Notes on GPRS Performance Issues

White Paper

ascom

Contents

1. Introduction 1

2. GPRS Applications 2
  2.1. Delay vs. Error Sensitivity in Wireless Network Services 2
  2.2. Applications Testable from within TEMS Investigation 3
    2.2.1. Speech 3
    2.2.2. FTP 3
    2.2.3. HTTP 3
    2.2.4. Ping 3
  2.3. Examples of Other Applications 3
    2.3.1. WAP 3
    2.3.2. SMS 4
    2.3.3. Video 4

3. Basics of GPRS Performance 5
  3.1. The Channel Coding Schemes 5
  3.2. Achievable Throughputs in GPRS 6

4. Factors Affecting or Limiting GPRS Performance 7
  4.1. GPRS Network Configuration 7
    4.1.1. Number of Available PDCHs (Timeslots) 7
    4.1.2. Phone Multislot Class 7
    4.1.3. Interference 7
    4.1.4. TBF Reallocations 7
  4.2. GPRS Specifications 7
    4.2.1. CS-1 for RLC/MAC Signaling 8
    4.2.2. TBF Reallocations 8
    4.2.3. Uplink TBF Allocation 8
  4.3. Limitations and Shortcomings of Underlying Technology 8
    4.3.1. TCP Timeouts 8
    4.3.2. TCP Slow Start 8
    4.3.3. Protocol Stalling 9
  4.4. User Application Characteristics 9
  4.5. Other External Factors 9
    4.5.1. Interference 9
    4.5.2. Server Problems 9

5. Other Performance Issues 10
  5.1. Handling QoS Settings and Negotiations 10
  5.2. Traffic Prioritizing 10
  5.3. Charging 11
  5.4. Speech Quality 11
    5.4.1. Dropped Calls, Blocked Calls, Bad Speech Quality 11
    5.4.2. Silent Calls 12
    5.4.3. Echo 12

5.4.4. Paging/SDCCH Dimensioning Problems 12
5.5. Teething Troubles 12
5.5.1. International Roaming 12
5.5.2. APN Settings 12
5.5.3. DNS Settings 13
5.5.4. GPRS Support in Individual Cells 13
5.5.5. SW/HW Design 13

6. GPRS Parameters and Measurements in TEMS Investigation 14

6.1. Mobile Station GPRS Capabilities 14
6.2. GPRS Events 14
6.3. GPRS Parameters 14
6.4. GPRS Measurements 15
6.4.1. Data Transfer Performance 15
6.4.2. Radio Link Quality 15
6.4.3. Details on BLER, Block Error Rate 15
6.4.4. Data on GPRS Connection 16
6.5. GPRS Features to Be Added in Future TEMS Product Versions 16
6.5.1. Features Planned for TEMS Investigation GSM 4.0 16
6.5.2. Other Features 16

1. Introduction

GPRS is a complex technology. It has proven more complex than the cellular industry initially realized, and it is very far indeed from being a straightforward add-on to voice-call GSM. Both infrastructure vendors and operators have underestimated the difficulties of developing the new network infrastructure and telephones required. Even today, in the first half of 2002, no vendor fully supports all the features specified in the standard, for example the packet BCCH (P-BCCH), and support of the CS-3 and CS-4 channel coding schemes is rare at best.

It is not the traditional speech service that gives GPRS operators a headache. Voice calls are hardly more demanding than in GSM, the same range of speech coders being offered. True, the upcoming introduction of the multimode speech coder AMR, which switches between coder modes to adapt to changing radio conditions, will entail extra work and add some complexity; but even so, securing voice performance is child’s play compared to managing packet traffic and the infrastructure required to support it.

Another challenge posed by GPRS is the multiplicity of services to be offered. While GSM was limited mostly to speech and slow, fairly primitive data transfer, GPRS with its much higher data rates enables a variety of further applications such as WAP, web browsing, and streaming. The operator has to simulate and continuously verify all these services – an exacting and time-consuming task.

Of course, the intricacies of setting up a GPRS network are reflected in the TEMS tools, with a wealth of new settable parameters, information elements, and events.

This document gives an overview of

- GPRS applications: chapter 2
- achievable GPRS throughput: chapter 3
- factors directly affecting GPRS throughput: chapter 4
- other aspects of GPRS network performance, including a discussion of common problems: chapter 5
- GPRS data presentable in TEMS Investigation GSM: chapter 6.

2. GPRS Applications

This section takes a look at common applications in GPRS and discusses application characteristics that dictate the choice of transport protocol.

2.1. Delay vs. Error Sensitivity in Wireless Network Services

Services for wireless networks differ in their sensitivity to delays on one hand and transmission errors on the other. In a noisy radio environment one cannot both transmit data in real time and guarantee data integrity: one must decide which is more important.

- *Delay sensitivity:* For audio and video, avoiding delays is imperative in order to preserve the continuity and order of segments in the transmission. Failing to do so will quickly render the transmission unacceptable or wholly unintelligible to the user. On the other hand, sound and pictures contain a great amount of redundancy, so scattered errors, causing disturbances of very short duration, are tolerable to a fairly great extent and may not even be noticed.
- *Error sensitivity:* For most other kinds of data intended to be read by humans or computers, preserving the exact content of the transfer must obviously be the top priority. Simple examples include e-mails, word processing documents, and files containing software code. On the other hand, there are usually no realtime requirements on such transmissions – delays can be accepted.

The Internet uses different transport protocols in the two cases:

- Delay-sensitive applications typically use UDP (User Datagram Protocol). This is a very simple protocol designed for the quickest possible forwarding of data blocks. UDP does not do any handshaking between the sending and receiving entities before the transmission begins, and there is no way to have faulty data blocks retransmitted.
- Error-sensitive applications typically use TCP (Transmission Control Protocol). TCP does much more work than UDP, establishing a connection between sender and receiver in such a fashion as to enable retransmissions when needed. TCP guarantees delivery of data and also guarantees that packets will be delivered in the same order in which they were sent. This comes, of course, at the price of increased transmission delays.

2.2. Applications Testable from within TEMS Investigation

Speech

Speech is realtime communication and is transferred by streaming over UDP.

It is conceivable that some sort of higher-quality speech service, perhaps using stereo, will be launched in the future.

FTP

FTP (File Transfer Protocol) is a common protocol for transferring files in the Internet. FTP assumes that the data is highly error-sensitive and therefore uses TCP to ensure data integrity.

FTP is characterized by continuous operation (uninterrupted stretches of downloading). Hence the average throughput over a period of time is likely similar to typical instantaneous throughput figures during that period. Compare HTTP.

HTTP

HTTP (Hypertext Transfer Protocol) is the protocol used on the World Wide Web. Like FTP, HTTP prioritizes data integrity and uses TCP. Unlike FTP, however, HTTP is characterized by bursty behavior with long stretches of no activity interspersed with short spurts of intense data transfer. The time-average of the throughput will be much lower than the peak throughput rate, reflecting, of course, the typical usage patterns of web browsing.

Ping

Ping (Packet Internet Groper) is a utility used to determine whether a specific IP address is accessible. It is a popular basic diagnostics tool for Internet connections. Ping uses the ICMP protocol which is embedded within the IP layer. It does not use TCP. Since Ping bounces short snatches of data between servers, it naturally has bursty characteristics.

2.3. Examples of Other Applications

WAP

In WAP, unlike the World Wide Web, fast delivery of data takes precedence over integrity. WAP therefore uses UDP rather than TCP. On the other hand, WAP does share the bursty characteristics of ordinary Web browsing.

SMS

SMS uses a reliable protocol that is built into GSM. Like web browsing it is bursty in nature.

Video

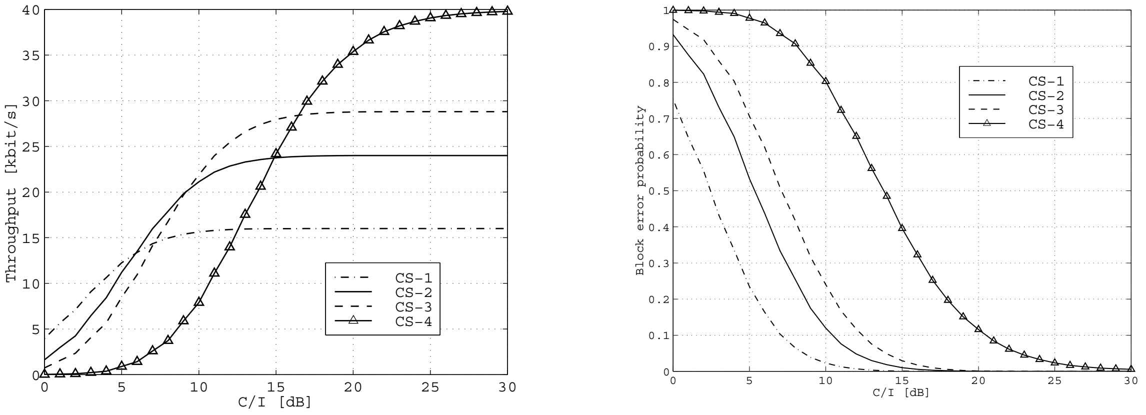
Like speech, video must be sent in real time and is therefore streamed over UDP.

3. Basics of GPRS Performance

3.1. The Channel Coding Schemes

GPRS offers a choice between four channel coding schemes, CS-1 through CS-4, differing in their level of error protection. CS-1 is the most conservative scheme, adding the largest amount of error protection bits; this scheme gives the lowest throughput in good radio conditions but on the other hand fares better as the radio environment deteriorates. CS-4 is the most optimistic scheme and accordingly works well only for high C/I's. (It should be noted that at present, support for CS-3 or CS-4 is not yet widespread.)

Below are two graphs showing the theoretical performance of each of the coding schemes as a function of C/I.



LLC throughput with a two-timeslot mobile (left) and LLC block error probability (right; see section 6.4.3) as a function of C/I for each of the GPRS channel coding schemes.

Assuming a C/I target of 12 dB, we can see that a tight (i.e. well-designed and near-optimal) speech network will be fine running CS-2: the degradation at C/I = 12 dB is very small and reduces throughput by no more than a few percent. In other words, the common practice of using CS-2 throughout for the speech payload is fully justified.

3.2. Achievable Throughputs in GPRS

The following tables list the maximum achievable throughputs at the RLC/MAC and LLC protocol levels.

RLC/MAC throughput in kbit/s (see EPL/N/TB-01:019, section 1.6.4):

Timeslots	**CS-1**	**CS-2**	**CS-3**	**CS-4**
1	9.2	13.2	15.6	21.2
2	18.4	26.4	31.2	42.4
3	27.6	39.6	46.8	63.6
4	36.8	52.8	62.4	84.8

LLC throughput in kbit/s (see EPL/N/TB-01:019, section 4.2):

Timeslots	**CS-1**	**CS-2**	**CS-3**	**CS-4**
1	8.0	12.0	14.4	20.0
2	16.0	24.0	28.8	40.0
3	24.0	36.0	43.2	60.0
4	32.0	48.0	57.6	80.0

4. Factors Affecting or Limiting GPRS Performance

Under this heading are collected a variety of factors that affect performance in a GPRS network. They are something of a mixed bag, and the chapter is intended to be one – though divided into a number of smaller compartments. Naturally the discussion is by no means exhaustive.

4.1. GPRS Network Configuration

These are factors that can be influenced directly by operators and/or infrastructure and phone vendors within the confines of the present specifications.

Number of Available PDCHs (Timeslots)

A PDCH (Packet Data Channel) is one timeslot which is reserved for GPRS traffic. The number of available PDCHs is the most important factor determining the achievable throughput.

Note that a PDCH can be shared among several GPRS users. They are also not permanently reserved for data services, but can be “stolen” to serve speech users when the need arises; cf. section 5.1 about traffic prioritizing.

Phone Multislot Class

Mobile phones differ as to the number of downlink and uplink timeslots they can handle (up to 4 on the downlink at the present time).

Interference

Interference from other GPRS users is naturally inescapable. Coding schemes CS-3 and CS-4, with the least error protection, are most sensitive to interference. CS-2, on the other hand, requires no remarkable carrier-to-interference ratio to give acceptable quality, provided the GPRS network is suitably dimensioned for voice (cf. section 3.1).

TBF Reallocations

See below, section 4.2.

4.2. GPRS Specifications

These are some mechanisms which are prescribed in the present-day GPRS technical specifications and which, while clearly desirable for other reasons, have the effect of limiting the achievable throughput.

CS-1 for RLC/MAC Signaling

RLC/MAC control signaling, such as packet assignments or packet acks/nacks (as opposed to payload data transfer) always uses coding scheme CS-1 to minimize the risk of non-correctable channel errors. Large amounts of such signaling will keep the throughput down.

TBF Reallocations

Every data transmission in GPRS requires setting up a so-called temporary block flow (TBF). The TBF has a predetermined life span, and when it times out a new one must be allocated. If the TBF is too short-lived – in some networks it used to last only a single IP packet –, throughput performance will be severely impacted. (The problem has more recently been remedied; one solution is to use a “Delayed TBF” which is maintained for a longer period than originally intended in the system design.)

Uplink TBF Allocation

Every time a GPRS subscriber downloads something, the downlink data transfer must be acknowledged on the uplink, requiring the setup of an uplink TBF. Like all signaling this takes up network resources, and may in case of heavy downloading have some effect on network throughput.

4.3. Limitations and Shortcomings of Underlying Technology

These issues concern protocols used in GPRS networks.

TCP Timeouts

After passing on data packets, the TCP protocol regularly requests an acknowledgement from the remote host saying that the packet was correctly received. If the acknowledgement does not arrive within a specified time, TCP will time out, causing delay and lowering the throughput rate. This can be a severe problem with slow client PCs; in such cases operators must simply upgrade their computer hardware.

TCP Slow Start

TCP is a protocol designed for stationary wireline networks and not for wireless radio. In a wireline network, when congestion has occurred, TCP will ramp up its transfer rate conservatively in order to avoid clogging the router. Now this behavior is meaningless in the radio environment, where interruptions of the data transmission are typically due to Rayleigh fading dips; as soon as the mobile has gotten out of the dip – which usually happens within a small fraction of a second – there is no reason not to resume the transmission at full blast. Still, TCP being TCP, a break in the transmission will be followed by a characteristic “slow start” which, of course, is suboptimal from a throughput point of view.

Protocol Stalling

The RLC and TCP protocols both make use of retransmissions. Before any GPRS networks were put into operation, many people foresaw serious problems with these two protocols stalling one another, each protocol waiting for one or more responses from the other. In this way the data throughput would be more or less choked.

Fortunately, these misgivings have so far turned out to be unjustified. Apparently the transfer rates are not high enough for such stalling effects to become significant. However, experts still fear considerable problems in WCDMA for similar reasons, because of the much higher data rates to be supported in WCDMA networks.

4.4. User Application Characteristics

The operator must be familiar with the characteristic behavior of user applications in order not to draw false conclusions about network performance. For example, FTP is characterized by a steady data flow and long stretches of nonzero (though perhaps far from constant) throughput. HTTP, in contrast, is much more bursty and has a lower average throughput owing to the generally longer intervals of inactivity. Cf. section 2.2.

4.5. Other External Factors

Interference

Interference was listed in section 4.1 as a parameter dependent on the network configuration. Naturally, however, the employed frequencies may also be disturbed by external sources.

Server Problems

While trivial in this context, problems with servers being down or overloaded (perhaps because they are wrongly dimensioned) must be noted as one of the chief obstacles to achieving good GPRS performance.

5. Other Performance Issues

Apart from the basic requirement of securing decent throughput rates, there are many other aspects of the performance of a GPRS network that must not be neglected. This chapter surveys a few of these.

5.1. Handling QoS Settings and Negotiations

GPRS mobiles are capable of requesting a specified quality of service from the network. Nothing of this kind is found in traditional GSM, nor is it really needed for the small palette of services offered there.

The QoS request is made as part of setting up a PDP context, a procedure roughly equivalent to the call setup in GSM. A “negotiation” follows between mobile and network, leading to establishment of the QoS. (The network naturally has the final say, but the negotiating metaphor is nevertheless an apt one). The negotiation involves a fair number of parameters, a considerable amount of signaling in both directions, and consequently a multitude of test cases for the operator. Making it all work is non-trivial.

5.2. Traffic Prioritizing

GPRS offers a variety of services. With a scarce radio resource, it is necessary and non-trivial to decide what traffic should take priority. What takes precedence, speech or data?

In the past, of course, the voice caller has reigned supreme in cellular networks, competing only with other voice callers for spectrum. A subscriber engaged in a call would be extremely annoyed to be disconnected just because someone else wanted to download, say, a preview of The Lord of the Rings. Indeed, operators universally agree that voice should have the highest priority: “No harm to speech users”.

Aside from the speech vs. data conflict, the order of precedence must also be established between different types of data services. This is a tricky task at present, because GPRS networks and services are still in such an early phase of development that traffic volumes are hard to predict. For speech there is more than 10 years of accumulated GSM experience to draw upon, but about the behavior of data service users very little is as yet known: choice of applications, duration of connections, movement patterns, etc. Observations from Japan, where the I-mode data service has been available for a number of years, suggest that data service users behave very differently from voice users: people talk on the phone while on the move, in the street and elsewhere, but use I-mode predominantly at home.

It should be noted that a speech connection takes up much less bandwidth than typical data services, in particular video and intensely interactive services.

5.3. Charging

Operators are clearly extremely interested in unambiguous and reliable charging mechanisms. Nevertheless, at the time of writing, charging in GPRS seems to be at least in some respects fuzzy and incompletely thought through. Specifically, it is sometimes not clear who pays for a particular data transfer, or indeed whether anybody will be charged at all.

Here is an example:

- A GPRS subscriber, Alice, is running WAP on her mobile. The mobile has a PDP context activated and has been assigned an IP address, say, 10.99.12.1. Alice cannot be pinged from a terminal outside the GPRS intranet, because all 10.x.x.x addresses are internal to the GPRS intranet (assigned by the GGSN), and not accessible outside it.
- Another GPRS subscriber, Bob, is using FTP. His mobile, too, has a PDP context activated and has received an IP address: 10.99.12.2. Like Alice, Bob cannot be pinged from outside the GPRS intranet.
- However, Alice can ping Bob, and vice versa (assuming that their mobiles have the Ping protocol built in). Now the question is: *Who pays for the ping response?*

It is obviously unreasonable that Bob should have to pay; a subscriber should not be charged for a communication that occurs without his knowledge as a forced response to an action by another subscriber. Still, in today’s GPRS networks, it is in fact not certain exactly what will happen. It is even unclear to what extent the user will be charged for signaling in general, i.e. including the signaling required to perform self-initiated actions – for example to establish a PDP context.

The potential charging problems are of course greatly aggravated if GPRS subscribers are not given network-internal but *ordinary* IP addresses, accessible to anyone with an Internet connection. (We mention this because it is known to have occurred in practice.)

5.4. Speech Quality

Although (as already noted) the speech service is not the main obstacle to obtaining a smoothly running GPRS network, this section is included cataloging the chief malfunctions and deficiencies typically encountered.

Dropped Calls, Blocked Calls, Bad Speech Quality

All of these are manifestations of bad traffic channel and control channel dimensioning. The former (allocation of TCHs) determines the capacity of the network; the latter (allocation of BCCHs) is crucial to securing geographical coverage. The interference level is naturally affected by both traffic and signaling.

Silent Calls

A silent call is one that is correctly set up but in which neither party can hear the other. This is generally due to a PCM link to an MSC or other higher-level network node being incorrectly established.

Echo

Echo is generally an intermittent problem in modern telephony networks: either objectionable echoes are almost non-existent, or they are a considerable nuisance. The reason for this is twofold. On one hand, today’s echo cancellers are very good and usually suppress echoes below the threshold of audibility. On the other hand, cellular networks are often designed in such a way that defined groups of network elements share a fixed pool of echo cancellers. Therefore, in exceptional circumstances where all the echo cancellers are busy, some connections will be left without any echo cancelling at all, thus being heavily affected by echo.

Paging/SDCCH Dimensioning Problems

The network capacity is also dependent on adequate dimensioning of paging channels (PCHs) and other signaling channels (such as SDCCHs used for call setup).

5.5. Teething Troubles

Under this heading are listed some imperfections typical of newly started-up GPRS networks, as well as a couple of more generally unresolved issues. These problems may be expected to fade away gradually as GPRS networks mature. The charging issues brought up in section 5.3 certainly also belong here, but they were put in a separate section to emphasize their importance.

International Roaming

Roaming for voice calls works excellently in present-day GSM networks, across Europe and beyond. In contrast, roaming for GPRS services is in its infancy at best. The progress in developing roaming interfaces also differs widely between operators and countries.

APN Settings

This is one example of a technicality on which operators take different tacks. The APN (Access Point Name) is the address to the GGSN. Some operators require the subscriber to know and state the APN in order to log onto the GPRS network; others do not. In other words, with some operators the subscriber must create a special data account (and thus remember yet another user name and password), whereas other operators do not impose this burden on the subscriber.

DNS Settings

Certain operators require the subscriber to know and explicitly indicate the IP address of the DNS server to be able to set up a GPRS connection. Other operators send this IP address to the subscribers so that they do not need to bother about it.

GPRS Support in Individual Cells

A silly but common problem in the start-up phase is that of operators simply forgetting to turn GPRS on in certain cells, for example after restarting network nodes.

SW/HW Design

Teething troubles understandably also include software and hardware design problems – in mobile stations, in GPRS nodes, and in interfaces between nodes. These may result in, for example, SGSN and GGSN malfunctions as well as PCU restarts.

6. GPRS Parameters and Measurements in TEMS Investigation

What follows is an overview of the GPRS-related data that is currently presentable in TEMS Investigation GSM (version 3.2).

6.1. Mobile Station GPRS Capabilities

- Multislot class: Number of timeslots available on downlink and uplink
- Switch-Measure and Switch-Measure-Switch times: Time required for switching to a new radio channel, performing a neighbor cell power measurement, and (in the latter case) switch channels once again
- GPRS class: Capability of running GPRS in parallel with other GSM services
- GPRS encryption
- SMS capability in GPRS mode

6.2. GPRS Events

- GPRS Attach, GPRS Attach Failure
- GPRS Authentication Failure
- GPRS Detach
- GPRS PDP Context Activation, GPRS PDP Context Activation Failure
- GPRS PDP Deactivation
- GPRS Routing Area Update, GPRS Routing Area Update Failure

6.3. GPRS Parameters

- PDP context settings
- GMM state
- GRR state
- RAC, Routing Area Code
- Assigned timeslots
- Channel type
- TFI
- TLLI
- USF

6.4. GPRS Measurements

For further details on application level measurements, see the document “Application-level Data Service Measurements in TEMS Investigation GSM”, EPL-02:000694 Uen.

6.4.1. Data Transfer Performance

- Throughput: User application level; LLC level, RLC/MAC level; uplink and downlink
- Mean throughput for current session at user application level
- Bytes received at application level: since dialup; during current session
- Bytes sent at application level: since dialup; during current session

6.4.2. Radio Link Quality

- BLER, block error rate (percentage of blocks resent on uplink; percentage of blocks erroneously decoded on downlink) at PDCH, LLC, and RLC/MAC levels: see section 6.4.3 below
- BER by timeslot (PDCH)
- C/I by timeslot (PDCH)
- C31, C32
- Neighbor C31, Neighbor C32

6.4.3. Details on BLER, Block Error Rate

Throughput and block error rate are related by the equation

$$\text{Throughput} = \text{Throughput}_{\text{max}} \cdot (1 - \text{BLER})$$

The block error rate is measured in TEMS Investigation GSM at several levels:

- PDCH BLER on downlink, information element **FER/Timeslot (%)**. This is the block error rate for one timeslot allocated for GPRS. Measured in packet transfer mode on all allocated PDCHs (i.e. timeslots). All blocks to all mobiles sharing each of the PDCHs are counted, not only the blocks to the TEMS mobile. No uplink PDCH BLER is available; only the downlink is monitored.
- RLC BLER on downlink, information element **RLC Decode Errors DL (%)**. This is the downlink block error rate as perceived by the TEMS mobile. Measured in packet transfer mode, counting only radio blocks addressed to the TEMS mobile. All blocks on each assigned PDCH are counted, and the overall BLER for the active PDCH set is calculated.
- RLC BLER on uplink, information element **RLC Retransmissions UL (%)**. This is the uplink block error rate as perceived by the TEMS mobile. Measured in packet transfer mode, counting only blocks transmitted by the TEMS mobile. All blocks

on each assigned PDCH are counted, and the overall BLER for the active PDCH set is calculated.

- LLC BLER on downlink and uplink, information elements **LLC Decode Errors DL (%)** and **LLC Retransmissions UL (%)**. These are similar to the RLC block error rates but concern the LLC layer.

6.4.4. Data on GPRS Connection

- Ping round-trip time
- Timeslots used
- Coding scheme usage on uplink and downlink
- Details on active PDP context or contexts: precedence class, delay class, peak throughput, etc.
- IP address assigned to RAS client
- Connection duration

6.5. GPRS Features to Be Added in Future TEMS Product Versions

6.5.1. Features Planned for TEMS Investigation GSM 4.0

- **Mobile station** features: 4+1 timeslots
- Monitoring of **PDCH utilization** at the TFI/USF level
- **Coding scheme usage** for application data (excluding signaling, which always uses CS-1)
- **Multiple data sessions** in parallel
- **RAS measurements**

6.5.2. Other Features

- **Tracing of TCP layer** from within TEMS Investigation. Today customers who want to investigate the goings-on in the TCP layer when running GPRS typically use some third-party application. This however has the drawback that the user must synchronize the TEMS data manually to the TCP trace (by exporting TEMS logfiles in text format).
- More **TBF information** (start/end of TBF event, TBF duration)
- **WAP, e-mail** (SMTP)