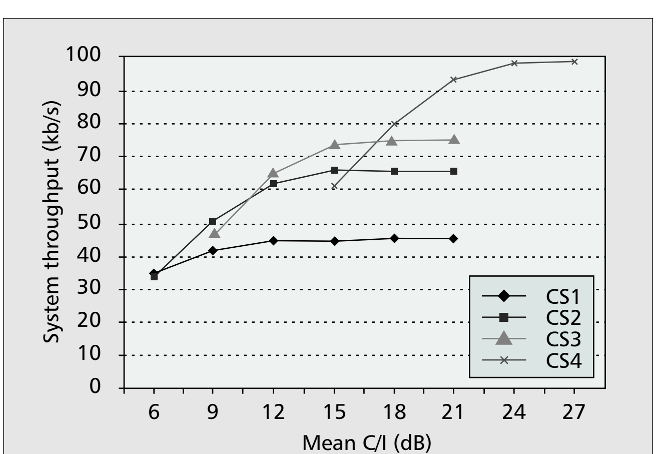
In order to investigate which system throughput can be expected for fewer than eight PDCHs, additional simulations with one, two, and four PDCHs have been carried out. Figure 10 shows system throughput vs. channel quality of one, two, four, and eight PDCHs with CS2 and MSC1. Note the nearly linear increase of throughput vs. the number of PDCHs. This suggests that the results can be extrapolated to larger numbers of PDCHs (i.e., if more than one transceiver is available in a cell).

Packet Bit Rate Performance

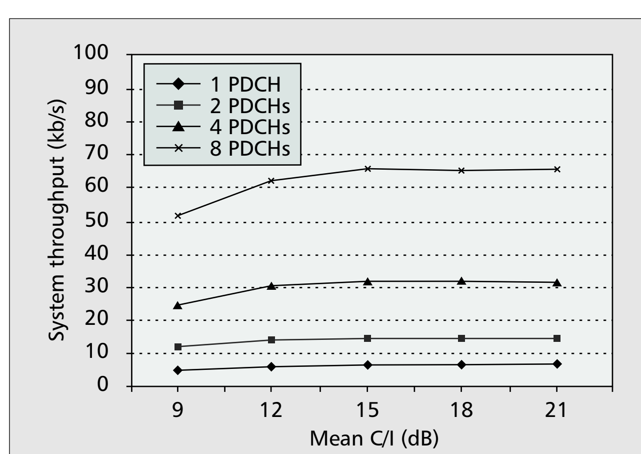
So far, only total system performance has been considered (i.e., the throughput from the operator's point of view). For end users it is more important to identify the throughput they can expect while they are active. Figure 11 gives the mean packet bit rate (object size divided by its transmission delay) measured at the application layer vs. the number of Web processes for different coding schemes and multislot capabilities. The underlying simulation is the same as for Fig. 8 based on the load iteration parameter set of Table 4. On the x-axis the simulation series were grouped according to the multislot capabilities with the four coding schemes. From Fig. 11 we observe that in low load cases, the multislot capability has a high influence on the packet bit rate. This is evident from the different levels of peaks in the surfaces when the number of users is small. On the other hand, for high system loads (i.e., more than 40 Web users) the multislot capability is negligible, and all surfaces have the same flat shape.

To further explore this behavior Fig. 12 presents the mean packet bit rate for coding scheme CS2 and different multislot capabilities. One can observe that in the case of few Web processes the mean packet bit rate is high and the distinction between multislot capabilities significant. For example, in the case of 10 Web users with MSC4 terminals, mean data rates of 22 kb/s can be achieved, whereas with MSC1 terminals data rates of only 8 kb/s can be accomplished. However, as the number of Web users is increased this effect disappears, and every user, independent of multislot capability, perceives the same low mean packet bit rate. Consequently, an important observation is that users can expect an acceptable speed for Web surfing when the network serves less than 30 users. Note that more users can be supported than the number of avail-

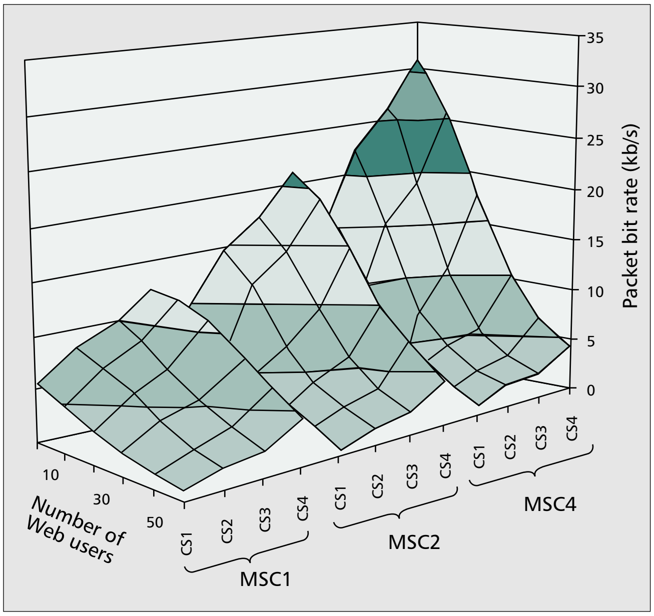
[6] *Note that the system is in all situations fully loaded.*



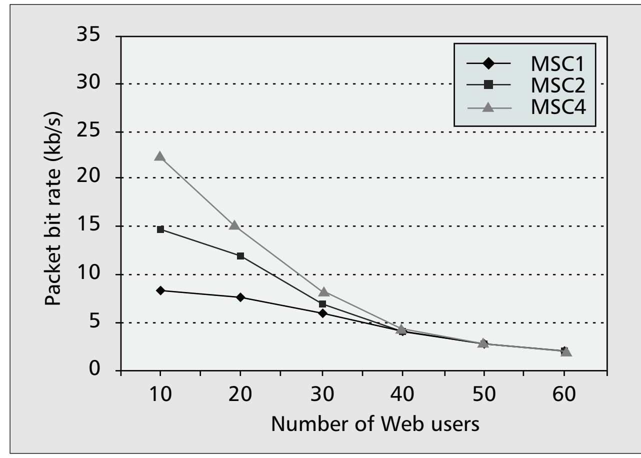
■ **Figure 9.** *Aggregated system throughput for varying channel conditions and different coding schemes.*



■ **Figure 10.** *Aggregated system throughput for varying channel conditions and certain allocated PDCHs.*



■ **Figure 11.** *Packet bit rate for varying load and different combinations of multislot capability and coding scheme.*



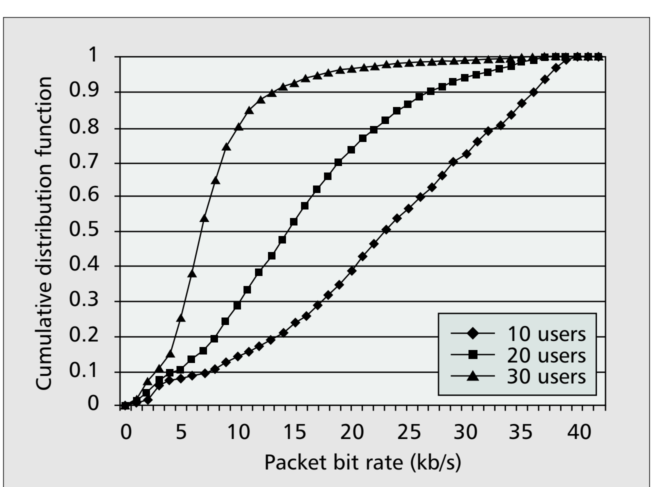
■ **Figure 12.** *Packet bit rate for varying load and different multislot capabilities.*

able PDCHs. This is in contrast to circuit-switched services in GSM, where the number of users cannot exceed the number of available TSs.

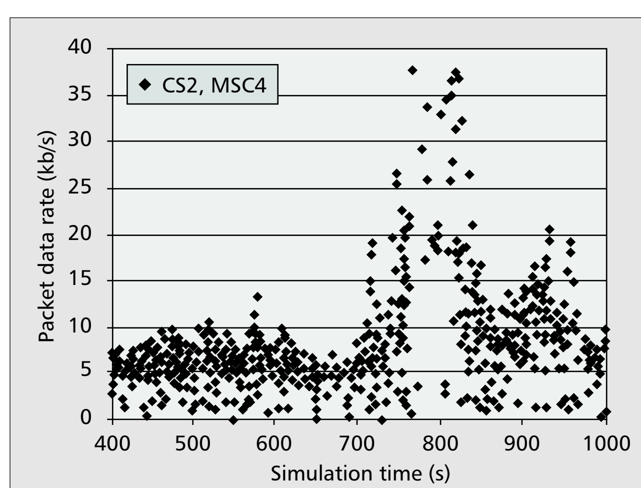
By the aid of Fig. 12 we can also derive the multiplexing gain. In the case of 20 ongoing Web sessions, each user with MSC4 obtains a mean packet bit rate of 15 kb/s. A straightforward computation neglecting the multiplexing gain would thus result in a total throughput of 300 kb/s. In contrast, from Fig. 7 the system capacity was determined to be 66 kb/s for CS2 and MSC4. The difference of 234 kb/s results from using the idle phases of some users for data transmission of others, which is impossible in the circuit-switched case. This example shows that the additional GPRS signaling overhead, which enables multiplexing, results in much more efficient usage of the scarce radio resources.

Figure 11 demonstrates the mean packet bit rate for different configurations, but provides no information about its distribution. Hence, Fig. 13 shows the cumulative distribution functions of the packet bit rate for the cases of 10, 20, and 30 Web users with CS2 and MSC4. As expected, the packet bit rates decrease with offered load. The curves reveal the percentage of transmitted objects that experience a packet bit rate higher than a certain threshold. For example, 90 percent of the Web objects are transmitted with at least 7 kb/s in the case of 10 users. In addition, Fig. 13 shows that the distributions are very similar for low packet bit rates. This is due to two things [16]. First, small objects feature a relatively larger header and protocol overhead (e.g., TCP connection setup), resulting in low packet bit rates on the application level. Second, certain IP packets are lost in the Internet, and time-consuming TCP retransmissions are necessary to recover from these losses. In general, the results indicate that for less than 20 users the large majority of Web objects are received with more than 10 kb/s. Large objects tend to be received with higher packet bit rates, since the relative protocol overhead is less [16].

In order to investigate how packet bit rates vary during a certain time interval, we consider Fig. 14. This figure depicts the packet bit rate for each application object during 10 minutes of simulation time. Thirty Web processes are active using CS2 and MSC4. Note that the packet bit rate varies in a wide range up to a maximum of 38 kb/s. We observe that the majority of objects reach the client with a throughput of more than 5 kb/s. For the simulation period between 400 and 700 s the maximum packet bit rate is only 13 kb/s. During this interval many users need to be multiplexed; hence, a higher data rate is not possible. However, between 700 and 1000 s the overall packet bit rate is higher. This indicates that more users are in idle phase and the active ones can achieve higher data rates.



■ **Figure 13.** *Cumulative distribution function of the packet bit rate for three load scenarios.*



■ **Figure 14.** *Packet bit rate for individual Web objects, 30 users.*

Conclusions

GSM networks will evolve in the next years with packet-switched GPRS. This service was developed for optimized support of Internet traffic, reflecting the growing market demand for data services and wireless Internet access. GPRS technology allows users to stay online for long periods and is designed for applications with bursty traffic characteristics.

This article outlines the fundamental architecture of GPRS and demonstrates, via simulation results, how efficient use of the scarce air interface resources can be achieved. One of the main advantages of GPRS is the dynamic and flexible allocation of radio resources, which results in efficient multiplexing and increased throughput. Those two beneficial characteristics are considered key components in the design and implementation of modern communication systems.

The article also discusses a dedicated simulator which has been developed to analyze the system behavior of GPRS from the application layer perspective. The simulation results presented provide an insight to the capacity of GPRS and to the number of Web users that can be supported with certain allocated resources. Moreover, the influence of different coding schemes under certain channel conditions is examined. Finally, it is shown that end users will experience a wide range of data rates when they use GPRS as the access medium for Web browsing. Assuming identical resources for circuit- and packet-switched services, GPRS can support at least two times the Web users with equal or even higher packet bit rates. In closing, we may argue that the results indicate GPRS is significantly superior to circuit-switched services for applications characterized by bursty traffic patterns such as typical Internet applications.

References

[1] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*, 1992.

[2] "Digital Cellular Telecommunications System (Phase 2+), High Speed Circuit Switched Data, Service Description, Stage 1," GSM 02.34 v. 6.0.0, Jan. 1999.

[3] "Digital Cellular Ttelecommunications System (Phase 2+), General Packet Radio Service, Service Description, Stage 1," GSM 02.60 v. 6.2.1, Aug. 1999.

[4] "Digital Cellular Ttelecommunications System (Phase 2+), General Packet Radio Service (GPRS), Service Description, Stage 2," GSM 03.60 v. 6.3.0, Apr. 1999.

[5] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, GPRS Tunneling Protocol across the Gn and Gp Interface," GSM 09.60 v. 6.2.2 Nov. 1998.

[6] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Subnetwork Dependent Convergence Protocol (SNDCP)," GSM 04.65 v. 6.3.0, Mar. 1999.

[7] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Logical Link Control (LLC) layer specification," GSM 04.64 v. 6.3.0, Mar. 1999.

[8] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Overall Description of the GPRS Radio Interface, Stage 2," GSM 03.64 v. 6.2.0, May 1999.

[9] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Radio Link Control/Medium Access Control (RLC/MAC)," GSM 04.60 v. 6.3.0, Mar. 1999.

[10] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Multiplexing and Multiple Access on the Radio Path," GSM 05.02 v. 6.6.0, Mar. 1999.

[11] "Digital Cellular Telecommunications System (Phase 2+), General Packet Radio Service, Channel Coding, Stage 2," GSM 05.03 v. 6.0.0, Jan. 1998.

[12] M. Meyer, "TCP Performance over GPRS," *IEEE Wireless Commun. and Networking Conf.*, New Orleans, LA, Sept. 1999.

[13] W. R. Stevens, *TCP/IP illustrated*, vol. 1, Addison Wesley, 1994.

[14] V. Paxson and S. Floyd, "Why We Don't Know How to Simulate the Internet," *Proc. 1997 Winter Simulation Conf.*, Dec. 1997.

[15] S. Hoff, M. Meyer, and J. Sachs, "Analysis of the General Packet Radio Service (GPRS) of GSM as Access to the Internet," *ICUPC '98*, Florence, Italy, Oct. 1998.

[16] M. Adams, M. Meyer, and J., Sachs, "Quality of Service Evaluation for Mobile Internet Access via the General Packet Radio Service (GPRS)," *5th Int'l. Wksp. Mobile Multimedia Commun.*, Berlin, Germany, Oct. 1998, pp. 295–300.

Biographies

ROGER KALDEN (Roger.Kalden@eed.ericsson.se) received his diploma in computer science from the Technical University of Aachen, Germany, in March 1998. He joined Ericsson Research in April 1998. His research interests are in the areas of performance analysis of wireless Internet access networks and the provisioning of quality of service in such systems.

INGO MEIRICK (Ingo.Meirick@eed.ericsson.se) received his diploma in electrical engineering from the Technical University of Aachen in November 1998. In December 1998 he joined the GPRS system design group at Ericsson Eurolab Deutschland. He currently focuses on optimizing the GPRS system for mobile Internet access and the provisioning of quality of service.

MICHAEL MEYER (Michael.Meyer@eed.ericsson.se) received his diploma and doctoral degree in electrical engineering from the University of Paderborn in 1991 and 1996, respectively. From 1992 to 1996 he worked as a research assistant with the Communications Engineering Department of the University of Paderborn. In May 1996 he joined Ericsson Research in Aachen. His current research interests include the performance analysis of radio link layer protocols, their interactions with Internet protocols, and appropriate simulation techniques.